



Universidade de Vigo

Trabajo Fin de Máster

---

# Efectos mixtos en análisis de supervivencia: unha revisión dos modelos de fraxilidade

---

Tomé Rodríguez Rodríguez

Máster en Técnicas Estadísticas

Curso 2023-2024



## Propuesta de Trabajo Fin de Máster

<p><b>Título en galego:</b> Efectos mixtos en análise de supervivencia: unha revisión dos modelos de fraxilidade</p>
<p><b>Título en español:</b> Efectos mixtos en análisis de supervivencia: una revisión de los modelos de fragilidad</p>
<p><b>English title:</b> Mixed effects in survival analysis: a review of frailty models</p>
<p><b>Modalidad:</b> A</p>
<p><b>Autor:</b> Tomé Rodríguez Rodríguez, Universidade de Vigo</p>
<p><b>Directora:</b> María Carmen Iglesias Pérez, Universidade de Vigo</p>
<p><b>Breve resumen del trabajo:</b></p> <p>Neste traballo búscase presentar e explicar os modelos de fraxilidade en análise de supervivencia centrándose na súa versión máis popular, os modelos de fraxilidade compartida, incluíndo unha descrición xeral destes, as diferentes distribucións propostas para o seu axuste, os distintos métodos de estimación e a implementación destes en R.</p> <p>En base a que áreas de investigación se encontre que están abertas, se engadirá á literatura sobre estes modelos unha ou varias das seguintes opcións: análise da eficiencia e precisión dos métodos de estimación a través de estudos de simulación, estudo de métodos de diagnóstico, estudo de tests de bondade de axuste, aplicación a datos, ...</p>
<p><b>Recomendaciones:</b></p>
<p><b>Otras observaciones:</b></p>



Doña María del Carmen Iglesias Pérez, Catedrática de E.U. de la Universidade de Vigo informa que el Trabajo Fin de Máster titulado

**Efectos mixtos en análise de supervivencia: unha revisión dos modelos de fraxilidade**

fue realizado bajo su dirección por don Tomé Rodríguez Rodríguez para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Pontevedra, a 22 de xullo de 2024.



La directora:  
Doña María del Carmen Iglesias Pérez

El autor:  
Don Tomé Rodríguez Rodríguez

---

**Declaración responsable.** Para dar cumplimiento a la Ley 3/2022, de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster (Artículo 11, Disposición 2978 del BOE núm. 48 de 2022, [<https://www.boe.es/eli/es/1/2022/02/24/3/dof/spa/pdf>]), **el autor declara** que el Trabajo Fin de Máster presentado es un documento original en el que se han tenido en cuenta las siguientes consideraciones relativas al uso de material de apoyo desarrollado por otros/as autores/as:

- Todas las fuentes usadas para la elaboración de este trabajo han sido citadas convenientemente (libros, artículos, apuntes de profesorado, páginas web, programas,... )
- Cualquier contenido copiado o traducido textualmente se ha puesto entre comillas, citando su procedencia.
- Se ha hecho constar explícitamente cuando un capítulo, sección, demostración,... sea una adaptación casi literal de alguna fuente existente.

Y, acepta que, si se demostrara lo contrario, se le apliquen las medidas disciplinarias que correspondan.



# Índice xeral

<b>Resumo</b>	<b>IX</b>
<b>Prefacio</b>	<b>XI</b>
<b>1. Análise de supervivencia</b>	<b>1</b>
1.1. Estimación da supervivencia . . . . .	2
1.1.1. Modelos paramétricos . . . . .	3
1.1.2. Modelos non paramétricos . . . . .	4
1.2. Modelos de regresión . . . . .	6
1.2.1. Modelo de riscos proporcionais de Cox . . . . .	6
1.2.2. Modelo de tempo de fallo acelerado . . . . .	8
1.3. Transformada de Laplace . . . . .	9
<b>2. Modelos de fraxilidade</b>	<b>11</b>
2.1. Descrición dos modelos de fraxilidade . . . . .	12
2.1.1. Fraxilidade univariante . . . . .	12
2.1.2. Fraxilidade compartida . . . . .	13
2.2. Distribucións da fraxilidade . . . . .	15
2.2.1. Fraxilidade gamma . . . . .	15
2.2.2. Fraxilidade gaussiana inversa . . . . .	17
2.2.3. Fraxilidade positiva estábel . . . . .	18
2.2.4. Fraxilidade potencia-varianza . . . . .	19
2.2.5. Fraxilidade log-normal . . . . .	22
2.3. Aplicación a datos . . . . .	23
<b>3. Métodos de estimación de modelos de fraxilidade</b>	<b>25</b>
3.1. Máxima verosimilitude paramétrica . . . . .	25
3.1.1. Implementación en R . . . . .	25
3.2. Algoritmo de esperanza-maximización . . . . .	26
3.2.1. Implementación en R . . . . .	27
3.3. Verosimilitude parcial penalizada . . . . .	28
3.3.1. Implementación en R . . . . .	29
3.4. Verosimilitude xerárquica . . . . .	29
3.4.1. Implementación en R . . . . .	30
3.5. Pseudoverosimilitude completa . . . . .	30
3.5.1. Implementación en R . . . . .	31
3.6. Comparación dos paquetes de estimación . . . . .	31
3.7. Robustez da distribución gamma . . . . .	32

<b>4. Métodos de diagnóstico dos modelos de fragilidade</b>	<b>35</b>
4.1. Tests de bondade de axuste do modelo de regresión . . . . .	35
4.1.1. Tests de Bagdonavičius e Nikulin e de Vonta . . . . .	35
4.1.2. Test de González-Manteiga, Martínez-Miranda e Van Keilegom . . . . .	36
4.2. Test de heteroxeneidade . . . . .	37
4.2.1. Test de Commenges-Andersen . . . . .	37
4.3. Tests de bondade de axuste da distribución de fragilidade . . . . .	39
4.3.1. Test de Shih e Louis . . . . .	39
4.3.2. Tests de Glidden e de Cui e Sun . . . . .	39
4.3.3. Test de Geerdens, Claeskens e Janssen . . . . .	40
4.3.4. Test de Munda e Legrand . . . . .	40
<b>5. Residuos de Li, Wu e Feng aplicados a modelos de fragilidade</b>	<b>43</b>
5.1. Descrición do método . . . . .	43
5.1.1. Test de heteroxeneidade . . . . .	44
5.1.2. Test de bondade de axuste do modelo de regresión . . . . .	46
5.2. Estudo de simulación . . . . .	49
5.2.1. Descrición do estudo . . . . .	49
5.2.2. Resultados . . . . .	50
5.3. Aplicación a datos . . . . .	54
5.3.1. Test de heteroxeneidade . . . . .	54
5.3.2. Test de bondade de axuste . . . . .	55
5.3.3. Modelo de regresión con interaccións . . . . .	56
<b>6. Conclusións</b>	<b>59</b>
<b>A. Distribucións relevantes</b>	<b>61</b>
A.1. Distribucións dos tempos de evento comunmente utilizadas . . . . .	61
A.2. Distribucións das fragilidades comunmente utilizadas . . . . .	62
<b>B. Código de R</b>	<b>63</b>
B.1. Nesgo da distribución das fragilidades preditas . . . . .	63
B.2. Estudo de simulación . . . . .	63
B.3. Aplicación dos métodos de diagnóstico propostos . . . . .	68
B.3.1. Test de heteroxeneidade . . . . .	68
B.3.2. Test de bondade de axuste do modelo de regresión . . . . .	69
B.3.3. Modelo de regresión con interaccións . . . . .	70
<b>C. Derivación da densidade posterior das fragilidades gamma</b>	<b>73</b>
<b>D. Táboas adicionais sobre o estudo de simulación</b>	<b>75</b>
D.1. Proporción de axustes que non converxeron . . . . .	75
D.2. Coincidencia entre tests . . . . .	76
<b>Bibliografía</b>	<b>77</b>



# Resumo

## Resumo en galego

Esta tese é una revisión dos modelos de fragilidade compartida en análise de supervivencia, que é un modelo de efectos aleatorios que asume que os membros dun clúster comparten certa susceptibilidade de experimentar o evento de interese, centrándose na súa integración co modelo de riscos proporcionais de Cox. Analízanse os modelos de fragilidade dende o punto de vista teórico, os métodos de estimación existentes e as súas implementacións en R, así como os métodos de diagnóstico.

Esta revisión identifica unha falta de investigación respecto dos métodos de diagnóstico para os modelos de fragilidade, á que responde coa proposición de dous tests baseados nun método de diagnóstico recentemente publicado para regresión con datos censurados. O primeiro dos tests analiza a presenza de fragilidade, mentres que o segundo avalía a bondade de axuste dun modelo de regresión con fragilidade, con dúas versións cada unha das cales é máis axeitada para un tamaño de clúster.

Por último, esta memoria presenta un estudo numérico para avaliar a validez empírica destes métodos, xunto cunha aplicación a datos reais. Os resultados indican que o primeiro test funciona correctamente, demostrando a súa eficacia, mentres que o outro amosa boa potencia malia algúns problemas baixo a hipótese nula. Á luz de todo o visto anteriormente, o traballo remata cunha discusión sobre a utilidade dos modelos de fragilidade e en concreto dos métodos de diagnóstico propostos e propóñense liñas de investigación futuras.

## English abstract

This paper is a revision of shared frailty models in survival analysis, which is a random effects model that assumes that cluster members share a certain susceptibility to experience the event of interest, focusing on its integration with the Cox proportional hazards model. Frailty models are analysed from a theoretical point of view, as well as the existing estimation methods and their implementations in R, and the diagnostic methods.

This revision identifies a shortage of research on diagnostic methods for frailty models, which is addressed by proposing two tests based on a recently published diagnostic method for regression with censored data. The first two tests analyse the presence of frailty, while the second assesses the goodness of fit of a regression model with frailty, with two versions each of which is more accurate for a cluster size.

Finally, this thesis presents a numerical study to assess the empirical validity of these methods, along an application to real data. The results indicate that the first test works correctly, demonstrating its effectiveness, while the other has good power despite some issues under the null hypothesis. In light of everything seen until then, this piece of research ends with a discussion on the usefulness of the frailty models and in particular the proposed diagnostic methods, and future lines of research are proposed.



# Prefacio

Os modelos mixtos en análise de supervivencia son un campo da estatística cunha literatura extensa e unha variedade de implementacións nos softwares estatísticos máis utilizados cunha interesante variedade de aplicacións prácticas. Así e todo, é unha rama da estatística cun uso limitado, mesmo en contextos en que podería aportar ferramentas útiles e frecuentemente non se inclúe nos cursos de análise de supervivencia clásica.

O obxectivo deste traballo é estudar os modelos de fraxilidade, un modelo para a inclusión de efectos aleatorios en análise de supervivencia que foi introducido orixinalmente por D. G. Clayton en 1978 e nomeado en 1979 por J. W. Vaupel et al., utilizando terminoloxía derivada da xerontoloxía, onde fraxilidade indica a maior probabilidade de certos individuos de enfermarse e sufrir certas lesións (Wienke, 2003). En concreto, estúdanse estes modelos no contexto dos modelos de riscos proporcionais. Dende a súa orixe, como se amosará ao longo deste traballo, existiu un gran desenvolvemento teórico destes métodos, acompañado por unha variedade de implementacións prácticas. O foco deste traballo está nos modelos de fraxilidade compartida, onde se asume que os individuos dunha poboación se distribúen en clusters que comparten a susceptibilidade de experimentar un evento de interese. Estes modelos trátanse no contexto dos riscos proporcionais de Cox, onde se asume que as covariábeis teñen un efecto multiplicativo na función de risco.

A estrutura deste traballo é a seguinte. En primeiro lugar, no **capítulo 1** introdúcense conceptos básicos da análise de supervivencia que son relevantes para a comprensión e estudo dos modelos de fraxilidade, principalmente a regresión en análise de supervivencia, e introdúcese tamén a transformada de Laplace que, como se verá, ten un lugar central no estudo dos modelos de fraxilidade

No **capítulo 2** introdúcese a teoría dos modelos de fraxilidade e as distribucións máis comunmente utilizadas para a modelización das fraxilidades, estudando brevemente a relación entre algunhas distribucións e mencionando propiedades interesantes dos modelos de fraxilidade.

A continuación, no **capítulo 3** vense os modelos de fraxilidade dende un enfoque máis práctico, xa que se estudan os distintos métodos de estimación, presentando en cada caso as implementacións existentes en R, inclúese unha comparación destas implementacións e discútense o efecto dunha incorrecta especificación da distribución de fraxilidade.

No **capítulo 4** pásase ao estudo dos métodos de diagnóstico existentes para detectar a incorrecta especificación do axuste do modelo de regresión ou da distribución das fraxilidades, e para comprobar a presenza de fraxilidade na poboación. Neste capítulo vese que os métodos de diagnóstico da bondade de axuste do modelo de regresión en modelos con fraxilidade é unha área que precisa de máis investigación

No **capítulo 5** achégase investigación neste campo, propondo unha serie de tests de diagnóstico de modelos de fraxilidade, facendo unha extensión dun método recentemente proposto para o estudo de modelos de regresión para datos censurados, e facendo un estudo de simulación para estudar a validez empírica dos mesmos. O capítulo remata cunha aplicación dos métodos propostos a datos reais.

Por último, á luz de todo o presentado, no **capítulo 6** discútense a utilidade dos modelos de fraxilidade,

e a aplicacións dos mesmos, posibles rumbos para a investigación así como os puntos fortes e limitacións da memoria.

# Capítulo 1

## Análise de supervivencia

Neste capítulo introdúcese as nocións básicas da análise de supervivencia, certa notación relevante, algúns modelos paramétricos para modelizar datos de supervivencia, os métodos de estimación máis populares e os modelos de regresión en análise de supervivencia, xunto cos seus métodos de diagnóstico. Ademais, introdúcese a transformada de Laplace no contexto deste traballo que, como se verá máis adiante, ten moita importancia na estimación dos modelos de supervivencia con fraxilidade. Este capítulo segue principalmente o libro de Wienke (2010) e engádense referencias cando se inclúe contido doutras fontes.

A análise de supervivencia define unha colección de procedementos de análise estatística onde a variábel de interese é o tempo ata a ocorrencia dun evento ben definido, a miúdo chamado tempo de evento ou de fallo. Este evento é frecuentemente a morte dun individuo, daí o nome da disciplina.

Dacordo con Clark et al. (2003), a motivación destes métodos é a distribución asimétrica dos datos, que non poden ser negativos, así como o evento da censura, que limita a información dispoñíbel sobre o valor dalgúns datos. Se a distribución dos tempos de fallo fose normal e estes fosen totalmente observados, poderían utilizarse os métodos da estatística clásica.

A censura é un fenómeno polo que o tempo do evento non se observa, senón que só se observa que non ocorreu durante certo período ou, en casos que non se consideran neste traballo, períodos. Isto pode ocorrer porque un individuo abandone o estudo, porque non experimente o evento durante o tempo de seguimento ou porque experimente outro suceso que impida a ocorrencia do evento de interese. O caso máis común é a censura pola dereita, onde se sabe que o evento non ocorreu ata certo tempo, pero non se sabe cando ocorreu. Na maioría dos casos adoita suporse que a censura non ten relación co evento de interese, como por exemplo cando un individuo non morreu ao final do tempo de seguimento dun estudo, ou morreu por unha causa distinta á de interese. Esta é coñecida como censura non informativa. Neste traballo o foco estará unicamente nesta clase de censura.

En análise de supervivencia polo xeral para un individuo observamos o par de datos  $(T, \delta)$

$$\begin{aligned} T &= \min(T^*, C) \\ \delta &= \mathbb{1}(T^* \leq C), \end{aligned}$$

onde  $T^*$  é o tempo ata o evento de interese e  $C$  o tempo de censura. É dicir, para un individuo sabemos o mínimo entre o tempo ata o evento e o tempo de censura e sabemos con cal das dúas opcións se corresponde ese tempo. A variábel  $\delta$  indica se un individuo experimentou o evento de interese,  $\delta = 1$ , ou se o dato dese individuo foi censurado,  $\delta = 0$ .

Como é habitual no campo da estatística, neste traballo úsanse letras maiúsculas, por exemplo  $T$ , para falar dunha variábel aleatoria e minúsculas, como  $t$ , para falar dunha realización da variábel. Faise

unha excepción con  $\Delta$  e  $\delta$  para as que se usa  $\delta$  indistintamente, xa que  $\Delta$  é unha notación pouco utilizada neste ámbito.

A distribución dunha variábel aleatoria, neste caso os tempos de evento, está completa e unicamente definida pola súa función de distribución, que se denota con  $F(\cdot)$ , neste caso,  $F(t)$ , que dá a probabilidade de que un valor aleatorio  $T^*$  sexa menor ou igual a  $t$ ,

$$F(t) = P(T^* \leq t).$$

$F(t)$  é non decrecente en  $t$ , continua pola dereita e tal que  $F(-\infty) = 0$  e  $F(\infty) = 1$ .

Outra función de interese é a función de densidade  $f(t)$ , que é a derivada de  $F(t)$ . Neste traballo asúmese que  $T^*$  é unha variábel aleatoria continua positiva, con función de densidade  $f(t) = dF(t)/dt$ .

Na análise de supervivencia adoitan estudarse, ademais, outras 3 funcións. A primeira destas é a función de supervivencia  $S(t)$ , que indica a probabilidade de que o suceso aínda non ocorrese en tempo  $t$

$$S(t) = P(T^* > t) = 1 - F(t) = \int_t^\infty f(u)du.$$

A segunda é a función de risco,  $\lambda(t)$ , que indica a probabilidade de que un individuo que está sendo observado en tempo  $t$  teña un evento nese exacto momento

$$\lambda(t) = \lim_{\epsilon \rightarrow 0} \frac{P(T^* \leq t + \epsilon | T^* > t)}{\epsilon} = \frac{f(t)}{S(t^-)} \quad (1.1)$$

Esta é a miúdo a función máis utilizada pola facilidade de interpretación. Esta función indica o número de eventos esperados en tempo  $t$ .

A terceira é a función de risco acumulado, que é a integral entre 0 e  $t$  da función de risco

$$\Lambda(t) = \int_0^t \lambda(u)du$$

Estas funcións están claramente vinculadas xa que:

$$\Lambda(t) = \int_0^t \lambda(u)du = \int_0^t \frac{f(u)}{1 - F(u)}du = -\log(1 - F(t))$$

e, polo tanto

$$S(t) = 1 - F(t) = e^{-\int_0^t \lambda(u)du} = e^{-\Lambda(t)}. \quad (1.2)$$

## 1.1. Estimación da supervivencia

Polo xeral, en base aos datos observados dunha mostra de tamaño  $n$ ,  $i = 1, \dots, n$ , trataremos de estimar as funcións de risco e supervivencia descoñecidas. Para iso usamos a función de verosimilitude dos datos, que é o produto da función de densidade de cada par de datos observados  $(T_i, \delta_i)$ . A densidade dos datos observados é:

$$f(T_i, \delta_i) = f(T_i)^{\delta_i} (1 - F(T_i))^{1 - \delta_i} \times g(T_i)^{1 - \delta_i} (1 - G(T_i))^{\delta_i}$$

onde  $g(T)$ ,  $G(T)$  son respectivamente as funcións de densidade e distribución dos tempos de censura (Teorema 2.1 de Wienke (2010)). Polo tanto a función de verosimilitude, denotada  $L$ , para un conxunto de datos observados  $\{(T_i, \delta_i), i = 1, \dots, n\}$  é

$$L = \prod_{i=1}^n f(T_i, \delta_i) = \prod_{i=1}^n f(T_i)^{\delta_i} (1 - F(T_i))^{1 - \delta_i} g(T_i)^{1 - \delta_i} (1 - G(T_i))^{\delta_i}.$$

Dado que consideramos exclusivamente censura non informativa, os tempos de censura non están relacionados cos tempos de evento e polo tanto a verosimilitude é proporcional ao produto das funcións de densidade dos tempos de evento

$$L = \prod_{i=1}^n f(T_i, \delta_i) \propto \prod_{i=1}^n f(T_i)^{\delta_i} (1 - F(T_i))^{1-\delta_i} \quad (1.3)$$

que contén tódolos parámetros ou valores que buscamos estimar. Co cal, podemos traballar unicamente con esta expresión para estimar os parámetros de interese.

Utilizando a expresión (1.2), a expresión (1.3) pode reescribirse como

$$\begin{aligned} L &= \prod_{i=1}^n f(T_i, \delta_i) \propto \prod_{i=1}^n (\lambda(T_i)S(T_i))^{\delta_i} S(T_i)^{1-\delta_i} \\ &\propto \prod_{i=1}^n \lambda(T_i)^{\delta_i} S(T_i) \end{aligned} \quad (1.4)$$

### 1.1.1. Modelos paramétricos

Para estimar a función de distribución dos tempos de supervivencia, no caso de dispór de información sobre a forma desta, pódense substituír as funcións de supervivencia e risco na expresión (1.4) e maximizala respecto dos parámetros da distribución para obter estimacións destes e construír a función de distribución estimada. Cando en efecto se dispón de información sobre a distribución subxacente aos tempos de evento, a utilización dunha expresión paramétrica para a función de distribución, en comparación con métodos non paramétricos (descritos na sección 1.1.2), é máis eficiente.

Algunhas distribucións popularmente utilizadas para estimar tempos de supervivencia son a exponencial, a Weibull, a Gompertz, a log-normal e a log-loxística.

Unha táboa coas funcións mencionadas anteriormente, a súa esperanza e varianza, cando existe e é sinxela, pode encontrarse no apéndice A.1. A continuación introdúcense de xeito sintético. A notación segue principalmente o libro de Klein e Moeschberger (2003).

#### Distribución exponencial

A distribución exponencial é a distribución máis sinxela de entre as utilizadas para modelizar tempos de evento. A distribución exponencial depende dun único parámetro  $\lambda$ , que é a súa función de risco constante,

$$\lambda(t) = \lambda, \quad (\lambda > 0).$$

A simpleza desta distribución leva a unha falta de flexibilidade que fai que en raras ocasións sexa axeitada para describir os datos.

#### Distribución de Weibull

Outra distribución popular é a distribución de Weibull, que é similar á exponencial pero máis flexíbel grazas á introdución dun segundo parámetro  $\alpha$ , que permite describir un crecemento ou decrecemento no risco. A función de risco Weibull é

$$\lambda(t) = \alpha \lambda t^{\alpha-1}, \quad (\lambda, \alpha > 0).$$

É doado ver que a distribución exponencial é un caso especial desta distribución con  $\alpha = 1$ .

### Distribución de Gompertz

A distribución de Gompertz é outra opción, caracterizada por unha función de risco crecente. A función de risco vén dada por

$$\lambda(t) = \lambda e^{\varphi t}, \quad (\lambda, \varphi > 0).$$

A función exponencial é un caso límite desta distribución cando  $\varphi \rightarrow 0$ .

### Distribución log-normal

Cando se usa a distribución log-normal suponse que o logaritmo dos tempos de distribución segue unha distribución normal de parámetros  $m$  e  $s$ ,  $\log(T) \sim N(m, s)$ . Isto leva á función de risco

$$\lambda(t) = \frac{\frac{1}{st} \phi\left(\frac{\log(t)-m}{s}\right)}{1 - \Phi\left(\frac{\log(t)-m}{s}\right)}, \quad (s > 0, m \in \mathbb{R}),$$

con  $\phi(\cdot)$  e  $\Phi(\cdot)$  as funcións de densidade e distribución dunha distribución normal estándar.

### Distribución log-loxística

Por último, a distribución log-loxística tamén é unha distribución de dous parámetros denotada  $\text{logL}(v, \kappa)$ . A distribución pode caracterizarse pola función de risco

$$\lambda(t) = \frac{v\kappa(vt)^{\kappa-1}}{1 + (vt)^\kappa}, \quad (v, \kappa > 0).$$

## 1.1.2. Modelos non paramétricos

Unha opción alternativa aos modelos paramétricos, cando non se dispón de información sobre a distribución dos tempos de evento, é axustar as funcións de interese de xeito non paramétrico.

Dous estimadores non paramétricos son especialmente coñecidos; o estimador da supervivencia de Kaplan Meier e o estimador do risco acumulado de Nelson-Aalen.

### Estimador de Kaplan-Meier

A curva de Kaplan-Meier é unha función discontinua, con discontinuidades de salto nos tempos nos que se observou un evento, que se usa para estimar a función de supervivencia. Pode probarse que é o estimador non paramétrico que maximiza a verosimilitude dada en (1.4) e a súa expresión é a seguinte:

$$\hat{S}(t) = \prod_{i:t_{(i)} \leq t} \left(1 - \frac{\delta_{(i)}}{n - i + 1}\right).$$

con  $(i)$  a posición ordinal dos tempos observados,  $t_{(i-1)} \leq t_{(i)} \leq t_{(i+1)}$ , e  $\delta_{(i)}$  o indicador asociado ao tempo  $t_{(i)}$ .

Dado que o tempo é unha variábel continua, dous eventos non deberían coincidir no tempo, pero na práctica pode suceder. Son os chamados empates. Para lidiar cos empates, usemos o índice  $l$  para denotar aos tempos de observación distintos,  $l = 1, \dots, L$ , con  $L \leq \sum_i \delta_i \leq n$ . Usando esta notación o estimador de Kaplan-Meier é

$$\hat{S}(t) = \prod_{l \in R(t)} \left(1 - \frac{d(t_l)}{\#R(t_l)}\right). \quad (1.5)$$

con  $d(t_l)$  o número de eventos en  $t_l$ ,  $R(t_l)$  os individuos aínda baixo risco en  $t_l$ , é dicir tódolos individuos que non experimentaron o evento de interese ou foron censurados antes de  $t_l$ , e  $\#R(t_l)$  o número destes.



Na figura 1.1 poden verse graficamente dous exemplos de estimadores de Kaplan-Meier para datos distribuídos segundo dúas distribucións Weibull. Un problema do estimador de Kaplan-Meier é que non

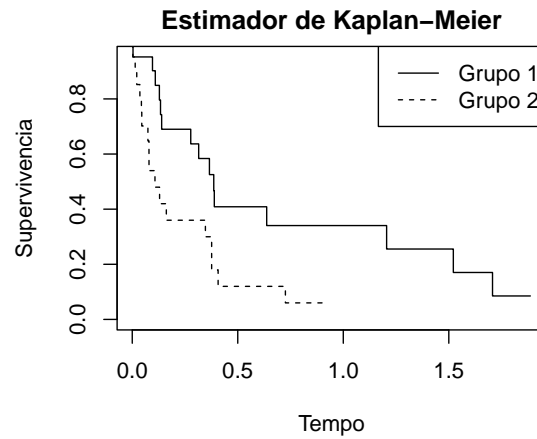


Figura 1.1: Dúas curvas de supervivencia estimadas usando o estimador de Kaplan-Meier

está definido máis alá do último tempo de evento observado cando a última observación é censurada. Habitualmente a función non se define máis alá deste límite.

### Estimador de Nelson-Aalen

O estimador de Nelson-Aalen da función de risco acumulado é, tamén, unha función con discontinuidades de salto nos tempos de evento. Usando a notación da expresión (1.5) escríbese

$$\hat{\Lambda}(t) = \sum_{l \in R(t)} \frac{d(t_l)}{\#R(t_l)}.$$

Dous exemplos de curvas de risco basal estimadas usando o estimador de Nelson-Aalen poden verse na figura 1.2.

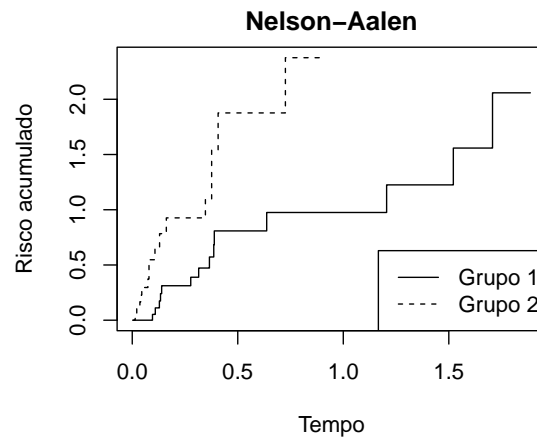


Figura 1.2: Dúas funcións de risco acumulado estimadas usando o estimador de Nelson-Aalen

## 1.2. Modelos de regresión

Ata agora consideramos unicamente o caso máis sinxelo de variábeis independentes e idénticamente distribuídas, baixo a suposición de que a poboación é homoxénea. Porén, en moitas ocasións terá interese estudar o efecto de certas diferenzas entre os individuos no tempo de fallo. Por exemplo, nun estudo sobre o tempo de supervivencia dos pacientes con certa enfermidade, os pacientes poden diferenciarse na idade, no grao da doenza ou nas terapias recibidas. Algunhas destas covariábeis poden ser de especial importancia, como o efecto dunha terapia, noutras situacións o interese pode estar en engadilas para obter estimacións máis axustadas. En ámbolos dous casos utilizaremos modelos de regresión.

Supoñemos que para cada individuo  $i$ , con  $i = 1, \dots, n$ , observamos unha serie de  $p$  covariábeis  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$ . Supoñemos ademais que estas covariábeis teñen un efecto na supervivencia, que é común a tódolos individuos. O interese estará en estimar este efecto; en que medida un valor distinto desta covariábel aumenta ou diminúe a probabilidade de supervivencia. Para estimalo, introducimos na expresión da verosimilitude (1.4) unha dependencia aos valores observados de  $\mathbf{X}$ .

$$L = \prod_{i=1}^n f(T_i, \delta_i | \mathbf{X}_i) \propto \prod_{i=1}^n \lambda(T_i | \mathbf{X}_i)^{\delta_i} S(T_i | \mathbf{X}_i).$$

Nótese que a **negriña** indica que se trata dun vector ou unha matriz en lugar de dun escalar.

De cara a utilizar esta verosimilitude, cómpre definir a forma que ten  $\lambda(t_i | \mathbf{X}_i)$  e, por extensión  $S(t_i | \mathbf{X}_i)$ . O modelo máis popular é o modelo de riscos proporcionais de Cox.

### 1.2.1. Modelo de riscos proporcionais de Cox

O modelo de riscos proporcionais supón que as covariábeis teñen un efecto multiplicativo sobre a función de risco, de xeito que para un individuo, a función de risco e supervivencia son

$$\lambda(t | \mathbf{X}) = \lambda_0(t) e^{\boldsymbol{\beta}^t \mathbf{X}} \quad (1.6)$$

$$S(t | \mathbf{X}) = e^{-\Lambda_0(t) \exp(\boldsymbol{\beta}^t \mathbf{X})}. \quad (1.7)$$

A función de risco,  $\lambda(t | \mathbf{X})$ , está formada polo risco basal  $\lambda_0(t)$  e o resultado do produto matricial  $\boldsymbol{\beta}^t \mathbf{X} = (\beta_1 X_1 + \dots + \beta_p X_p)$  (Klein & Moeschberger, 2003; Bradburn et al., 2003).

$\Lambda_0(t)$  é a integral de 0 a  $t$  de  $\lambda_0(u)$ , do mesmo xeito que en (1.2). É sinxelo ver que o risco basal se corresponde coa función de risco medio no caso de estaren as variábeis centradas, e co risco para un individuo que tivese valor 0 para tódalas covariábeis se non están centradas.

Este modelo recibe o nome de modelo de riscos proporcionais dado que se miramos ao ratio das funcións de risco de dous individuos con covariábeis  $\mathbf{X} = \mathbf{x}$  e  $\mathbf{X} = \mathbf{x}^*$  temos:

$$\frac{\lambda(t | \mathbf{x})}{\lambda(t | \mathbf{x}^*)} = \frac{\lambda_0(t) e^{\boldsymbol{\beta}^t \mathbf{x}}}{\lambda_0(t) e^{\boldsymbol{\beta}^t \mathbf{x}^*}} = \exp[\boldsymbol{\beta}^t (\mathbf{x} - \mathbf{x}^*)],$$

que é un valor constante no tempo. Polo tanto, describe funcións de risco proporcionais entre os individuos (Klein & Moeschberger, 2003).

De cara a estimación dos parámetros de regresión  $\boldsymbol{\beta}$  no modelo de Cox, utilízase a verosimilitude parcial. Na verosimilitude parcial cada elemento do produto é a probabilidade de que cada individuo experimente o evento, dado o número de individuos que están en risco nese momento. Asumindo que

non hai empates para simplificar a presentación, a función resultante é o produto destas probabilidades para tódolos individuos que experimentaron o evento

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^n \left( \frac{\lambda_0(T_i) e^{\boldsymbol{\beta}^t \mathbf{X}_i}}{\sum_{j \in R(T_i)} \lambda_0(T_i) e^{\boldsymbol{\beta}^t \mathbf{X}_j}} \right)^{\delta_i} \\ &= \prod_{i=1}^n \left( \frac{e^{\boldsymbol{\beta}^t \mathbf{X}_i}}{\sum_{j \in R(T_i)} e^{\boldsymbol{\beta}^t \mathbf{X}_j}} \right)^{\delta_i} \end{aligned} \quad (1.8)$$

onde  $\mathbf{X}_i$  e  $\mathbf{X}_j$  son os vectores de covariábeis dos individuos  $i$  e  $j$ , e  $R(t_i)$  os índices dos individuos en risco en tempo  $t_i$  (Klein & Moeschberger, 2003). Dado que o risco basal está en tanto no divisor como no dividendo pódese eliminar de xeito que as estimacións das  $\boldsymbol{\beta}$  non dependen da forma do risco basal.

### Modelo de riscos proporcionais semiparamétrico

Logo de obter as estimacións dos parámetros  $\boldsymbol{\beta}$  pódese obter unha estimación do risco usando

$$\hat{\lambda}(t_l) = \frac{d(t_l)}{\sum_{h \in R(t_l)} e^{\hat{\boldsymbol{\beta}}^t \mathbf{X}_h}}. \quad (1.9)$$

No caso de  $\hat{\boldsymbol{\beta}} = 0$  convértese no estimador de Nelson-Aalen. A función de supervivencia basal estimada é polo tanto

$$\hat{S}_0(t) = e^{-\hat{\Lambda}_0(t)},$$

onde  $\hat{\Lambda}_0(t)$  é o estimador de Breslow da función de risco acumulado

$$\hat{\Lambda}_0(t) = \sum_{l=1}^r \frac{\mathbb{1}(t_l \leq t) d(t_l)}{\sum_{h \in R(t_l)} e^{\boldsymbol{\beta}^t \mathbf{X}_h}} \quad (1.10)$$

A estimación da supervivencia dun individuo en concreto sería

$$\hat{S}(t|\mathbf{X}) = \hat{S}_0(t)^{\exp(\hat{\boldsymbol{\beta}}^t \mathbf{X})}.$$

### Modelo de riscos proporcionais paramétrico

Unha alternativa ao modelo semiparamétrico en (1.9) é supór que os tempos de evento se distribúen seguindo algunha distribución, como as descritas anteriormente e estimar os parámetros da familia de distribucións correspondente a partir da función de verosimilitude dada por:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1}^n (\lambda_0(T_i|\boldsymbol{\theta}) \exp(\boldsymbol{\beta}^t \mathbf{X}_i))^{\delta_i} e^{-\Lambda_0(T_i|\boldsymbol{\theta}) \exp(\boldsymbol{\beta}^t \mathbf{X}_i)},$$

que no caso dunha distribución exponencial con parámetro lambda é

$$L(\boldsymbol{\beta}, \lambda) = \prod_{i=1}^n (\lambda e^{\boldsymbol{\beta}^t \mathbf{X}_i})^{\delta_i} e^{-\lambda T_i \exp(\boldsymbol{\beta}^t \mathbf{X}_i)}$$

e para a Weibull é

$$L(\boldsymbol{\beta}, \lambda, \alpha) = \prod_{i=1}^n (\alpha \lambda T_i^{\alpha-1} e^{\boldsymbol{\beta}^t \mathbf{X}_i})^{\delta_i} e^{-\lambda T_i^\alpha \exp(\boldsymbol{\beta}^t \mathbf{X}_i)}$$

Este método pode presentar certas vantaxes en tanto que é máis eficiente cando se escolle unha familia de distribucións axeitada aos datos, e que permite obter predicións das probabilidades de supervivencia en tempos de evento non observados (Clark et al., 2003).

### Métodos de diagnóstico dun modelo de riscos proporcionais

No coñecido libro *Survival analysis; techniques for censored and truncated data* (Klein & Moeschberger, 2003), recóllense 2 métodos para analizar a bondade de axuste dun modelo de riscos proporcionais, os residuos de Cox-Snell e os residuos de martingala. Ademais destes preséntanse métodos para o contraste da suposición de proporcionalidade e a detección de datos influíntes e/ou atípicos, pero estes van alén do foco desta monografía.

**Residuos de Cox-Snell:** Os residuos de Cox-Snell,  $\mathbf{r}$ , defínense como

$$r_i = \hat{\Lambda}_0(T_i) \exp \left( \sum_{h=1}^p \hat{\beta}_h^t X_{ih} \right) \quad i = 1, \dots, n, \quad (1.11)$$

onde  $\hat{\Lambda}_0(T_i)$  e  $\hat{\beta}$  son as estimacións do risco basal e dos parámetros de regresión respectivamente.

Se o modelo é correcto e as  $\hat{\beta}$  se achegan ao valor real das  $\beta$ , os valores de  $\mathbf{r}$  deberan ser unha mostra censurada dunha distribución exponencial. Para comprobar se en efecto é así pode computarse o estimador de Nelson-Aalen do risco basal, e facer un gráfico deste contra os  $\mathbf{r}$ . Este debería ser aproximadamente unha liña recta con orixe en 0 e pendente 1.

**Residuos martingala:** Para datos con censura pola dereita e covariábeis constantes no tempo, que é o enfoque desta monografía, os residuos martingala,  $\hat{M}$  son

$$\hat{M}_i = \delta_i - r_i \quad i = 1, \dots, n$$

con  $r$  o residuo de Cox-Snell.

Para un tamaño de mostra importante os  $\hat{M}$  son unha mostra incorrelada dunha distribución de media 0. No caso de coñecer e usar  $\beta$  e  $\Lambda_0$  en lugar das estimacións, os residuos serían unha martingala. Se se fai unha gráfica destes residuos contra os valores de cada covariábel, debería amosar ruído branco. Se se observa unha tendencia pode indicar que é preciso aplicar unha transformación a covariábel para encontrar a forma do seu efecto.

### 1.2.2. Modelo de tempo de fallo acelerado

O modelo de tempo de fallo acelerado (AFT, polo seu acrónimo en inglés) é outro enfoque popular para os modelos de regresión. Aínda que este traballo trata principalmente co modelo de Cox, o modelo de fallo de tempo acelerado preséntase para dar completude ao traballo.

O modelo toma a forma

$$S(t) = S_0(\exp(-\beta^t \mathbf{X})t)$$

de xeito que  $\exp(\beta^t \mathbf{X})$  actúa como factor de aceleración (Bradburn et al., 2003). Este modelo exprésase a miúdo na súa forma log-linear como

$$\log(t) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

con  $\varepsilon$  un termo de erro. Neste modelo un valor das  $\beta$  maior a 1 *ralentiza* a chegada do evento, mentres que un valor menor ten un efecto *acelerador*. Usando este enfoque asúmese que os tempos de supervivencia seguen unha distribución como a Weibull, log-normal ou log-loxística.

Dado que é un modelo completamente paramétrico poden obterse as estimacións dos parámetros de xeito sinxelo usando a verosimilitude en (1.4).

### 1.3. Transformada de Laplace

É importante, como veremos máis adiante, introducir un concepto relevante no estudo dos modelos de fraxilidade; a transformada de Laplace ou laplaciana e as súas derivadas.

Dada certa función dunha variábel aleatoria  $Z$ ,  $f(Z)$ , como por exemplo unha función de densidade, a transformación de Laplace defínese como

$$\mathcal{L}(u) = \int_0^{\infty} f(z)e^{-zu} dz, \quad (1.12)$$

con  $\mathcal{L}(u)$  unha función que depende só de  $u$ . Nos modelos de fraxilidade a transformada de Laplace ten interese porque permite obter a esperanza en  $Z$  de  $e^{-Zu}$ , sendo  $Z$  a fraxilidade.

A transformada de Laplace existe sempre que a integral no lado dereito da igualdade converxa para algún valor de  $u$  (Spiegel, 1965).

Por exemplo, asumindo que  $Z$  segue unha distribución gamma de parámetros  $\theta$  e  $k$ ,  $f(z) = \frac{1}{\Gamma(k)\theta^k} z^{k-1} e^{-z/\theta}$ , e comezando con  $u = \lambda - \frac{1}{\theta}$  como se fai a miúdo para simplificar a derivación da laplaciana, temos que

$$\begin{aligned} \mathcal{L}\left(\lambda - \frac{1}{\theta}\right) &= \int_0^{\infty} \frac{z^{k-1} e^{-z\lambda}}{\Gamma(k)\theta^k} dz \\ &= \frac{1}{\Gamma(k)\theta^k} \int_0^{\infty} z^{k-1} e^{-z\lambda} dz \\ &= \frac{\Gamma(k)}{\Gamma(k)\theta^k \lambda^k} \\ &= (\theta\lambda)^{-k}, \end{aligned}$$

dado que  $\lambda = u + \frac{1}{\theta}$ , temos que a transformada laplaciana para  $u$  é

$$\begin{aligned} \mathcal{L}(u) &= \left(\theta\left(u + \frac{1}{\theta}\right)\right)^{-k} \\ &= (1 + \theta u)^{-k}. \end{aligned} \quad (1.13)$$

Igualmente importantes son as  $k$  derivadas da transformada de Laplace, que teñen a seguinte forma explícita (Duchateau & Janssen, 2008)

$$\begin{aligned} \mathcal{L}^{(k)}(u) &= \frac{\partial^k}{\partial u^k} \int_0^{\infty} f(z)e^{-zu} dz \\ &= (-1)^k \int_0^{\infty} f(z)z^k e^{-zu} dz \end{aligned} \quad (1.14)$$

No apéndice A.2, como referencia, poden encontrarse as transformadas de Laplace dalgunhas distribucións relevantes neste traballo.



## Capítulo 2

# Modelos de fragilidade

Neste capítulo mótvanse e explícanse os modelos de fragilidade e, a continuación, introdúcense algunhas das distribucións máis populares para a modelización das fragilidades. Tamén se comentan algúns detalles sobre a natureza da dependencia axustada usando distintas distribucións para modelizar as fragilidades. Este capítulo segue principalmente os libros de Duchateau e Janssen (2008) e de Wienke (2010), aínda que segue unha estrutura diferente. Mentres que ambos libros presentan en primeiro lugar o modelo de fragilidade gamma, e explican usando esta distribución a modo de exemplo os modelos de fragilidade, este traballo opta por explicar en primeiro lugar certas características dos modelos de fragilidade, como a función de verosimilitude e a estimación dos valores individuais das fragilidades, independentemente da distribución das fragilidades.

A noción de fragilidade proporciona un concepto interesante para a introdución dos modelos de efectos aleatorios na análise de supervivencia. Na súa forma máis común, a fragilidade é unha variábel aleatoria  $Z$  non observada que aumenta ou reduce o risco dun individuo ou dun grupo de individuos de experimentar certo evento (Wienke, 2003). O termo provén do campo da xerontoloxía, onde se usa para indicar que certas persoas teñen maior risco de morbilidade e/ou mortandade (Wienke, 2010).

A maioría da literatura existente na fragilidade está baseada nas seguintes dúas suposicións de conveniencia matemática: (1) que a fragilidade é constante no tempo e (2) que actúa de xeito multiplicativo no risco (Wienke, 2010). Así, a función de risco é

$$\lambda(t|Z) = \lambda_0(t)Z,$$

o que indica que o perfil de risco dos individuos é o mesmo a excepción dun termo de fragilidade que aumenta o risco de xeito proporcional. A fragilidade é por definición unha variábel non negativa,  $Z \geq 0$ . Ademais, por convención, adoita fixarse a media das  $Z$  en 1, de xeito que a función de risco basal reflecta o risco medio na poboación.

Unha extensión natural deste modelo é a inclusión de covariábeis que teñen certo efecto sobre a función de risco

$$\lambda(t|\mathbf{X}, Z) = \lambda(t|\mathbf{X})Z,$$

por exemplo usando variábeis coa forma do modelo de riscos proporcionais de Cox (1.9)

$$\lambda(t|\mathbf{X}, Z) = \lambda_0(t)e^{\beta^t \mathbf{X}} Z. \tag{2.1}$$

Neste sentido o modelo de fragilidade pode ser visto como unha extensión do popular modelo de Cox. Os modelos de fragilidade serían con esta idea en mente un modelo de Cox cunha serie de covariábeis

$\mathbf{X}_1$  observadas, e toda unha serie de variábeis  $\mathbf{X}_2$  inobservábeis que levan ao modelo

$$\lambda(t|\mathbf{X}_1, \mathbf{X}_2) = \lambda_0(t)e^{\beta_1^t \mathbf{X}_1 + \beta_2^t \mathbf{X}_2}.$$

Dado que non podemos axustar este modelo polo descoñecemento das variábeis  $\mathbf{X}_2$ , asúmese que  $Z = e^{\beta_2^t \mathbf{X}_2}$  e axustamos o modelo en (2.1).

## 2.1. Descrición dos modelos de fraxilidade

Os métodos de análise de supervivencia, por exemplo o modelo de Cox, supoñen polo xeral que as observacións son independentes e están identicamente distribuídas, alén do efecto das covariábeis consideradas. Existen casos en que estas suposicións non representan a realidade dos datos dispoñíbeis. Wienke (2007) menciona 3 exemplos:

- Nun famoso estudo sobre o tempo ata a cegueira de cada ollo dunha persoa con diabetes retinopática é lóxico asumir que os dous ollos dunha persoa compartirán certas características que modificarán o risco da perda de visión.
- Cando incluímos nun estudo sobre a incidencia dunha enfermidade a membros dunha mesma familia podemos asumir que a familia pode compartir certos rasgos xenéticos, que modifican o risco de contraela.
- Cando estudamos o tempo ata o desenvolvemento dunha doenza recorrente é razoábel supór que os tempos para unha mesma persoa están correlados.

Outro exemplo, por amosar que o tamaño do clúster pode ser moito maior que nos exemplos anteriores, é asumir que en cada centro médico que participa nun ensaio clínico haberá patróns comúns no tratamento que inflúen na supervivencia dos individuos.

### 2.1.1. Fraxilidade univariante

A versión máis sinxela dos modelos de fraxilidade é a fraxilidade univariante, que asigna un termo de fraxilidade diferente  $Z$  a cada individuo. Estes modelos utilízanse en situacións nas que existe heteroxeneidade na poboación. Nesta situación, a diferenza nas fraxilidades  $Z$  leva a un menor tempo de evento promedio para aqueles individuos máis fráxiles, de xeito que esta parte da poboación abandona o estudo e a poboación observada é progresivamente menos fráxil. Nas situacións nas que existe heteroxeneidade, ignorala leva a un nesgo nas estimacións do risco (Wienke, 2010).

Na figura 2.1 pódese ver un exemplo do nesgo que se introduce na función de risco estimada ao estimala sen considerar as fraxilidades. Aínda que non se trata neste traballo, aquí se considera a efectos ilustrativos unha distribución das fraxilidades discreta que só toma valores 0.5, 1 e 2. Vemos que a función de risco incondicional medra a medida que o risco da poboación máis fráxil aumenta, pero comeza despois decrecer a medida que esta abandona a poboación total. Así, semella que a distribución do risco ten a forma dunha campá asimétrica, co risco decrecendo para aqueles que sobreviven ata certo tempo. En realidade o risco crece constantemente e a forma vén explicada por diferentes perfís de risco de distintos grupos da poboación.

Estes modelos presentan unha dificultade en tanto que a distribución das fraxilidades é extremadamente difícil de estimar en base a datos univariantes, e a selección da distribución da fraxilidade pode influír considerablemente nos resultados (Balan & Putter, 2020).

Estes modelos encontraron aplicacións prácticas á hora de explicar resultados anómalos, como a maior incidencia de cancro testicular en homes novos cun decrecemento a partir de certa idade, que podería ser o resultado dunha maior susceptibilidade de algúns homes de sufrir este cancro, que resulta nunha doenza temperá nestes homes, e unha redución dos casos a medida que quedan excluídos do



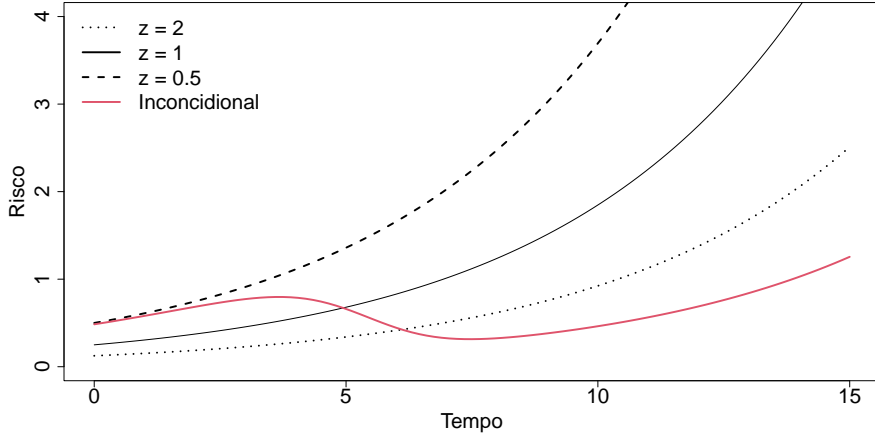


Figura 2.1: Risco condicional ás fraxilidades e incondicional dunha poboación ficticia con distribución de fraxilidade discreta

grupo (Wienke, 2010). Porén, estes modelos principalmente foron desenvolto con obxectivos teóricos (Govindarajulu & D'Agostino, 2020).

### 2.1.2. Fraxilidade compartida

O modelo máis comunmente utilizado na práctica é o modelo de fraxilidade compartida, que supón que os individuos forman grupos ou clusters que comparten características non observadas que modifican o seu risco. Polo tanto, en lugar de teren riscos individuais, tódolos membros dun grupo comparten certo risco  $Z$ . Os modelos de fraxilidade univariante poden verse como un caso especial destes modelos onde tódolos clusters teñen tamaño 1.

Formalmente, se consideramos unha poboación formada por  $n$  clusters ( $i = 1, \dots, n$ ), con  $n_i$  individuos en cada clúster ( $j = 1, \dots, n_i$ ), de xeito que a poboación total está formada por  $\sum_{i=1}^n n_i$  individuos, o modelo fórmulase como

$$\lambda(t_{ij}|\mathbf{X}_{ij}, Z_i) = \lambda_0(t_{ij})e^{\beta^t \mathbf{X}_{ij} Z_i}, \quad (2.2)$$

con  $Z_i \geq 0$  e i.i.d., con función de densidade  $f(z_i)$ , e  $\lambda_0(t_{ij})e^{\beta^t \mathbf{X}_{ij}}$  unha función de risco específica a cada individuo  $j$  do clúster  $i$  coa forma da expresión (1.6).

Os tempos de fallo dos membros dun clúster  $i$  asúmense independentes condicionalmente a  $Z_i$ . Dada esta independencia condicional tense que a función de supervivencia conxunta condicional (a  $Z_i$ ) para un clúster  $i$  é igual ao produto das funcións de supervivencia dos seus membros  $j = 1, \dots, n_i$

$$S(t_{i1}, \dots, t_{in_i}|\mathbf{X}_{ij}, Z_i) = \prod_{j=1}^{n_i} \exp\left(-\Lambda_0(t_{ij})e^{\beta^t \mathbf{X}_{ij} Z_i}\right) = \exp\left(-Z_i \sum_{j=1}^{n_i} \Lambda_0(t_{ij})e^{\beta^t \mathbf{X}_{ij}}\right). \quad (2.3)$$

No sucesivo usarei a  $\Lambda_{\text{clúster}}(\mathbf{t}_i|\mathbf{X}_i)$  para indicar o risco acumulado conxunto condicionado a  $Z_i$ , que é

$$\sum_{j=1}^{n_i} \Lambda_0(t_{ij})e^{\beta^t \mathbf{X}_{ij}} = \Lambda_{\text{clúster}}(\mathbf{t}_i|\mathbf{X}_i). \quad (2.4)$$

O risco acumulado nun clúster  $i$  alén da fraxilidade  $Z_i$  é igual á suma dos riscos dada a independencia condicional a  $Z_i$ , como se explicita na expresión (2.3).

Usando (2.2) e (2.3) e a expresión da verosimilitude en (1.3) temos que a contribución dun clúster  $i$  á función de verosimilitude condicional aos valores de  $Z_i$  é

$$\begin{aligned} L_i(\boldsymbol{\beta}|Z_i) &= \prod_{j=1}^{n_i} \left[ \left( \lambda_0(T_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} Z_i \right)^{\delta_{ij}} \right] \exp(-\Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)Z_i) \\ &= \prod_{j=1}^{n_i} \left[ \left( \lambda_0(T_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} \right)^{\delta_{ij}} \right] Z_i^{D_i} \exp(-\Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)Z_i) \end{aligned} \quad (2.5)$$

onde  $D_i$  é o número de eventos no clúster  $i$ ,  $D_i = \sum_{j=1}^{n_i} \delta_{ij}$ .

É evidente que sen observar as fraxilidades  $\mathbf{Z} = (Z_1, \dots, Z_n)$  non podemos usar a función de verosimilitude construída como o produto destas contribucións. No seu lugar teremos que maximizar a súa esperanza en  $Z$ , que é o que se observa. Así pasamos da función de verosimilitude condicional á función de verosimilitude incondicional ou marxinal

$$\begin{aligned} L_i(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \prod_{j=1}^{n_i} \left( \lambda_0(t_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} \right)^{\delta_{ij}} \mathbb{E} \left[ Z_i^{D_i} \exp(-Z_i \Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)) | \boldsymbol{\theta} \right] \\ &= \prod_{j=1}^{n_i} \left( \lambda_0(T_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} \right)^{\delta_{ij}} \int_0^\infty f(z_i|\boldsymbol{\theta}) z_i^{D_i} \exp(-z_i \Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)) dz_i, \end{aligned}$$

onde  $f(z_i|\boldsymbol{\theta})$  é o valor da función densidade das fraxilidades de parámetro (ou parámetros)  $\boldsymbol{\theta}$  en  $z_i$  e os límites da integral veñen dados polo dominio de  $Z_i$ .

Podemos ver que a contribución do clúster  $i$  sen eventos á verosimilitude incondicional  $L_i(\cdot)$  é

$$L_i(\boldsymbol{\beta}, \boldsymbol{\theta}|D_i = 0) = \int_0^\infty f(z_i|\boldsymbol{\theta}) \exp(-z_i \Lambda_{\text{clúster}}(\mathbf{t}_i|\mathbf{X}_i)) dz_i.$$

Esta igualdade é a transformada de Laplace vista en (1.12) con  $u = \Lambda_{\text{clúster}}(\mathbf{t}_i|\mathbf{X}_i)$ , co que temos que a contribución dun clúster sen eventos na anterior ecuación pódese escribir

$$L_i(\boldsymbol{\beta}, \boldsymbol{\theta}|D_i = 0) = \mathcal{L}(\Lambda_{\text{clúster}}(\mathbf{t}_i|\mathbf{X}_i)). \quad (2.6)$$

onde  $\boldsymbol{\theta}$  é o parámetro ou vector de parámetros da densidade  $f(z_i)$  que dan forma á transformada de Laplace.

Visto doutro xeito, temos que a función de supervivencia incondicional dun clúster é

$$S(\mathbf{t}_i|\mathbf{X}_i, D_i = 0) = \mathcal{L}(\Lambda_{\text{clúster}}(\mathbf{t}_i|\mathbf{X}_i)). \quad (2.7)$$

Se agora consideramos o caso con  $D_i \geq 0$ , temos

$$\begin{aligned} L_i(\boldsymbol{\beta}, \boldsymbol{\theta}) &= \prod_{j=1}^{n_i} \left[ \left( \lambda_0(T_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} \right)^{\delta_{ij}} \right] \int_0^\infty z_i^{D_i} \exp(-z_i \Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)) f(z_i|\boldsymbol{\theta}) dz_i \\ &= \prod_{j=1}^{n_i} \left[ \left( \lambda_0(T_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} \right)^{\delta_{ij}} \right] (-1)^{D_i} (-1)^{D_i} \int_0^\infty z_i^{D_i} \exp(-z_i \Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)) f(z_i|\boldsymbol{\theta}) dz_i \end{aligned}$$

e agora ao final do lado dereito temos a derivada da transformada de Laplace en (1.14) coa mesma  $u$  que antes e con  $k = D_i$ . Polo tanto

$$L_i(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{j=1}^{n_i} \left[ \left( \lambda_0(T_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} \right)^{\delta_{ij}} \right] (-1)^{D_i} \mathcal{L}^{(D_i)}(\Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)). \quad (2.8)$$

Nótese que a expresión (2.8) se reduce a (2.6) cando  $D_i = 0$ .

O produto das contribucións de cada clúster á función de verosimilitude resulta na función de verosimilitude incondicional. A maximización da función de verosimilitude respecto dos parámetros de regresión  $\beta$  e do parámetro ou parámetros da función de distribución das fraxilidades  $\theta$  permite obter estimacións dos mesmos,  $(\hat{\beta}, \hat{\theta})$  e, en base a estas, estimar a función de risco e de supervivencia  $S(t|\mathbf{X})$  e da función de densidade  $f(t|\mathbf{X})$  incondicionais. Como se verá no capítulo 3 esta maximización é máis complexa do habitual cando o modelo de regresión é semiparamétrico.

Logo de obter estas estimacións, usando a regra de Bayes, como se explica na páxina 66 de Duchateau e Janssen (2008), e na 62 de Wienke (2010) pódese estimar a densidade posterior de  $Z_i$  usando

$$f(z_i|\mathbf{T}_i, \delta_i, \hat{\beta}, \hat{\theta}, \hat{\lambda}, \mathbf{X}_i) = \frac{f(\mathbf{T}_i, \delta_i|\hat{\beta}, \hat{\lambda}, \mathbf{X}_i, z_i)f(z_i|\hat{\theta})}{f(\mathbf{T}_i, \delta_i|\hat{\beta}, \hat{\theta}, \hat{\lambda}, \mathbf{X}_i)} \quad (2.9)$$

e predicir a fraxilidade de cada clúster como o valor medio de  $Z_i$  dada a densidade estimada. Nótese que  $\hat{\lambda}$  é a función de risco estimada, que pode ser paramétrica ou non paramétrica.

## 2.2. Distribucións da fraxilidade

Unha característica importante á hora de axustar un modelo de fraxilidade é escoller a distribución que se supón xeneradora das fraxilidades. En vista da importancia da transformada de Laplace, nos inicios do desenvolvemento destes modelos optouse principalmente por distribucións para as que esta existise, de xeito que a variábel de fraxilidade se puidese extraer da función de supervivencia a través da integración e así obter unha función de supervivencia non condicional que dependa só dos valores observados, como se trata máis adiante, o que facilita as rutinas de estimación (Wienke, 2010). Algúns exemplos de distribucións non negativas con transformada laplaciana sinxela son a gamma, gaussiana inversa, positiva estábel, e potencia-varianza (Duchateau & Janssen, 2008).

Nos últimos anos gañou popularidade a distribución da fraxilidade log-normal que ofrece maior flexibilidade e cuxa maior desvantaxe, a falta de transformada de Laplace que leva a rutinas de estimación máis complexas, non supón xa un grande obstáculo grazas ao incremento na capacidade computacional nos últimos anos (Wienke, 2010).

Á hora de optar por certa distribución é importante considerar que as distintas distribucións xeran distintos tipos de dependencia entre os membros dun grupo ou clúster (Duchateau & Janssen, 2008).

### 2.2.1. Fraxilidade gamma

O modelo de fraxilidade compartida gamma segue o modelo explicado anteriormente con fraxilidades  $Z$  que se distribúen segundo unha distribución gamma. A distribución gamma,  $\Gamma(k, \theta)$ , é unha distribución flexíbel que toma unha variedade de formas segundo o valor da variábel  $k$ . Con  $k = 1$  é a distribución exponencial, mentres que para  $k$  grande se achega a unha distribución con forma de campá semellante á distribución normal. Mentres que a distribución gamma está entre as máis utilizadas nas publicacións neste campo, non existe ningunha razón biolóxica que indique que sexa máis axeitada, senón que se escolle principalmente por conveniencia matemática (Wienke, 2010).

A densidade dunha variábel  $Z$  distribuída segundo unha gamma de parámetros  $k$  e  $\theta$  é

$$f(z) = \frac{1}{\Gamma(k)\theta^k} z^{k-1} e^{-z/\theta}.$$

A media dunha variábel con distribución gamma é  $\mathbb{E}[Z] = \theta k$ , mentres que a varianza é  $\text{Var}[Z] = \theta^2 k$ .

Para facer o modelo identificábel fíxase a media das fraxilidades a 1,  $\mathbb{E}[Z] = 1$ , (Wienke, 2010) de xeito que  $k = 1/\theta$  e  $\text{Var}[Z] = \theta$ .  $Z$  está polo tanto distribuída segundo unha  $\Gamma(1/\sigma^2, \sigma^2)$ , con  $\sigma^2 = \text{Var}[Z]$ . A densidade desta distribución é

$$f(z) = \frac{1}{\Gamma(\frac{1}{\sigma^2})\sigma^2/\sigma^2} z^{\frac{1}{\sigma^2}-1} \exp\left(-\frac{z}{\sigma^2}\right). \quad (2.10)$$

Usando a expresión (1.13) vemos que a transformada de Laplace de  $\Gamma(1/\sigma^2, \sigma^2)$  é

$$\mathcal{L}(u) = \frac{1}{(1 + \sigma^2 u)^{1/\sigma^2}}.$$

Polo tanto usando (2.7) temos que a función de supervivencia conxunta incondicional gamma é

$$S(\mathbf{t}_i | \mathbf{X}_i) = \frac{1}{(1 + \sigma^2 \Lambda_{\text{clúster}}(\mathbf{t}_i | \mathbf{X}_i))^{1/\sigma^2}}.$$

Unha vantaxe da distribución gamma para a estimación das fraxilidades, ademais de ter transformada de Laplace, é a expresión pechada das súas derivadas que dan Duchateau e Janssen (2008)

$$\mathcal{L}^{(k)}(u) = \frac{(-1)^k \prod_{l=0}^{k-1} [1 + l\sigma^2]}{(1 + \sigma^2 u)^{1/\sigma^2 + k}}$$

Usando esta relación e (2.8) temos que a verosimilitude incondicional para estimar un modelo coa forma de (2.2) é

$$L(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^n \left[ \prod_{j=1}^{n_i} \left[ \left( \lambda_0(T_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} \right)^{\delta_{ij}} \right] \frac{\prod_{l=0}^{D_i-1} [1 + l\sigma^2]}{(1 + \sigma^2 \Lambda_{\text{clúster}}(\mathbf{T}_i | \mathbf{X}_i))^{1/\sigma^2 + D_i}} \right] \quad (2.11)$$

Agora temos en (2.5, 2.10 e 2.11) as funcións para construír a densidade condicional das  $Z_i$  como vimos en (2.9), que poderemos usar para estimar as fraxilidades individuais. Como se recolle no apéndice C usando certas operacións alxebraicas chegamos a

$$f(z_i | \mathbf{T}_i, \boldsymbol{\delta}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\lambda}, \mathbf{X}_i) = \frac{z_i^{1/\hat{\sigma}^2 + D_i - 1} \exp(-z_i(1/\hat{\sigma}^2 + \sum_{j=1}^{n_i} \hat{\Lambda}_0(T_{ij}) e^{\hat{\boldsymbol{\beta}}^t \mathbf{X}_{ij}}))}{(1/\hat{\sigma}^2 + \sum_{j=1}^{n_i} \hat{\Lambda}_0(T_{ij}) e^{\hat{\boldsymbol{\beta}}^t \mathbf{X}_{ij}})^{-(1/\hat{\sigma}^2 + D_i)} \Gamma(1/\hat{\sigma}^2 + D_i)} \quad (2.12)$$

con  $\hat{\sigma}^2$  a estimación da varianza da fraxilidade.

A expresión (2.12) é a densidade dunha distribución gamma de parámetros  $k = 1/\hat{\sigma}^2 + D_i$  e  $\theta = 1/(1/\hat{\sigma}^2 + \sum_{j=1}^{n_i} \hat{\Lambda}_0(t_{ij} | X_{ij}))$ , polo tanto, dado que a media dunha distribución gamma é  $k\theta$ , podemos obter a predición da fraxilidade do clúster  $i$  usando:

$$\mathbb{E}[Z_i | \mathbf{T}_i, \boldsymbol{\delta}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\lambda}, \mathbf{X}_i] = \frac{1/\hat{\sigma}^2 + D_i}{1/\hat{\sigma}^2 + \sum_{j=1}^{n_i} \hat{\Lambda}_0(T_{ij}) e^{\hat{\boldsymbol{\beta}}^t \mathbf{X}_{ij}}}$$

A distribución gamma modeliza unha dependencia tardía como se pode apreciar gráficamente na figura 2.2 da sección 2.2.4.

### 2.2.2. Fraxilidade gaussiana inversa

A distribución gaussiana inversa de parámetros  $\mu$  e  $\alpha$ ,  $\text{IG}(\mu, \alpha)$  ten como función de densidade

$$f(z) = \sqrt{\frac{\alpha}{2\pi z^3}} \exp\left(-\frac{\alpha(z-\mu)^2}{2z\mu^2}\right)$$

con  $\mu, \alpha > 0$  (Duchateau & Janssen, 2008).

A distribución  $\text{IG}(\mu, \alpha)$  ten media  $\mathbb{E}[Z] = \mu$  e varianza  $\text{Var}[Z] = \mu^3/\alpha$ . Fixando  $\mathbb{E}[Z] = \mu = 1$  para facer a distribución identificábel, a distribución é  $\text{IG}(1, 1/\sigma^2)$ , con  $\sigma^2$  a varianza das  $Z$ , e a súa densidade é

$$f(z) = \sqrt{\frac{1}{2\pi\sigma^2 z^3}} \exp\left(-\frac{(z-1)^2}{\sigma^2 2z}\right). \quad (2.13)$$

A transformada de Laplace desta distribución é

$$\mathcal{L}(u) = \exp\left(\frac{1 - \sqrt{1 + 2\sigma^2 u}}{\sigma^2}\right). \quad (2.14)$$

Polo tanto, usando a expresión (2.7), tense que a función de supervivencia conxunta para un vector arbitrario de tempos para un clúster  $i$  vén dada por

$$S_{\text{clúster}}(\mathbf{t}|\mathbf{X}_i) = \exp\left(\frac{1 - \sqrt{1 + 2\sigma^2 \Lambda_{\text{clúster}}(\mathbf{t}|\mathbf{X}_i)}}{\sigma^2}\right).$$

De cara a obter a función de verosimilitude precisamos das  $k$  derivadas da transformada de Laplace. Desafortunadamente, neste caso non teñen unha forma pechada. Duchateau e Janssen (2008) indican que as 3 primeiras derivadas son:

$$\begin{aligned} \mathcal{L}'(u) &= -\exp\left(\frac{1 - \sqrt{1 + 2\sigma^2 u}}{\sigma^2}\right) \frac{1}{\sqrt{1 + 2\sigma^2 u}} \\ \mathcal{L}''(u) &= \exp\left(\frac{1 - \sqrt{1 + 2\sigma^2 u}}{\sigma^2}\right) \frac{1}{1 + 2\sigma^2 u} \left(1 + \frac{\sigma^2}{\sqrt{1 + 2\sigma^2 u}}\right) \\ \mathcal{L}^{(3)}(u) &= -\exp\left(\frac{1 - \sqrt{1 + 2\sigma^2 u}}{\sigma^2}\right) \frac{1}{\sqrt{(1 + 2\sigma^2 u)^3}} \left(1 + \frac{3\sigma^2}{\sqrt{1 + 2\sigma^2 u}} + \frac{3\sigma^4}{1 + 2\sigma^2 u}\right). \end{aligned}$$

A contribución do clúster  $i$  á función de log-verosimilitude incondicional é

$$\begin{aligned} \log(L(\boldsymbol{\beta}, \sigma^2)) &= \sum_{j=1}^{n_i} \left[ \delta_{ij} \log\left(\lambda_0(T_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}}\right) \right] + \frac{1 - \sqrt{1 + 2\sigma^2 \Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)}}{\sigma^2} \\ &\quad - \frac{D_i}{2} \log(1 + 2\sigma^2 \Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)) + C(D_i) \end{aligned} \quad (2.15)$$

con  $C(D_i)$  unha función que toma unha forma distinta segundo o número de eventos no clúster, que

toma as seguintes formas para  $D_i \leq 3$

$$\begin{aligned} C(0) &= C(1) = 0 \\ C(2) &= \log \left( 1 + \frac{\sigma^2}{\sqrt{1 + 2\sigma^2 \Lambda_{\text{clúster}}(T_i | \mathbf{X}_i)}} \right) \\ C(3) &= \log \left( 1 + \frac{3\sigma^2}{\sqrt{1 + 2\sigma^2 \Lambda_{\text{clúster}}(T_i | \mathbf{X}_i)}} + \frac{3\sigma^4}{1 + 2\sigma^2 \Lambda_{\text{clúster}}(T_i | \mathbf{X}_i)} \right) \\ &\dots \end{aligned}$$

En base ás estimacións dos parámetros de regresión,  $\hat{\beta}$ , da varianza das fraxilidades,  $\hat{\sigma}^2$ , e da función de risco basal  $\hat{\lambda}$ , as fraxilidades individuais poden predicirse usando as funcións de densidade da gaussiana inversa (2.13), e as funcións de densidade conxuntas condicional e incondicional ás  $Z$  dos tempos de evento usando a regra de Bayes como se viu en de (2.9).

A media de  $Z_i$  con esta función de densidade é a a predición da fraxilidade do clúster  $i$ ,  $\hat{Z}_i$ . A diferenza do caso gamma, esta densidade non leva a unha expresión pechada das predicións das fraxilidades. A distribución gaussiana inversa modeliza unha fraxilidade máis temperá que a distribución gamma. Esta dependencia pódese observar graficamente na figura 2.2, incluída na sección 2.2.4.

### 2.2.3. Fraxilidade positiva estábel

Unha distribución considérase estábel se ten a seguinte propiedade: tendo  $Y_1, \dots, Y_n$  variábeis i.i.d., para cada  $n$  existe unha constante normalizadora  $\mathcal{C}(n)$  tal que  $F(\sum_{i=1}^n Y_i) = F(\mathcal{C}(n)Y_1)$ , con función de distribución  $F(y)$ . A constante  $\mathcal{C}(n)$  toma a forma  $n^{1/\theta}$  con  $\theta \in (0, 2]$ . As distribucións estábeis na semirecta positiva teñen  $\theta \in (0, 1]$

Unha variábel  $Z$  con distribución positiva estábel de parámetros  $\theta \in (0, 1]$  e  $k$ ,  $PS(\theta, k)$ , ten densidade

$$f(z) = -\frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{\Gamma(k\theta + 1)}{k!} \left(-\frac{1}{z^\theta}\right)^k \text{sen}(\theta k \pi). \quad (2.16)$$

A distribución positiva estábel caracterízase por ter media infinita e, consecuentemente, varianza indeterminada.

A transformación de Laplace ten a forma sinxela

$$\mathcal{L}(u) = \exp(-u^\theta), \quad (2.17)$$

o que leva á seguinte función de supervivencia conxunta incondicional

$$S_{\text{clúster}}(\mathbf{t} | \mathbf{X}_i) = \exp\left(-\Lambda_{\text{clúster}}(\mathbf{t} | \mathbf{X}_i)^\theta\right).$$

Por outra parte, a función de risco incondicional para un individuo é

$$\lambda(\mathbf{t} | \mathbf{X}_{ij}, \theta) = \frac{d}{dt} \left( -\log(S(\mathbf{t} | \mathbf{X}_{ij})) \right) = \theta \lambda_0(t) \Lambda_0^{\theta-1}(t) e^{\beta^t \mathbf{X}_{ij} \theta}. \quad (2.18)$$

Podemos ver que para dous individuos con covariábeis  $\mathbf{X} = \mathbf{x}$  e  $\mathbf{X} = \mathbf{x}^*$  o ratio das funcións de risco con (2.18) é

$$\frac{\theta \lambda_0(t) \Lambda_0^{\theta-1}(t) e^{\beta^t \mathbf{x} \theta}}{\theta \lambda_0(t) \Lambda_0^{\theta-1}(t) e^{\beta^t \mathbf{x}^* \theta}} = \exp(\theta \beta^t (\mathbf{x} - \mathbf{x}^*))$$

que é un valor constante no tempo. A distribución positiva estábel para as fraxilidades é a única que mantén a proporcionalidade dos riscos, aínda que con  $e^{\theta\beta^t(x-x^*)}$  en lugar de  $e^{\beta^t(x-x^*)}$ .

Do mesmo xeito que coa distribución gaussiana inversa, as derivadas da transformada de Laplace non teñen unha expresión pechada. Duchateau e Janssen (2008) dan as 3 primeiras derivadas:

$$\begin{aligned}\mathcal{L}'(u) &= -\theta u^{\theta-1} \exp(-u^\theta) \\ \mathcal{L}''(u) &= \theta^2 u^{2(\theta-1)} \exp(-u^\theta) \left(1 + \frac{1-\theta}{\theta u^\theta}\right) \\ \mathcal{L}^{(3)}(u) &= -\theta^3 u^{3(\theta-1)} \exp(-u^\theta) \left(1 + \frac{3(1-\theta)}{\theta u^\theta} + \frac{(1-\theta)(2-\theta)}{\theta^2 u^{2\theta}}\right)\end{aligned}$$

Polo tanto, a contribución do clúster  $i$  á log-verosimilitude incondicional pode presentarse como:

$$\begin{aligned}\log(L_i(\boldsymbol{\beta}, \theta)) &= \sum_{j=1}^{n_i} \left[ \delta_{ij} (\lambda(T_{ij} | \mathbf{X}_{ij})) \right] - \Lambda_{\text{clúster}}(\mathbf{T}_i | \mathbf{X}_i)^\theta + D_i \log(\theta) \\ &\quad + D_i(\theta - 1) \log(\Lambda_{\text{clúster}}(\mathbf{T}_i | \mathbf{X}_i)) + C(D_i)\end{aligned}\tag{2.19}$$

con  $C(D_i)$  un termo que depende do número de eventos no clúster que toma as seguintes formas para  $D_i \leq 3$

$$\begin{aligned}C(0) &= C(1) = 0 \\ C(2) &= \log\left(1 + \frac{1-\theta}{\theta \Lambda_{\text{clúster}}(\mathbf{T}_i | \mathbf{X}_i)^\theta}\right) \\ C(3) &= \log\left(1 + \frac{3(1-\theta)}{\theta \Lambda_{\text{clúster}}(\mathbf{T}_i | \mathbf{X}_i)^\theta} + \frac{(1-\theta)(2-\theta)}{(\theta \Lambda_{\text{clúster}}(\mathbf{T}_i | \mathbf{X}_i)^\theta)^2}\right) \\ &\dots\end{aligned}$$

Do mesmo xeito que no caso da distribución gaussiana inversa, a densidade das fraxilidades  $Z$  condicionada aos datos non ten unha forma sinxela.

A distribución gaussiana inversa modelizaba unha dependencia previa no tempo á da distribución gamma. A distribución positiva estábel á súa vez modeliza unha dependencia previa á da gaussiana inversa. Vese na figura 2.2 da sección 2.2.4.

#### 2.2.4. Fraxilidade potencia-varianza

A familia de distribucións potencia-varianza, tamén coñecida como de Tweedie polo seu descubridor, é unha familia de distribucións que ten como casos especiais á distribución gamma, gaussiana inversa e positiva estábel.

Esta distribución está caracterizada por 3 parámetros  $\theta$ ,  $\mu$  e  $\nu$ , e ten densidade

$$\begin{aligned}f(z) &= \exp\left(-\frac{\nu}{\theta} \left(\frac{z}{\mu} + \frac{1}{\nu-1}\right)\right) \\ &\quad \times \frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{(\nu/\theta)^{k\nu} (z/\mu)^{k(\nu-1)} \Gamma(1-k(\nu-1)) \text{sen}(\pi k(\nu-1))}{k!(\nu-1)^k}\end{aligned}\tag{2.20}$$

con  $0 < \mu, \theta$ , e  $0 < \nu \leq 1$ .

A transformación de Laplace desta distribución é

$$\mathcal{L}(u) = \exp\left(\frac{\nu}{\theta(1-\nu)} \left(1 - \left(1 + \frac{\theta\mu u}{\nu}\right)^{1-\nu}\right)\right) \quad (2.21)$$

Esta distribución ten media  $\mathbb{E}[Z] = \mu$  e varianza  $\text{Var}[Z] = \theta\mu^2$  cando estas existen (por exemplo non existen para a distribución positiva estábel). Polo xeral fixaremos  $\mu = 1$  de xeito que  $\theta$  sexa á varianza da fraxilidade.

En base a isto a función de supervivencia conxunta incondicional para un vector arbitrario de tempos  $\mathbf{t}$  sería

$$S_{\text{clúster}}(\mathbf{t}|\mathbf{x}_i) = \exp\left(\frac{\nu}{\theta(1-\nu)} \left(1 - \left(1 + \frac{\theta\Lambda_{\text{clúster}}(\mathbf{t}, |\mathbf{x}_i)}{\nu}\right)^{1-\nu}\right)\right)$$

De cara a calcular a contribución dun clúster á (log-)verosimilitude incondicional precisamos das derivadas desta función como se viu anteriormente. As derivadas non teñen unha expresión pechada. Duchateau e Janssen (2008) dan as 3 primeiras derivadas

$$\begin{aligned} \mathcal{L}'(u) &= -\frac{\mathcal{L}(u)}{(1 + \theta u/\nu)^\nu} \\ \mathcal{L}''(u) &= \mathcal{L}(u) \left( \frac{1}{(1 + \theta u/\nu)^{2\nu}} + \frac{\theta}{(1 + \theta u/\nu)^{\nu+1}} \right) \\ \mathcal{L}^{(3)}(u) &= -\mathcal{L}(u) \left( \frac{1}{(1 + \theta u/\nu)^{3\nu}} + \frac{3\theta}{(1 + \theta u/\nu)^{2\nu+1}} + \frac{\theta^2(1 + 1/\nu)}{(1 + \theta u/\nu)^{\nu+2}} \right) \end{aligned}$$

polo tanto, a contribución dun clúster á log-verosimilitude é

$$\begin{aligned} \log(L_i(\boldsymbol{\beta}, \theta, \nu)) &= \sum_{j=1}^{n_i} \delta_{ij} \left( \log(\lambda_0((T_{ij})e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}})) \right) + \left( \frac{\nu}{\theta(1-\nu)} \left(1 - \left(1 + \frac{\theta\Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)}{\nu}\right)^{1-\nu}\right) \right) \\ &\quad - D_i \nu \log\left(\frac{1 + \theta\Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)}{\nu}\right) + C(D_i) \end{aligned} \quad (2.22)$$

con  $C(D_i)$  un termo que depende do número de eventos no clúster. Para os clusters que teñen de 0 a 3 eventos  $C(D_i)$  é

$$\begin{aligned} C(0) &= C(1) = 0 \\ C(2) &= \log\left(1 + \theta \left(1 + \frac{\theta\Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)}{\nu}\right)^{\nu-1}\right) \\ C(3) &= \log\left(1 + 3\theta \left(1 + \frac{\theta\Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)}{\nu}\right)^{\nu-1} + \sigma^4(1 + 1/\nu) \left(1 + \frac{\theta\Lambda_{\text{clúster}}(\mathbf{T}_i|\mathbf{X}_i)}{\nu}\right)^{2\nu-2}\right) \end{aligned}$$

Pódense obter estimacións para as fraxilidades individuais do mesmo xeito que para as distribucións gaussiana inversa e positiva estábel. Polo xeral a función de densidade utilizada na estimación non ten unha expresión pechada.



### Relación da distribución potencia-varianza coas distribucións gamma, gaussiana inversa e positiva estábel

Vexamos que, como se comentou, as distribucións gamma, gaussiana inversa e positiva estábel son casos especiais da distribución potencia-varianza.

Para  $\nu = 0.5$  a transformada de Laplace correspóndese coa da gaussiana inversa, onde o parámetro  $\alpha$  da gaussiana inversa toma o valor  $\alpha = 1/\theta$ . Visto doutro xeito,  $\theta$  sería a varianza das fraxilidades,  $\theta = \sigma^2$ . Pódese ver que a transformada laplaciana en (2.21) con estes valores e  $\mu = 1$

$$\mathcal{L}(u|\mu = 1, \theta = \sigma^2, \nu = 0.5) = \exp\left(\frac{0.5}{\sigma^2(1-0.5)}\left(1 - \left(1 + \frac{\sigma^2 u}{0.5}\right)^{1-0.5}\right)\right) = \exp\left(\frac{1 - \sqrt{1 + 2\sigma^2 u}}{\sigma^2}\right)$$

que é a transformada laplaciana da gaussiana inversa en (2.14).

Respecto da distribución positiva estábel, chegamos a ela se facemos que  $\mu$  e  $\mu^2\theta$  da distribución P-V tendan ao infinito de forma axeitada. Para obter a transformación de Laplace da distribución positiva estábel tomamos

$$1 - \nu = \frac{\mu^{1-\nu}\nu^\nu}{\theta^\nu}$$

ou, equivalentemente

$$\theta = \frac{\nu}{(1-\nu)^{1-\nu}}\mu^{1/\nu-1}$$

Dado que  $0 < \nu < 1$ ,  $\theta \rightarrow \infty$ , e  $\mu \rightarrow \infty$ . Ademais,

$$\frac{\nu}{\mu\theta} \rightarrow 0$$

A transformada de Laplace en (2.21) con  $\mu = 1$  pode reescribirse como

$$\mathcal{L}(u) = \exp\left(\left(\frac{\nu}{\theta}\right)^{1-\nu} - \left(\frac{\nu}{\theta} + u\right)^{1-\nu}\right)$$

e polo tanto

$$\lim_{\theta \rightarrow \infty} \mathcal{L}(u) = \exp(-u^{1-\nu})$$

que é a transformación de Laplace en (2.17) con  $\theta = 1 - \nu$ .

Respecto da transformación de gamma, pódese probar usando a regra de l'Hôpital (como fan Duchateau e Janssen (2008)) que cando  $\nu$  tende a 1, o logaritmo da transformada de Laplace da distribución P-V tende ao logaritmo da transformada de Laplace da gamma

$$\lim_{\nu \rightarrow 1} \frac{\nu}{\theta(1-\nu)} \left(1 - \left(1 + \frac{\theta u}{\nu}\right)^{1-\nu}\right) = -\frac{1}{\theta} \log(1 + \theta u).$$

En xeral, o valor de  $\nu$  na distribución potencia-varianza indica o tipo de dependencia. Valores maiores de  $\nu$  axustan unha dependencia máis tardía, e viceversa.

Na figura 2.2 pódese ver graficamente a relación entre a dependencia e a distribución escollida. A figura amosa o ratio entre a función de supervivencia conxunta para dous individuos e o produto das dúas funcións de supervivencia, sen considerar a dependencia para distintas distribucións, mantendo constante o valor da dependencia, medido usando a  $\tau$  de Kendall. Duchateau e Janssen (2008) achegan a función para obter esta medida de dependencia para cada unha das distribucións. Neste caso úsase  $\tau = 0.25$ .

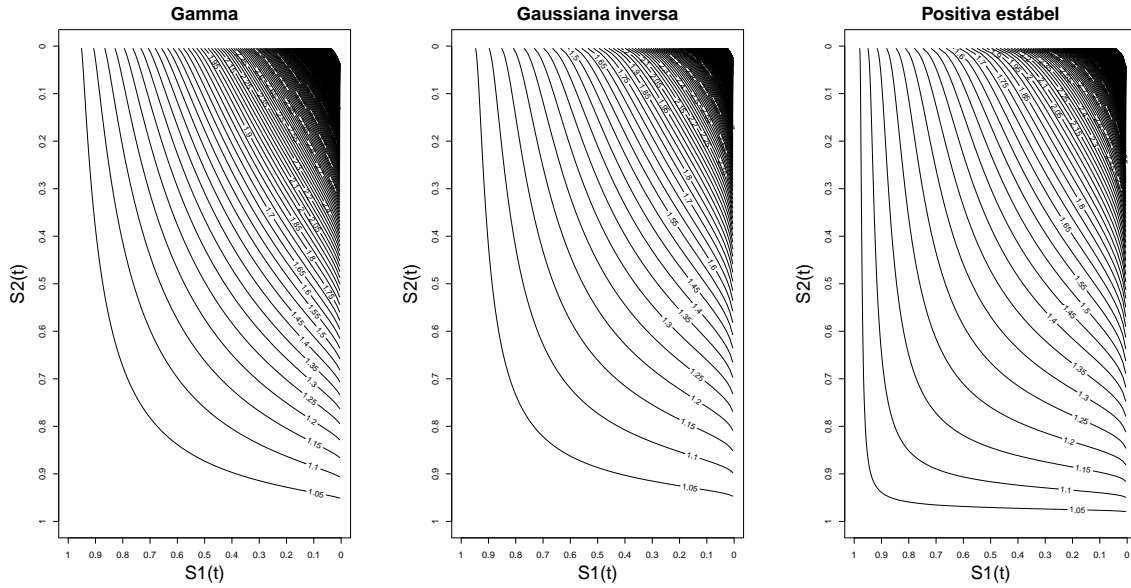


Figura 2.2: Ratio da función de supervivencia conxunta e do produto das supervivencias individuais para dous individuos con fraxilidade con nivel de dependencia  $\tau = 0.25$  para distintas distribucións. Adaptación de Duchateau e Janssen (2008)

O valor 1, que é o valor mínimo para este ratio, indica independencia mentres que valores máis altos amosan unha dependencia crecente. Podemos ver claramente que é para a distribución gamma para a que o ratio máis tarda en desviarse significativamente do 1, mentres que o fai bastante cedo no caso da distribución positiva estábel.

### 2.2.5. Fraxilidade log-normal

A distribución log-normal permite utilizar metodoloxía para axustar modelos de fraxilidade que é similar á teoría clásica dos modelos mixtos lineares. O modelo proposto é

$$\lambda(t|\mathbf{X}, W) = \lambda_0(t)e^{\beta^t \mathbf{X} + W}, \quad (2.23)$$

onde  $W = \log(Z)$  é o efecto aleatorio que segue unha distribución normal con varianza  $\gamma > 0$ .

Para tódalas distribucións anteriores supúxose que  $\mathbb{E}[Z] = 1$ , porén para a distribución log-normal, é común supór que os efectos aleatorios  $W = \log(Z)$  teñen media 0. Polo tanto, a media aritmética das fraxilidades non é normalmente 1.

É doado amosar que

$$f(z) = \frac{\exp\left(-\frac{(\log(z))^2}{2\gamma}\right)}{z\gamma\sqrt{2\pi}}.$$

A media e varianza dos efectos aleatorios son

$$\begin{aligned} \mathbb{E}[W] &= \exp(\gamma/2) \\ \text{Var}[Z] &= \exp(2\gamma) - \exp(\gamma). \end{aligned}$$

A distribución log-normal non ten transformada laplaciana, co cal non se pode extraer  $Z$  da supervivencia incondicional e no seu lugar deberán usarse métodos de integración numérica ou outros métodos de cara á estimación dos parámetros do modelo.

## 2.3. Aplicación a datos

Para exemplificar o uso dos modelos de fragilidade, aplícanse aos datos sobre o tempo ata a perda da visión en cada un dos ollos de pacientes con retinopatía dispoñíbeis no conxunto de datos `retinopathy` do paquete `survival`. Como se mencionou no inicio deste capítulo son datos que motivan a utilización dos modelos de fragilidade, creando o clúster ao nivel do individuo, de xeito que se considere a propensión de cada individuo a perder a visión por esta doenza.

A selección das variábeis a incluír no modelo fíxose axustando tódolos modelos posibles sen termos polinómicos nin interaccións e escollendo aquel co maior AIC. Esta selección fíxose usando a función `glmulti` do paquete homónimo. As variábeis incluídas finalmente son `trt`, que indica se o ollo foi tratado e `risk` que é unha medida médica do risco de perder a visión no ollo, que na mostra toma valores entre 6 e 12, e para a que valores máis altos indican maior risco.

En primeiro lugar axústase o modelo sen considerar a fragilidade, utilizando a función `coxph` do paquete `survival`, a modo de comparación co resultado ao considerar a fragilidade. A continuación, aínda que `survival` dispón dunha función para engadir fragilidade a un modelo, como se explica na sección 3.3.1, usouse a función `emfrail` do paquete `frailtyEM` que contén ferramentas adicionais como o test de heteroxeneidade de Commenges-Andersen e o gráfico da relación entre o ratio de riscos condicional e marxinal. Este paquete está descrito na sección 3.2.1.

```
cox <- coxph(Surv(futime, status)~trt + risk, data = ret)
cox

## Call:
## fitfunc(formula = as.formula(x), data = data)
##
##           coef  exp(coef)  se(coef)  z      p
## trt  -0.77792   0.45936   0.16881  -4.608 4.06e-06
## risk  0.14600   1.15720   0.05588  2.613 0.00898
##
## Likelihood ratio test=29.33 on 2 df, p=4.282e-07
## n= 394, number of events= 155

library(frailtyEM)
cox.frail <- emfrail(Surv(futime, status)~trt + risk + cluster(id), data = ret)
cox.frail

## Call:
## emfrail(formula = Surv(futime, status) ~ trt + risk + cluster(id),
##          data = ret)
##
## log-likelihood: -848.0778
## frailty variance: 0.7895216
## theta: 1.26659
##
##           coef  exp(coef)  se(coef)  adj. se  z      p
## trt  -0.903330  0.405218   0.174770  0.180410 -5.007098 0.0000
## risk  0.166907  1.181644   0.069049  0.069192  2.412246 0.0159
##
## Score test for heterogeneity: p-val 0.000329
```

Respecto da estimación dos regresores, vemos que ao considerar a fragilidade aumenta o tamaño do efecto de cada regresor, cun decrecemento do risco do 60% grazas ao tratamento entre ollos coa mesma fragilidade, en lugar de 55% sen considerar a fragilidade, e un crecemento da función de risco de 18% por cada aumento do nivel de risco, `risk`, para a mesma fragilidade, 2% máis que o 16% do aumento do risco por nivel ao non considerar fragilidade. Vemos que, cando non se considera a fragilidade, se infraestima a variación de risco achegada polo tratamento ou por un menor nivel de risco.

Podemos ver que o test de heteroxeneidade de Commenges-Andersen (Score test for heterogeneity) resulta claramente significativo, o que evidencia a necesidade de utilizar un modelo de fragilidade.

A estimación dos parámetros tamén permite ver como a fragilidade afecta á estimación da función de risco, xa que a composición da poboación estudada varía a medida que os individuos con maior

fraxilidade experimentan o evento e abandonan o estudo. O paquete **frailtyEM** permite xerar a figura 2.3, que compara o ratio de riscos condicional ás  $Z$ , e o ratio incondicional ou marxinal estimados usando o seguinte código:

```
autoplot(cox.frail, type = "hr", lty = 2,
         newdata = data.frame(trt = c(1, 0), risk = rep(mean(ret$risk), 2)))
```

É doado ver que o *hazard ratio* condicionado ás  $Z$  respecta a proporcionalidade suposta no modelo de Cox, pero que o aumento do risco marxinal crece co tempo, a medida que os individuos máis fráxiles abandonan o estudo. É dicir, a medida que se deixa de considerar a aqueles individuos máis susceptíbeis de perder a visión, a diferenza do tratamento magnifícase.

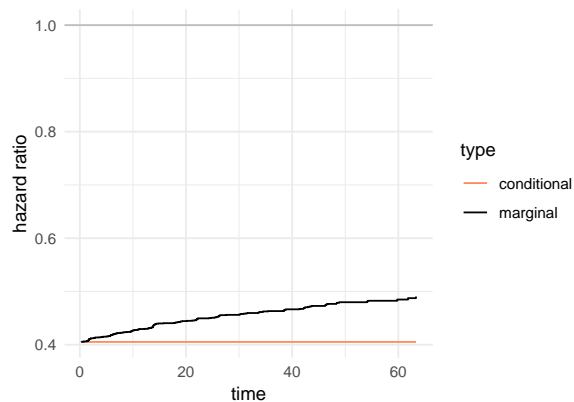


Figura 2.3: *Hazard ratio* condicional e incondicional entre un ollo tratado e un ollo non tratado

## Capítulo 3

# Métodos de estimación de modelos de fraxilidade

### 3.1. Máxima verosimilitude paramétrica

A maximización da (log-)verosimilitude, calculada a partir da expresión (2.8) e que se amosou para as distintas distribucións de fraxilidade, respecto dos parámetros de interese é o método de estimación evidente á hora de axustar modelos completamente parametrizados como o modelo de Cox cun risco basal paramétrico visto na sección 1.2.1.

#### 3.1.1. Implementación en R

##### **parfm**

Os modelos de Cox paramétricos poden axustarse usando a función **parfm** do paquete **parfm** desenvolto por Munda, Rotolo e Legrand (2012) usando a seguinte sintaxe

```
parfm(Surv(time, status)~cov1 + cov2 + ..., cluster = "frailty.var", data = dat, dist = "weibull",  
      frailty = "gamma")
```

onde **time** é o tempo observado, é dicir o mínimo entre o tempo de evento e o de censura, **status** é unha variábel binaria que indica se se observou o evento de interese, **cov1**, **cov2**, ... son as covariábeis das que dispoñemos para cada individuo, **frailty.var** é a variábel de fraxilidade e **dat** contén os datos.

Na variábel **dist** indícase a distribución dos tempos de evento, algunhas opcións son exponencial (**exponential**), weibull (**weibull**), weibull inversa (**invweibull**), log-normal (**lognormal**) e log-loxística (**loglogistic**). Na variábel **frailty** indícase a distribución das fraxilidades que pode ser gamma (**gamma**), gaussiana inversa (**ingau**), positiva estábel (**possta**) e log-normal (**lognormal**). No caso da fraxilidade con distribución log-normal, dada a ausencia de transformada de Laplace, utilízase a aproximación *saddlepoint* (para máis detalles, consúltese Rotolo et al. (2023)).

Este paquete engade ademais unha función chamada **select.parfm** que permite axustar varias combinacións de modelos con distintas distribucións para os tempos de evento e/ou as fraxilidades. A función devolve o AIC e o criterio de información de Bayes (BIC) de cada axuste.

### 3.2. Algoritmo de esperanza-maximización

Á hora de estimar modelos semiparamétricos nos que o risco basal non se especifica, que como se indicou son os máis populares en análise de supervivencia, precísase doutros métodos. O algoritmo de esperanza-maximización (EM) é unha opción para a estimación destes modelos, que recollen tanto Duchateau e Janssen (2008) como Wienke (2010).

Este algoritmo itera entre dous pasos. Nun primeiro paso, chamado de esperanza, obtéñense os valores esperados das fraxilidades. No segundo paso, chamado de maximización, considérase que esta é a información observada e obtéñense estimacións dos parámetros de regresión e da función de distribución das fraxilidades.

Comecemos co logaritmo da verosimilitude condicionada ás  $\mathbf{Z}$ , vista en (2.5), que é

$$\log(L(\boldsymbol{\beta}|\mathbf{Z})) = \sum_{i=1}^n \left[ \sum_{j=1}^{n_i} \left[ \delta_{ij} \left( \log(\lambda_0(T_{ij})) + \log \left( e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} Z_i \right) \right) - \Lambda_0(T_{ij}) e^{\boldsymbol{\beta}^t \mathbf{X}_{ij}} Z_i \right] \right]. \quad (3.1)$$

No caso de coñecer os valores das  $\mathbf{Z}$ , poderíamos obter estimacións das  $\boldsymbol{\beta}$  reescribindo  $\exp(\boldsymbol{\beta}^t \mathbf{X}_{ij}) Z_i$  como  $\exp(\boldsymbol{\beta}^t \mathbf{X}_{ij} + \log(Z_i))$  e maximizando respecto de  $\boldsymbol{\beta}$  unha versión parcial desta verosimilitude. Tamén sería sinxelo obter estimacións de  $\boldsymbol{\theta}$  maximizando a función de densidade das fraxilidades respecto destes parámetros.

Porén, como non coñecemos as  $\mathbf{Z}$ , teremos que usar unha predición das mesmas. Para obtelas úsase a expresión (2.9), que depende das estimacións dos parámetros de interese.

Como é obvio que non se poden obter simultaneamente as predicións das  $\mathbf{Z}$  que dependen das estimacións dos parámetros e as estimacións dos parámetros que dependen das predicións, teremos que iterar entre os dous pasos que dan nome ao método.

#### Maximización

Consideremos en primeiro lugar o paso da maximización.

Estendendo a idea da función verosimilitude parcial no modelo de Cox en (1.8), a función de verosimilitude en (2.5) podería adaptarse para conseguir:

$$L(\boldsymbol{\beta}|\mathbf{Z}) = \prod_{i=1}^n \prod_{j=1}^{n_i} \left( \frac{\exp(\boldsymbol{\beta}^t \mathbf{X}_{ij} + \log(Z_i))}{\sum_{qr \in R(t_{ij})} Z_q \exp(\boldsymbol{\beta}^t \mathbf{X}_{qr})} \right)^{\delta_{ij}}.$$

Tomando o logaritmo desta expresión e substituíndo as  $\mathbf{Z}$  e os  $\log(\mathbf{Z})$  polo seu valor esperado (dadas as estimacións dos parámetros máis recentes),  $\mathbb{E}[\mathbf{Z}]$  e  $\mathbb{E}[\log(\mathbf{Z})]$ , chegamos a

$$\log(L(\boldsymbol{\beta}|\mathbf{Z})) = \sum_{i=1}^n \left[ \sum_{j=1}^{n_i} \left[ \delta_{ij} \left( \boldsymbol{\beta}^t \mathbf{X}_{ij} + \mathbb{E}[\log(Z_i)] - \log \left( \sum_{qw \in R(t_{ij})} e^{\boldsymbol{\beta}^t \mathbf{X}_{qw}} \mathbb{E}[Z_q] \right) \right) \right] \right],$$

que pode ser maximizado para obter as estimacións dos parámetros cos métodos habituais.

Por outra parte maximizamos o logaritmo da función de densidade das  $\mathbf{Z}$ ,  $\log(f(\mathbf{Z}|\boldsymbol{\theta}))$ , substituíndo de novo as  $\mathbf{Z}$  polas súas predicións, para obter estimacións dos parámetros da distribución das fraxilidades  $\boldsymbol{\theta}$ .

### Esperanza

No paso da esperanza obteremos predicións das fraxilidades usando a expresión (2.9) dadas as estimacións dos parámetros de regresión e as estimacións dos parámetros da distribución das fraxilidades.

De cara a utilizar esta expresión son precisas estimacións do risco basal e do risco basal acumulado. Obtemos estas estimacións usando o estimador recollido en (1.10), coas estimacións máis recentes dos parámetros de interese  $\hat{\beta}$  e as predicións anteriores das fraxilidades.

$$\hat{\lambda}_0(t_{(k)}) = \frac{D(t_{(k)})}{\sum_{qw \in R(t_{(k)})} \mathbb{E}[Z_q] \exp(\beta^t \mathbf{X}_{qw})},$$

$$\hat{\Lambda}_0(t) = \sum_{t_{(k)} \leq t} \hat{\lambda}_0(t_{(k)}).$$

con  $D(t_{(k)})$  o número de eventos en  $t_{(k)}$ ,  $k = 1, \dots, K$  e neste caso  $K$  o número de tempos observados, non necesariamente de evento, diferentes  $K \leq \sum_{i=1}^n n_i$

### Algoritmo

Iterando entre estes dous pasos obtéñense as estimacións dos parámetros.

Como estimacións iniciais dos parámetros de regresión, Duchateau e Janssen (2008) propoñen usar o resultado dun axuste de Cox sen fraxilidade, mentres que as dos parámetros da distribución de fraxilidade se fixarían a 1. Despois, comezando polo paso da esperanza, propoñen facer un mínimo de tres iteracións, e a partir daí comprobar a converxencia vendo se a diferenza absoluta entre a verosimilitude incondicional (ou marxinal) na última iteración e a anterior é menor a un valor predeterminado, e deter o algoritmo cando se cumpra esta condición.

#### 3.2.1. Implementación en R

##### frailtyEM

Balan e Putter (2019) desenvolveron o paquete **frailtyEM** que utiliza unha versión algo máis complexa e eficiente desta rutina de estimación. Esta rutina consta de dous bucles, no bucle interior utilízanse os pasos E e M para a obtención das estimacións das fraxilidades dado certo valor da varianza das fraxilidades e en base a estas estímense os parámetros da distribución. No bucle exterior maximízase a verosimilitude da varianza das fraxilidades para os valores dos parámetros fixos estimados ata entón. Itérase entre estes dous bucles ata obter converxencia. Esta rutina pode utilizarse coa función `emfrail`, usando a seguinte sintaxe

```
emfrail(Surv(time, status) ~ cov1 + cov2 + ... + cluster(frailty.var), data = dat,
        distribution = emfrail_dist(dist = "gamma"))
```

Neste caso `distribution` indica a distribución das fraxilidades a través da función `emfrail_dist`. A distribución das fraxilidades pode ser gamma (`gamma`), positiva estábel (`stable`) ou potencia-varianza (`pvf`). No caso de usar a distribución potencia-varianza, `pvf`, debe indicarse usando `pvfm` o valor de  $\nu$ .

### Outros paquetes

Unha opción alternativa viña dada por o paquete **phmm**, pero foi retirado do repositorio CRAN en 2020 (Comprehensive R Archive Network, 2020).

### 3.3. Verosimilitude parcial penalizada

A verosimilitude parcial penalizada leva ás mesmas estimacións que o algoritmo de esperanza-maximización cando se usa, por exemplo, para estimar fraxilidades con distribución gamma. Pero ademais permite estimar modelos de fraxilidade cando a distribución subxacente é log-normal.

Neste apartado resúmese a explicación deste método de Duchateau e Janssen (2008), utilizando algunha notación de McGilchrist (1993), por ser máis detallada.

A log-verosimilitude consta de dúas partes, a primeira é a verosimilitude condicional dadas as fraxilidades, a segunda corresponde á distribución das fraxilidades. Esta expresión pode ser utilizada para a estimación da verosimilitude parcial penalizada comezando pola representación alternativa do modelo de fraxilidade de Cox

$$\lambda(t|\mathbf{X}_{ij}, W_i) = \lambda_0(t) \exp(\boldsymbol{\beta}^t \mathbf{X}_{ij} + W_i)$$

con  $W_i = \log(Z_i)$ . Neste contexto chámase ás  $\mathbf{W}$  efectos aleatorios e ás  $\mathbf{Z}$  fraxilidades, e denótase  $\gamma$  a varianza dos efectos aleatorios,  $\mathbf{W}$ , e  $\sigma^2$  á das fraxilidades.

Este método utiliza un enfoque similar ao anterior, usando a verosimilitude parcial vista anteriormente e considerando á densidade das fraxilidades como un termo de penalización, o que leva a:

$$L_{\text{ppl}}(\gamma, \boldsymbol{\beta}|\mathbf{w}) = L_{\text{part}}(\boldsymbol{\beta}, \mathbf{w}) - L_{\text{pen}}(\gamma, \mathbf{w})$$

con

$$L_{\text{part}}(\boldsymbol{\beta}, \mathbf{w}) = \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij} \left( \boldsymbol{\beta}^t \mathbf{X}_{ij} + w_i - \log \left( \sum_{qr \in R(t_{ij})} \exp(\boldsymbol{\beta}^t \mathbf{X}_{qr} + w_q) \right) \right)$$

e

$$L_{\text{pen}}(\gamma) = - \sum_{i=1}^n \log f(w_i|\gamma).$$

No bucle interior maximízase  $L_{\text{ppl}}(\gamma, \boldsymbol{\beta}|\mathbf{w})$  en  $\boldsymbol{\beta}$  e  $\mathbf{w}$ , para un valor provisional de  $\gamma$ , usando un procedemento Newton-Raphson, para obter estimacións de  $\boldsymbol{\beta}$  e  $\mathbf{w}$ . No bucle exterior obtense un estimador de máxima verosimilitude residual de  $\gamma$  usando as estimacións de  $\boldsymbol{\beta}$  e  $\mathbf{w}$ . O proceso itérase ata acadar converxencia.

Usemos  $l$  para denotar o número de iteracións no bucle exterior e  $k$  ás do bucle interior, de xeito que  $\hat{\boldsymbol{\beta}}_{(l,k)}$  sexan os estimadores de  $\boldsymbol{\beta}$  na  $l$ -ésima iteración do bucle exterior e a  $k$ -ésima do bucle interior.

Comezando no bucle interior, as iteracións do proceso de Newton-Raphson veñen dadas por

$$\begin{bmatrix} \boldsymbol{\beta}_{(l,k)} \\ \mathbf{W}_{(l,k)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_{(l,k-1)} \\ \mathbf{W}_{(l,k-1)} \end{bmatrix} - \mathbf{V}^{-1} \begin{bmatrix} 0 \\ \gamma_{(l)}^{-1} \mathbf{w}_{(l,k-1)} \end{bmatrix} + \mathbf{V}^{-1} \begin{bmatrix} \mathbf{X}(\mathbf{T}, \boldsymbol{\delta}) \\ \frac{dL_{\text{part}}(\boldsymbol{\beta}, \mathbf{W})}{d\mathbf{W}} \end{bmatrix},$$

con

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^t \\ (\mathbf{T}, \boldsymbol{\delta})^t \end{bmatrix} \left( \frac{-d^2 L_{\text{part}}(\boldsymbol{\beta}, \mathbf{W})}{d\mathbf{W} d\mathbf{W}^t} \right) \begin{bmatrix} \mathbf{X} & (\mathbf{T}, \boldsymbol{\delta}) \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \gamma_{(l)}^{-1} \mathbf{I} \end{bmatrix}$$

con  $\mathbf{I}$  a matriz identidade.

Cando o proceso de Newton-Raphson converxa para  $\gamma_{(l)}$ , pasamos á seguinte iteración do bucle exterior, que achega un novo estimador de máxima verosimilitude restrinxida para  $\gamma$  usando

$$\gamma_{(l+1)} = \sum_{i=1}^n \frac{(W_{i,(l,k)})^2}{n - \text{traza}(\mathbf{V}_{22})/\gamma_{(l)}}.$$



Este bucle exterior repítase ata que a diferenza absoluta entre dúas estimacións consecutivas de  $\gamma$ ,  $|\gamma_{(l)} - \gamma_{(l-1)}|$  sexa menor a un valor predeterminado.

### 3.3.1. Implementación en R

#### survival

A verosimilitude parcial penalizada é claramente a rutina máis popular para a estimación destes modelos e, consecuentemente, R contén unha variedade de paquetes que implementan esta rutina. O paquete **survival** permite axustar un modelo de fragilidade semiparamétrico gamma ou log-normal engadindo un termo **frailty** á función **coxph**.

```
coxph(Surv(time, status) ~ cov1 + cov2 + ... + frailty(frailty.var, distribution = "gamma", method = "em"), data = dat)
```

**distribution** indica a distribución das fragilidades que ademais de gamma pode ser log-normal (**gaussian**) ou  $t$  de Student (**t**). A variábel **method** indica o método usado para o axuste. **em** é específico á distribución gamma e **reml** é o método equivalente para a distribución log-normal. Usando un destes dous métodos coa distribución correspondente, axústase o logaritmo anterior. Ademais destes dous métodos, as tres distribucións permiten usar **fixed**, que permite ao usuario indicar o valor da varianza das fragilidades, **df**, co que o que se indicaría é o número de graos de liberdade a utilizar, ou **aic** que maximiza o AIC (Therneau & Lumley, 2024).

#### coxme

A función **coxme** do paquete **coxme** permite tamén a estimación de modelos semiparamétricos de fragilidade (Therneau, 2024), neste caso unicamente con distribución log-normal

```
coxme(Surv(time, status) ~ cov1 + cov2 + ... + (1|frailty.var), data = dat)
```

#### gss

Ademais destes dous paquetes, a función **sscox** do paquete **gss** permite axustar un modelo de Cox semiparamétrico usando splines cúbicos para a estimación dos parámetros. Este paquete permite introducir un parámetro de fragilidade con distribución log-normal (Gu, 2014)

```
sscox(Surv(time, status) ~ cov1 + cov2 + ..., random = ~ 1|frailty.var, data = dat)
```

#### Outros paquetes

O paquete **R2BayesX**, a través da función **bayesx** tamén permite axustar esta clase de modelos segundo indican Gorfine e Zucker (2023) pero non se encontrou literatura sobre o seu funcionamento.

O paquete **frailtyPack** era outra opción para o axuste destes modelos pero recentemente foi retirada do repositorio CRAN (Comprehensive R Archive Network, 2024).

## 3.4. Verosimilitude xerárquica

Ha et al. (2001) propoñen outro método chamado verosimilitude xerárquica. Este método non se detalla nesta memoria porque faría preciso explicar os modelos xeralizados xerárquicos de Poisson, que non é o obxecto deste traballo, e porque a innovación respecto dos métodos mencionados anteriormente para o modelos nos que se centra este traballo, non é moi importante.

De acordo con Ha et al. (2001; 2003) para o modelo de riscos proporcionais con distribución gamma das fragilidades, os estimadores dos parámetros con este método, serían equivalentes aos dous métodos anteriores. Por outra parte, para un modelo de riscos proporcionais con distribución das fragilidades

log-normal, este método sería unha extensión da verosimilitude parcial penalizada para datos con empates nos tempos de evento. Os autores non dan ningunha indicación sobre a equivalencia ou non dos métodos para fraxilidades distribuídas segundo unha gaussiana inversa, positiva estábel ou outras distribucións potencia-varianza.

### 3.4.1. Implementación en R

#### frailtyHL

Ha et al. (2019) implementaron este método de estimación para o caso semiparamétrico no paquete **frailtyHL** de R coa función homónima.

```
frailtyHL(Surv(time, status) ~ cov1 + cov2 + ... + (1|frailty.var), data = dat, RandDist = "Normal")
```

A distribución das fraxilidades indícase en `RandDist` e pode ser normal, `Normal`, ou gamma, `Gamma`.

A modo de mención, coméntase que este paquete permite tamén axustar modelos de fraxilidade con tempo de fallo acelerado usando a función `mlmfit`.

## 3.5. Pseudoverosimilitude completa

O último dos métodos de estimación foi proposto por Monaco et al. (2018). Para estudar este método, comezan por reescribir a verosimilitude incondicional (2.8) como

$$L(\beta, \sigma^2 | \tau) = \prod_{i=1}^n \left[ \prod_{j=1}^{n_i} \left[ \left( \lambda_0(T_{ij}) e^{\beta^t \mathbf{X}_{ij}} \right)^{\delta_{ij}} \right] (-1)^{n_i(\tau)} \mathcal{L}^{(n_i(\tau))}(\Lambda_{\text{clúster}}(\mathbf{Y}_i(\tau) | \mathbf{X}_i)) \right],$$

onde  $\tau$  é o tempo de seguimento e  $Y_{ij}(\tau)$ , o tempo de supervivencia observado ata  $\tau$ ,  $Y_{ij}(\tau) = T_{ij} \wedge \tau$ .  $\Lambda_{\text{clúster}}(\mathbf{t}_i | \mathbf{X}_i)$  é a suma dos riscos de cada individuo dun clúster, a excepción de  $Z$ , como se viu en (2.4). Por último,  $n_i(t)$  é o número de individuos no clúster  $i$  que seguen sendo observados en tempo  $t$ ,  $n_i(t) = \sum_{j=1}^{n_i} \mathbb{1}(T_{ij} \geq t)$

A log-verosimilitude correspondente é

$$\begin{aligned} \log(L(\beta, \sigma^2 | \tau)) &= \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij} [\log(\lambda_0(T_{ij})) + \beta^t \mathbf{X}_{ij}] \\ &\quad + \sum_{i=1}^n \log \left( \mathcal{L}^{(n_i(\tau))}(\Lambda_{\text{clúster}}(\mathbf{Y}_i(\tau) | \mathbf{X}_i)) \right). \end{aligned} \quad (3.2)$$

Para obter estimacións de  $\beta$  e  $\sigma^2$  é preciso un estimador de  $\Lambda_0$ . Ata agora consideramos un único tempo de seguimento, pero poderíamos considerar varios, engadindo un subíndice a  $\tau$ ,  $\tau_k$ . Usando esta notación, para certo valor de  $\beta$  e  $\sigma^2$ ,  $\Lambda_0$  estímase a través dunha función discontinua con saltos nos tempos de evento observados,  $\tau_l, l = 1, \dots, L$  con

$$\begin{aligned} \hat{\lambda}_0(\tau_l) &= \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \mathbb{1}(T_{ij} = \tau_l)}{\sum_{i=1}^n \psi \left( \gamma, \hat{\Lambda}_0, \tau_{l-1} \right) \sum_{j=1}^{n_i} \mathbb{1}(T_{ij} \geq \tau_l) e^{\beta^t \mathbf{X}_{ij}}} \\ \hat{\Lambda}_0(\tau) &= \sum_{\tau_l \leq \tau} \hat{\lambda}_0(\tau_l), \\ \psi \left( \gamma, \hat{\Lambda}_0, \tau_l \right) &= \frac{\mathcal{L}^{(n_i(\tau_l))}(\Lambda_{\text{clúster}}(\mathbf{Y}_i | \mathbf{X}_i, \tau_l))}{\mathcal{L}^{(n_i(\tau_l)+1)}(\Lambda_{\text{clúster}}(\mathbf{Y}_i | \mathbf{X}_i, \tau_l))}. \end{aligned}$$

Con este enfoque, pode verse  $\psi\left(\gamma, \hat{\Lambda}_0, \tau_{l-1}\right) Y_{ij}(\tau_l) e^{\beta^t \mathbf{x}_{ij}}$  como unha covariábel dependente do tempo e usarse o seguinte algoritmo de estimación

1. Utilizar a regresión de Cox estándar para obter estimadores de  $\beta$  e fixar  $\sigma^2 = 0$ .
2. Usar os valores actuais de  $\beta$  e  $\sigma^2$  para estimar  $\Lambda_0$  usando (3.2).
3. En base á estimación de  $\Lambda_0$ , obter estimacións de  $\beta$  e  $\sigma^2$  maximizando  $\log(L(\beta, \sigma^2))$ .
4. Reiterar entre os pasos 2 e 3 ata obter converxencia.

Unha das maiores desvantaxes que presenta este método é que de cara á estimación do risco en tempo  $\tau_l$  só se utiliza a información dispoñíbel e ata o tempo  $\tau_l$ , co que pode achegar estimadores menos eficientes que outros métodos que utilizan toda a información dispoñíbel (Gorfine et al., 2006).

### 3.5.1. Implementación en R

#### frailtySurv

Os autores do anterior artigo (Monaco et al., 2018) desenvolveron asimesmo un paquete de R para a utilización deste método de estimación, **frailtySurv**. A función `fitfrail` deste paquete permite axustar modelos de fragilidade usando esta rutina

```
fitfrail(Surv(time, status) ~ cov1 + cov2 + ... + cluster(frailty.var), dat = dat, frailty = "gamma")
```

A variábel `frailty` indica a distribución da fragilidade que pode ser `gamma`, gaussiana inversa (`invgauss`), positiva estábel (`posstab`), potencia-varianza (`pvf`) ou log-normal (`log-normal`). No caso da distribución log-normal, pola ausencia de transformada de Laplace, a integral debe avaliarse numericamente en cada iteración o que aumenta considerabelmente o tempo de computación que se precisa para a estimación.

## 3.6. Comparación dos paquetes de estimación

Wu et al. (2023), nun artigo pendente da revisión por pares, fixeron un estudo comparativo do funcionamento dalgúns dos paquetes presentados neste capítulo en termos de eficiencia e precisión para o axuste de modelos de fragilidade con distribución gamma a través do axuste de resultados simulados. Nesta sección preséntanse os resultados do seu estudo excluindo os resultados para o paquete `frailtyPack` que, como se mencionou anteriormente, foi recentemente retirado do repositorio CRAN.

Os autores simularon tempos de evento usando unha distribución de Weibull de parámetro de forma  $\alpha = 3$  e parámetro de escala  $\lambda = 0.007$  ao que engadiron tres covariábeis coa forma dos modelos de Cox en (1.6) e un termo de fragilidade, isto é

$$t_{ij} = \left( \frac{-\log(u_{ij})}{\lambda \exp(\beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij}) z_i} \right)^{1/\alpha}$$

con  $u_{ij}$  simulada da distribución uniforme estándar,  $U[0, 1]$ .

A distribución das  $x_1$  tamén é uniforme estándar,  $U[0, 1]$ , a das  $x_2$  é a normal estándar  $N(0, 1)$  e por último  $x_3$  segue unha distribución de Bernoulli de parámetro 0.25, *Bernoulli*(0.25). Os auténticos valores dos parámetros de regresión fixéronse en  $\beta = 1$ ,  $\beta = -1$  e  $\beta = 0.5$ . Xeráronse tamén tempos de censura dunha exponencial de parámetro  $\theta$  escollido de xeito que a proporción de observacións censuradas fose en media 20 %, 50 % e 80 %, respectivamente. Por último, a fragilidade xerouse dunha distribución gamma de media 1 e varianza 0.5.

Respecto do tamaño da poboación e dos clusters, consideráronse 5 escenarios diferentes:

- unha mostra de tamaño 100, con  $n = 10$  clusters de tamaño  $n_i = 10$ ,
- unha mostra de tamaño 400 con
  - $n = 40$  clusters de tamaño  $n_i = 10$  ou
  - $n = 10$  e  $n_i = 40$ , e
- unha mostra de tamaño 800 con
  - $n = 80$  e  $n_i = 10$ , ou
  - $n = 10$  e  $n_i = 80$ .

Considerando os 5 escenarios para o tamaño da poboación e os clusters e tres escenarios para a taxa de censura, consideráronse un total de 15 escenarios. Para cada escenario simuláronse 1000 mostras e axustouse o modelo con tódolos paquetes en CRAN que permiten axustar un modelo de fraxilidade con esa distribución.

O estudo non indica se, á hora de axustar os modelos asumindo unha forma paramétrica do risco basal co paquete **parfm**, se supuxo coñecida a distribución dos tempos de evento ou, no caso contrario, cal foi o criterio de selección.

Os autores do artigo mencionado conclúen que entre os paquetes dispoñíbeis para o axuste destes modelos (**survival**, **parfm**, **frailtyEM**, **frailtySurv** e **frailtyHL**), o paquete **survival** representa a mellor opción xa que é o máis rápido, cun tempo de computación que reduce en máis dun 90 % o que precisa calquera dos outros paquetes, e tamén é o que ofrece unha maior taxa de converxencia, medida como a proporción de veces que a rutina de estimación converxe e se obteñen estimacións, situándose en ou moi preto de 100 % en tódolos escenarios, comportándose moito mellor que outros paquetes.

Prestando atención ao nesgo, o paquete **survival** tamén ofrece un rendemento bo. Outro paquete que tamén ofrece un bo rendemento á hora de minimizar o nesgo é o paquete **frailtyEM**. Ademais, pode dicirse que **frailtyEM** ofrece o mellor rendemento á hora de minimizar o erro cadrático medio das estimacións.

Dado que a diferenza no tempo preciso para o axuste entre **frailtyEM** e **survival** é só duns segundos, podería ser recomendábel usar este paquete para axustar distribucións gamma, e só usar **survival** nas raras ocasións nas que esta rutina de estimación non converxa, especialmente cando o tamaño de mostra non sexa excesivo e a diferenza poida medrar. Ademais dos argumentos anteriores, este paquete achega información adicional útil como a desviación estándar da varianza da fraxilidade ou o resultado do test de heteroxeneidade Commenges-Andersen.

Sería interesante ter datos para outras das distribucións. De feito, considerouse facelo como parte deste traballo, pero dadas as características do estudo de simulación e o tempo de axuste medio obtido polos autores, que excedía nalgúns casos os dous minutos, claramente non é factíbel co tempo e medios dispoñíbeis.

### 3.7. Robustez da distribución gamma

Como se comentou anteriormente neste traballo, a distribución gamma é a máis utilizada historicamente para o axuste dos modelos de fraxilidade. En moitos casos, a falta de información sobre a distribución das fraxilidades, axústanse modelos con fraxilidade gamma. Isto é algo frecuente xa que, malia existir métodos de bondade de axuste, estes son limitados e a miúdo é difícil saber que outra distribución é mellor para modelizar as fraxilidades. Dado que na práctica se usa a distribución gamma, mesmo cando é significativamente diferente da verdadeira distribución das fraxilidades, é importante saber cal é a robustez das estimacións obtidas baixo a suposición de fraxilidade con distribución gamma cando a distribución verdadeira das fraxilidades é outra.

Gorfine et al. (2012) estudaron a robustez a través dun estudo de simulación, no que supuxeron que o risco basal era exponencial de parámetro 1, incluíndo unha única covariábel coa forma dos modelos de Cox, e que os tempos de censura seguían unha distribución uniforme entre 0 e ben 0.3 ou ben 3, de xeito que a taxa de censura se situase respectivamente entre o 60 % e o 80 % e entre o 30 % e o 40 %. A covariábel  $X$  xerouse dunha  $Bernoulli(0.5)$  e o valor real do efecto fixo  $\beta$  fixouse en  $\beta = \log(2) \approx 0.693$ . Ademais, xeraron unha fragilidade compartida dunha distribución gamma, gaussiana inversa e positiva estábel co parámetro escollido de xeito que o nivel de dependencia, medido usando a  $\tau$  de Kendall mencionada antes nesta memoria, se mantivese constante en 0.45. Xeráronse 20,000 tempos independentes. De entre estes escolléronse 500 e xunto con outros 500 formáronse 500 parellas ou clusters de tamaño 2, para os que se xerou o tempo dada a fragilidade observada. Cada escenario foi analizado xerando e axustando 1,000 mostras aleatorias.

Os resultados da simulación amosan que o nesgo e a perda de eficiencia introducidos pola incorrecta especificación da distribución de fragilidade son mínimos. Daí, conclúen que convén usar a distribución de fragilidade gamma cando se descoñece a distribución das fragilidades.

Outro estudo de simulación similar foi feito por Hsu et al. (2007). Neste caso consideráronse como distribucións da fragilidade a gaussiana inversa, a positiva estábel e a discreta que toma un de entre dous valores. Alén da fragilidade engadiu tamén a mesma covariábel Bernoulli e o parámetro de regresión correspondente. Neste estudo consideráronse dous conxuntos de parámetros, para inducir a dous niveis de dependencia diferentes,  $\tau = 0.20$  e  $\tau = 0.45$ . Ademais consideráronse respectivamente taxas de censura de 20 %, 50 % e 80 %. Os resultados e as conclusións dos autores son similiares ás obtidas no estudo mencionado anteriormente.

Anteriormente, Glidden e Vittinghoff (2004) fixeran outra análise similar con outro estudo de simulación. No seu caso, en lugar de modificar a taxa de censura modificaron o tamaño da mostra e o tamaño dos clusters, considerando un tamaño de mostra de 100 ou 400, e un tamaño de clúster de 2, 10 ou 20. O risco basal simulouse dunha Weibull de parámetros  $\alpha = 1.3$  e  $\lambda = 5$ . As fragilidades xeráronse dunha distribución gamma, gaussiana inversa ou positiva estábel, escollendo o parámetro de xeito que a medida de dependencia de Kendall,  $\tau$ , fose 0.5 en tódolos casos. O efecto das covariábeis tamén tiña a forma dos modelos de Cox, con valor  $\beta = \log(2) = 0.693$ . A covariábel  $X$  tomaba valor 1 ou 0, pero non se xerou aleatoriamente senón que cada metade dos membros de cada clúster se asignou a unha ou outra opción, reflectindo o que sucede co tratamento en ensaios clínicos levados a cabo nun número de centros.

Os resultados destes estudos indican que a incorrecta especificación da distribución de fragilidade non introduce importantes perturbacións nas estimacións dos parámetros de regresión.



## Capítulo 4

# Métodos de diagnóstico dos modelos de fragilidade

Neste capítulo recóllese algúns métodos de diagnóstico para os modelos de regresión con fragilidade, resúmense os resultados de estudos numéricos que estudan a súa validez e inclúese, cando esta existe, a implementación en R do método. O capítulo divídese en tres seccións, a primeira trata os modelos de diagnóstico de bondade de axuste do modelo de regresión, en segundo lugar trátanse os tests para detectar heteroxeneidade, é dicir, para detectar a presenza de fragilidade e, por último, preséntanse os tests de bondade de axuste da distribución suposta para as fragilidades.

### 4.1. Tests de bondade de axuste do modelo de regresión

Nos manuais e revisións dos modelos de fragilidade consultados (Balan & Putter, 2020; Duchateau & Janssen, 2008; Gorfine & Zucker, 2023; Govindarajulu & D’Agostino, 2020; Wienke, 2010) non se presta moita atención aos tests de bondade de axuste da forma do risco basal ou da forma dos preditores lineares, poñendo, como se verá nas seguintes seccións, o foco en comprobar a presenza de heteroxeneidade e na bondade de axuste da distribución da fragilidade. A menor importancia que se lle dá está en liña coa escasa literatura nesta área. A continuación, recóllese os 3 tests identificados na literatura.

#### 4.1.1. Tests de Bagdonavičius e Nikulin e de Vonta

Este apartado presenta de forma conxunta os tests propostos por Bagdonavičius e Nikulin (2011) e por Vonta (2020), xa que comparten a mesma idea.

Ámbolos dous tests propoñen dividir o tempo de observación da mostra en  $k$  intervalos coa forma  $(t_{l-1}, t_l]$ ,  $l \in 1, \dots, k$ , con  $t_0 = 0$  e  $t_k = \max(t_{ij})$  e comparar o número de eventos observado no intervalo,

$$d_l = \sum_{i=1}^n \sum_{j=1}^{n_i} \delta_{ij} \mathbb{1}(t_{l-1} < t_{ij} \leq t_l),$$

co número de eventos esperado

$$\Lambda_l = \sum_{ij: T_{ij} > t_{l-1}} [\Lambda(t_l \wedge T_{ij} | \mathbf{X}) - \Lambda(t_{l-1} | \mathbf{X})]$$

O test de Bagdonavičius e Nikulin propón en base a esta idea construír o estatístico

$$T_{BN} = \sum_{l=1}^k \frac{(d_l - \Lambda_l)^2}{d_l} + Q.$$

A forma de  $Q$  é excesivamente complexa para describirla nun resumo breve do método, pero depende das derivadas da función de risco baixo a hipótese nula respecto dos parámetros estimados, e leva a que o estatístico  $T_{BN}$  teña distribución chi-cadrado cun parámetro que tamén depende destas derivadas, co que se rexeita a hipótese nula de que a distribución siga o modelo de regresión proposto para valores excesivamente grandes do estatístico.

O estatístico proposto por Vonta non está completamente claro xa que tanto este como outro artigo anterior con Karagrigoriou (Vonta & Karagrigoriou, 2014) conteñen un erro na definición dos termos do estatístico. Ao noso entender, o estatístico proposto é

$$T_{V,\varphi} = \arg \min_{\theta \in \Theta} \sum_{l=1}^k n_l \left( \hat{F}_l \cdot \varphi \left( \frac{d_l/n_l}{\hat{F}_l} \right) + \hat{S}_l \cdot \varphi \left( \frac{(n_l - d_l)/n_l}{\hat{S}_l} \right) \right)$$

onde  $\hat{F}_l$  e  $\hat{S}_l$  son respectivamente a función de distribución e supervivencia estimada dos eventos no intervalo  $l$ ,

$$\hat{S}_l = 1 - \hat{F}_l = 1 - \frac{e^{-\Lambda_{l-1}} - e^{-\Lambda_l}}{e^{-\Lambda_{l-1}}}$$

e  $\varphi(\cdot)$  é unha función que mide a diverxencia, que ten as seguintes propiedades:

- $\varphi(x)$  é continua, diferenciable e convexa en  $x \geq 0$ ;
- $\varphi(1) = 0$
- $\varphi'(1) = 0$

como a función de Kullback-Leiber. A forma final do estimador dependerá da función escollida. Con unha das opcións achegadas chégase, de novo, á distribución chi-cadrado de Pearson.

Respecto do número e xeito de formar os  $k$  grupos, propóñense distintos métodos entre ambos artigos, que teñen como obxectivo levar a un número de eventos esperados en cada intervalo similar, e a unha maximización da potencia do test.

#### 4.1.2. Test de González-Manteiga, Martínez-Miranda e Van Keilegom

González-Manteiga, Martínez-Miranda e Van Keilegom (2023) propuxeron un test para detectar o efecto non linear das covariábeis na supervivencia para modelos con fraxilidade.

O test proposto consiste en axustar o modelo baixo a hipótese nula de efecto linear e baixo a hipótese alternativa usando bases ortogonais. A continuación, propóñense usar como estatístico de test a diferenza entre a log-verosimilitude baixo a hipótese nula e baixo a hipótese alternativa multiplicada por -2. En terceiro lugar, estimaríase a distribución do estatístico de test a través do seguinte algoritmo de remostraxe bootstrap:

1. Baixo a hipótese nula, axústase o modelo e obtéñense estimacións do risco basal,  $\lambda_0$ , dos parámetros de regresión,  $\beta$ , e o parámetro da distribución das fraxilidades  $\theta$ .
2. Obtéñense os efectos aleatorios  $W'_i = \exp(Z'_i)$ ,  $i = 1, \dots, n$  da súa distribución suposta substituíndo o parámetro descoñecido  $\theta$  pola súa estimación.



3. Xéranse tempos de supervivencia  $T_{ij}^*$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, n_i$  a través da función de supervivencia estimada

$$\hat{S}(t|\mathbf{X}_{ij}, W_i') = \exp\left(-\Lambda_0(t)e^{\hat{\beta}^t \mathbf{X}_{ij} + W_i'}\right),$$

con  $\hat{S}(t|\cdot)$  a función de supervivencia estimada.

4. Xéranse tempos de censura  $C_{ij}'$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, n_i$  usando o estimador de Cox da distribución de censura

$$\hat{G}(c|\mathbf{X}_{ij}) = \hat{G}_0(c)^{\exp(\hat{\beta}_\delta^t \mathbf{X}_{ij})},$$

con  $\hat{\beta}_\delta$  os estimadores dos parámetros de regresión da función de distribución da censura.

5. Fíxanse  $T_{ij}' = \min(T_{ij}^*, C_{ij}')$  e  $\delta_{ij}' = \mathbb{1}(T_{ij}^* \leq C_{ij}')$ . A mostra bootstrap é

$$\left\{ \{T_{ij}', \delta_{ij}', \mathbf{X}_{ij}\}; i \in 1, \dots, n; j = 1, \dots, n_i \right\}.$$

6. Recálculase o número de bases ortogonais óptimo usando o criterio de información de Akaike para chegar á distribución baixo a hipótese alternativa e obtense o estatístico de test.

Despois de repetir os pasos 1 a 6 un número grande de veces, estímase o valor crítico do test para un nivel de significación  $\alpha$  como o cantil  $(1 - \alpha)$  dos estatísticos na remostraxe e rexéitase o test para un valor do estatístico maior ao valor crítico.

Os autores conclúen a súa publicación cun estudo numérico das propiedades do test proposto. O test proposto amosa un bo comportamento baixo a hipótese nula, co nivel de rexeitamentos aproximadamente igual a  $100\alpha\%$  mais un rendemento moderado baixo a hipótese alternativa, precisando dun tamaño de mostra importante para detectar as desviacións da hipótese nula a maioría das veces, mesmo cando a hipótese alternativa é moi distante da nula e a variabilidade das mostras é limitada.

## 4.2. Test de heteroxeneidade

O único test que foi identificado na literatura para comprobar a presenza de heteroxeneidade é o proposto por Commenges e Andersen (1995).

### 4.2.1. Test de Commenges-Andersen

O test de Commenges-Andersen é un test para a detección de heteroxeneidade nunha mostra de datos de supervivencia coa forma dun modelo de fragilidade multiplicativo, como o que se vén tratando nesta memoria. O test foi publicado nun artigo destes autores (Commenges & Andersen, 1995).

Proponse escribir o termo de fragilidade como  $Z_i = \exp(W_i\sigma)$ , con  $W_i$  variábeis aleatorias independentes e identicamente distribuídas segundo unha distribución sen especificar de media 0 e varianza 1. A hipótese nula polo tanto é  $H_0 : \sigma = 0$  e a hipótese alternativa  $H_1 : \sigma > 0$ .

En base ao axuste dun modelo de riscos proporcionais de Cox (sen fragilidade) e á función de verosimilitude pode chegarse ao seguinte estatístico

$$T = \sum_{i=1}^n \left[ \sum_{j \neq k}^{n_i} M_{ij} M_{ik} \right] + \left( \sum_{i=1}^n \left[ \sum_{j=1}^{n_i} M_{ij}^2 - N \right] \right) + C,$$

O estatístico é a suma de tres termos. O primeiro, con  $M_{ij} = \delta_{ij} - \hat{\Lambda}_0(T_{ij}) \exp(\hat{\beta}^t \mathbf{X}_{ij})$ , é a suma das correlacións entre cada par de individuos nun clúster. O segundo é unha medida da dispersión excesiva

e o terceiro,

$$C = \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\delta_{ij}}{S^{(0)}(T_{ij})^2} \sum_{k=1}^n \left[ \sum_{l=1}^{n_i} \mathbb{1}(T_{kl} > T_{ij}) \exp(\hat{\beta}^t \mathbf{X}_{kl}) \right]^2$$

con  $S^{(0)}(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} \mathbb{1}(T_{ij} > t) \exp(\hat{\beta}^t \mathbf{X}_{ij})$ , é dicir a suma dos preditores lineais dos pacientes baixo risco en tempo  $t$ , é un termo de corrección que tende a cero a medida que  $N$  medra, sendo  $N$  o tamaño total da mostra,  $N = \sum_{i=1}^n n_i$ .

Cabe destacar que, no caso de avaliar a existencia de fraxilidade univariante, é dicir un termo de fraxilidade por individuo, o estatístico redúcese a unha medida de dispersión excesiva corrixida.

Para estimar a varianza de  $T$  chamemos  $\hat{\Sigma}$  á matriz de varianzas-covarianzas estimada dos  $\hat{\beta}$  con elementos  $\hat{\sigma}_{kl}$ ,  $1 \leq k \leq p$ ,  $1 \leq l \leq p$ . Denotemos

$$p_{ij}(t) = \frac{\mathbb{1}(T_{ij} < t) e^{\hat{\beta}^t \mathbf{X}_{ij}}}{S^{(0)}(t)}$$

e

$$\bar{p}_i(t) = \sum_{j=1}^{n_i} p_{ij}(t), \quad 1 \leq i \leq n$$

para cada tempo de evento distinto  $t_l$   $l = 1, \dots, L$ . Denotemos por  $D(t_l)$  o número de mortes observadas en tempo  $t_l$ . Lembremos que o residuo martingala para a observación  $ij$  en tempo  $t_l$  é

$$M_{ij}(t_l) = \begin{cases} \delta_{ij} - \Lambda_0(T_{ij}) \exp(\hat{\beta}^t \mathbf{X}_{ij}) & \text{se } T_{ij} \leq t_l \\ -\Lambda_0(t_l) \exp(\hat{\beta}^t \mathbf{X}_{ij}) & \text{se } T_{ij} > t_l. \end{cases}$$

Sexa  $\Lambda_{\text{clúster } i}(t_l | \mathbf{X}_i) = \sum_{j=1}^{n_i} \Lambda(t_l | \mathbf{X}_{ij})$  e  $\Lambda_{\text{clúster } i}(t_0 | \mathbf{X}_i) = 0$ . Para cada grupo, calculemos,

$$Q_i(t_l) = 2 \left[ \Lambda_{\text{clúster } i}(t_{l-1} | \mathbf{X}_i) - \sum_{g=1}^n \Lambda_{\text{clúster } g}(t_{l-1}) \bar{p}_g(t_l) - \bar{p}_i(t_l) + n \bar{p}_i(t_l) \right]$$

Finalmente para cada unha das  $p$ -covariábeis  $h = 1, \dots, p$  calculamos

$$\theta_h = \sum_{i=1}^n \sum_{j=1}^{n_i} Q_i(t_k) d(t_k) \left\{ \sum_{j=1}^{n_i} X_{ijh} \right\}$$

A varianza estimada do estatístico vén dada por

$$V = \sum_{i=1}^n \sum_{j=1}^{n_i} Q_i(t_k)^2 \bar{p}_i(t_k) d(t_k) \sum_{h=1}^p \sum_{k=1}^p \theta_h \theta_k \hat{\sigma}_{hk}.$$

Polo tanto, o estatístico de test  $T$  ten distribución asintótica  $N(0, \sqrt{V})$  e rexéitase a hipótese nula para

$$\Phi \left( \left| \frac{T}{\sqrt{V}} \right| \right) > \alpha$$

### Implementación en R

Como se viu anteriormente, este test está implementado no paquete **frailtyEM**. O test aplícase por defecto ao axustar un modelo usando a función **emfrail** do mesmo paquete. Tamén se pode usar aplicando a función **ca\_test** do mesmo paquete, **frailtyEM**, á saída da función **coxph** do paquete **survival**.

### 4.3. Tests de bondade de axuste da distribución de fraxilidade

#### 4.3.1. Test de Shih e Louis

O primeiro método desenvolvido co obxectivo de probar a validez da distribución utilizada para modelizar as fraxilidades que se inclúe nunha maioría de manuais é un método gráfico proposto por Shih e Louis (1995). Este método permite contrastar a hipótese nula de que a distribución de fraxilidade é gamma.

Se o modelo é correcto, para  $t \in [0, \tau]$ ,  $i = 1, \dots, n$ ,

$$\mathbb{E} \left[ \hat{Z}(t|\theta) \right] = \mathbb{E}_\theta \left[ \frac{1 + D_i(t)}{1 + \sum_{j=1}^{n_i} [\Lambda_0(T_{ij} \wedge t) e^{\beta^t \mathbf{x}_{ij}} Z_i]} \right] = 1$$

con  $D_i(t)$  o número de eventos no clúster  $i$  en tempo  $t$ .

En  $t = 0$ , a predición de  $Z_i$  condicionada a  $\theta$  é igual a 1,  $\hat{Z}_i(0|\theta) = 1$ . Por outra parte, por definición, usando toda a información observada no estudo,  $t = \tau$ , a media das predicións das fraxilidades  $Z_i$  condicionadas a  $\hat{\theta}$  é igual a 1,  $\bar{Z}(\tau|\hat{\theta}) = 1$ . A variábel  $\hat{\theta}$  é a solución ás ecuacións de estimación  $U(\theta|\tau) = 0$  con

$$U(\theta|\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \log(L_i(\theta|\tau))}{\partial \theta}.$$

Isto indica que o estatístico de diagnóstico  $\bar{Z}(t|\theta) - 1$  pode unificarse coa función score escalar  $U(\beta_0|t)$  para o termo de compensación ou offset  $\beta_0$  e dá pé a unha interpretación interesante. Se o modelo é correcto, a función  $U(\beta_0|t)$  no verdadeiro valor de  $\theta$  ten esperanza 0 para calquera  $0 \leq t \leq \tau$ , e polo tanto  $\hat{\beta}_0(t)$ , a solución de  $U(\beta_0|t)$ , debera estar preto do verdadeiro valor do parámetro  $\beta_0$ . En caso contrario, toda variación de  $t$  non explicada polo modelo será absorbida polo intercepto dependente do tempo,  $\hat{\beta}_0(t)$ , que non converxe a  $\beta_0$ .

Daí, baixo a especificación correcta da distribución de fraxilidade,  $\bar{Z}(t|\theta)$  debería fluctuar ao redor de 1. Pola lei dos grandes números de Markov, como amosan os autores no artigo,  $\bar{Z}(t|\theta)$  converxe a 1 en probabilidade. Pode probarse tamén que  $\sqrt{n}(\bar{Z}(t|\theta) - 1)$  converxe a unha normal de media 0 e varianza limitada. Substituíndo  $\theta$  por  $\hat{\theta}$  mantense o valor da media pero redúcese a varianza de xeito proporcional á forza da correlación entre  $(\bar{Z}(t, \theta) - 1)$  e  $U(\theta|t)$ .

Os autores propoñen utilizar estas propiedades para construír unha ferramenta de diagnóstico gráfica, facendo unha gráfica de  $\bar{Z}(t|\hat{\theta}) - 1 \pm 2SE$  sobre  $t$ , co erro estándar axudando a discernir a variación na gráfica. Baixo a especificación correcta do modelo e da distribución das fraxilidades,  $\bar{Z}(t|\theta)$  fluctúa preto do 1, achegándose a medida que crece  $t$ .

#### 4.3.2. Tests de Glidden e de Cui e Sun

Glidden (1999) propón un test para contrastar que a distribución de fraxilidade dunha poboación é gamma que ofrece a vantaxe respecto do anterior método de non precisar a correcta especificación da distribución do risco basal para contrastar a hipótese nula de que a distribución das fraxilidades é gamma. O artigo inclúe, ademais, o resultado dun estudo de simulación analizando o funcionamento empírico do test. Posteriormente, Cui e Sun (2004) adaptaron o método para modelos de regresión con covariábeis e realizaron outro estudo de simulación sobre a súa validez empírica.

Os autores amosan, de xeito similar ao test anterior e en liña co que se vén tratando neste traballo, que a predición das fraxilidades usando a información dispoñíbel ata o tempo  $\tau$  menos a súa media,

que por definición é 1,

$$\hat{z}_i(\tau) - 1 = \frac{1/\sigma^2 + D_i(t)}{1/\sigma^2 + \sum_{j=1}^{n_i} \Lambda_0(\tau \wedge T_{ij}) e^{\beta^t \mathbf{X}_{ij}}} - 1,$$

con  $D_i(t)$  definido do mesmo xeito que no método anterior, é unha variábel de media 0.

Como amosan os autores, a suma normalizada destes valores das fraxilidades pode ser separada nunha parte con distribución idéntica e independente entre os clusters,  $\Phi_i(\tau)$ , e unha segunda parte que converxe a 0 en probabilidade

$$\hat{W}(\tau) = n^{-1/2} \sum_{i=1}^n [\hat{z}_i(\tau) - 1] = n^{-1/2} \sum_{i=1}^n [\Phi_i(\tau)] + o_p(1)$$

ten distribución normal multivariante e converxe de xeito feble a un proceso gaussiano.

Dada a dificultade de obter a matriz de varianzas-covarianzas, os autores propoñen obter a distribución baixo a hipótese nula de  $\hat{W}(\tau)$  por simulación usando

$$W'(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\Phi}_i(\tau) \xi_i,$$

onde as  $\xi$  son variábeis con distribución normal estándar.

Cando a distribución das fraxilidades é correcta, as  $\hat{W}(\tau)$  nun gráfico sobre  $t$  estarán preto de 0. Engadindo ao gráfico un número de simulacións (os autores propoñen 20 pero coa capacidade computacional actual poderían ser máis) poderemos ver se as desviacións de 0 son significativas. Os autores propoñen asimesmo usar procedementos como o método do supremo para obter un test numérico en base ao mesmo principio.

Cabe comentar que, de acordo con Glidden (1999), para  $n \rightarrow \infty$  este test tamén se encontra baixo a hipótese nula cando a distribución das fraxilidades non sexa gamma, pero a varianza das fraxilidades tenda a 0.

Tanto o estudo de simulación de Glidden (1999), como o de de Cui e Sun (2004) amosan que o test ten unha potencia razoábel para tamaños de mostra moi pequenos (50), especialmente cando a censura é baixa ( $\leq 20\%$ ) ou o tamaño de clúster é maior (5) e unha potencia moi boa para tamaños de mostra algo maiores (200), mais tamén que é excesivamente conservador para distribucións de fraxilidade con varianza pequena, cun porcentaxe de rexeitamento dos tests inferior ao nivel de significación.

### 4.3.3. Test de Geerdens, Claeskens e Janssen

Geerdens, Claeskens e Janssen (2013) propuxeron un test numérico para comprobar se a distribución das fraxilidades é gamma, baseado na comparación entre a verosimilitude do axuste dun modelo con distribución de fraxilidade gamma e a verosimilitude de modelos axustados cunha clase de distribucións gamma xeneralizadas co uso de expansións ortogonais.

Os autores fixeron tamén un estudo de simulación para estudar o funcionamento do test. Os resultados amosan unha potencia limitada á hora de detectar que a distribución das fraxilidades é gaussiana inversa no canto de gamma, e en cambio moita potencia cando a distribución real é positiva estábel.

### 4.3.4. Test de Munda e Legrand

O test de Munda e Legrand (2013) é diferente dos anteriores porque non permite unicamente contrastar se a distribución das fraxilidades é gamma, senón que simultaneamente amosa se a positiva estábel ou gaussiana inversa achegan un mellor axuste. O test está formulado para mostrase en covariábeis con

clusters de tamaño 2 aínda que, como se indica máis adiante, se propoñen adaptacións para outras situacións.

O test baséase na distinta forma da dependencia entre un par de observacións nun clúster segundo a distribución das fraxilidades, como se amosou na figura 2.2. Proponse amosar nunha gráfica os valores de dependencia observados. Estes calcúlanse como a función de supervivencia conxunta empírica en  $t_1$  e  $t_2$ , con  $\hat{S}(t_1) = \hat{S}(t_2) = q$  para distintos valores  $q$  con  $\hat{S}(t)$  o estimador Kaplan-Meier da supervivencia, dividida por  $q$ ,  $\frac{\hat{S}(t_1, t_2)}{q}$ , con

$$\hat{S}(t_1, t_2) = \mathcal{L}(\mathcal{L}^{-1}(\hat{S}(t_1)) + \mathcal{L}^{-1}(\hat{S}(t_2))).$$

Estes valores úsanse nunha gráfica na que as ordenadas indican a probabilidade de supervivencia conxunta observada, e as abscisas indican o valor da supervivencia individual de cada membro do clúster  $q$ . Baixo independencia os puntos seguen a diagonal con ordeada na orixe 0 e pendente 1. Ademais dos valores observados amósanse na gráfica, a forma esperada para distintas distribucións de fraxilidade, que corresponderían ao valor da figura 2.2 na diagonal.

Á hora de expandir este test a mostras de datos con covariábeis, os autores propoñen usar dous estimadores de Kaplan-Meier cando se trate dunha única covariábel binaria, e substituír o estimador Kaplan-Meier por un estimador kernel noutras situacións. De cara a expandilo a tamaños de clúster maiores a 2, propoñen seleccionar aleatoriamente unha parella de cada clúster e facer a estimación anterior, repetir este exercicio un número de veces e utilizar o valor promedio para cada clúster.

Nun estudo de simulación para comprobar a súa validez, no que se busca escoller cal é a distribución das fraxilidades entre a gamma e a positiva estábel minimizando a suma dos cadrados das diferenzas entre os valores observados na gráfica e os valores preditos para cada distribución, os autores demostran que o test ten gran capacidade predictiva, levando á escolla correcta da distribución das fraxilidades entre un 38% e un 100% das veces, segundo o escenario, coa maioría de casos achegándose máis ao segundo dos límites.



## Capítulo 5

# Residuos de Li, Wu e Feng aplicados a modelos de fraxilidade

Li, Wu e Feng (2021) propuxeron un método para o diagnóstico dos modelos de regresión en análise de supervivencia clásico (sen fraxilidade) baseado na utilización de probabilidades de supervivencia aleatorizadas. Este método foi utilizado por Kats e Gorfine (2023) nun contexto específico de regresión para datos illness-death. Pero, ata o límite do noso coñecemento, non se consideraron estes residuos no diagnóstico dun modelo xeral de datos agrupados en clusters.

Neste capítulo preséntanse os residuos de Li, Wu e Feng, tamén chamados probabilidades de supervivencia aleatorizadas normalizadas ou NRSP polas súas siglas en ingles, e propóñense dous tests usando estes residuos, un de heteroxeneidade, e un de bondade de axuste do modelo de regresión, que ten dúas versións. Cada unha das versións usa unha adaptación diferente destes residuos aos modelos de fraxilidade:

1. por unha parte utilízase unha a adaptación inspirada pola proposta de Kats e Gorfine, e
2. por outra, unha opción alternativa que, segundo o noso coñecemento, é proposta por primeira vez neste traballo.

Posteriormente, explícase por que estes dous métodos son complementarios. Finalmente preséntanse un estudo de simulación para estudar a validez empírica dos residuos propostos e os resultados do mesmo.

### 5.1. Descrición do método

O método proposto por Li, Wu e Feng (2021) pode entenderse partindo dos tradicionais residuos de Cox-Snell. Cox e Snell propuxeron utilizar os riscos estimados  $r_i = \hat{\Lambda}(t_i | \mathbf{X}_{ij})$ ,  $i = 1, \dots, n$  como residuos. Como se víu en 1.11 se o modelo estimado é correcto e os valores de  $\hat{\beta}$  están próximos ás verdadeiras  $\beta$ , estes distribuiránse como unha mostra censurada dunha exponencial de parámetro 1,  $\exp(1)$ . Pola transformación integral da probabilidade temos que, se o modelo estimado fose correcto, a función de distribución resultante de aplicar a función de distribución suposta ás  $r$ , é dicir a distribución de  $1 - e^{-r}$ , ten distribución uniforme estándar censurada. Polo tanto, se aplicamos como transformación a función de supervivencia suposta  $e^{-r}$ , teremos tamén unha distribución uniforme. Definamos polo tanto estes valores como

$$SP(t) = \hat{S}(t|\mathbf{x}) = e^{-r}, \quad (5.1)$$

onde  $r$  é o residuo de Cox-Snell e  $SP$  vén de *survival probability*.

Os autores do artigo propoñen facer unha transformación relativamente sinxela das  $SP$  para lidiar co fenómeno da censura e obter residuos uniformes. As probabilidades de supervivencia aleatorizadas  $RSP$ , con distribución uniforme estándar, como veremos máis adiante, veñen dadas por

$$RSP(t, \delta, U) = U^{1-\delta} SP(t) \quad (5.2)$$

onde  $U$  se distribúe seguindo unha distribución uniforme en  $[0, 1]$ . Para a proba da uniformidade das  $RSP(t, \delta, U)$  con censura non informativa pola dereita, replícase a achegada polos autores.

### Proba de que os residuos RSP seguen unha distribución uniforme estándar

Para comezar reescíbese a fórmula das RSP como

$$RSP(T^*, C, U) = \begin{cases} S(T^*|\mathbf{X}), & \text{se } T^* \leq C \\ US(C|\mathbf{X}), & \text{se } T^* > C. \end{cases}$$

Para certa  $C = c$  temos que, se o tempo de fallo é menor que  $c$ ,  $T^* \leq c$  entón  $RSP(T^*, C, U) = S(T^*|\mathbf{x})$  que está uniformemente distribuída en  $U[S(c|\mathbf{x}), 1]$ . En cambio, se  $T^* > c$ ,  $RSP(T^*, C, U) = US(c|\mathbf{x})$  e está uniformemente distribuída en  $U[0, S(c|\mathbf{x})]$ . Con  $L(B)$  denotando a lonxitude do intervalo en  $[0, 1]$ , pódense derivar as seguintes igualdades:

$$\begin{aligned} & P(RSP(T^*, C, U) \in L(B)|C = c) \\ &= P(S(T^*|\mathbf{X}) \in B|C = c, T^* \leq c) \times P(T^* \leq c|\mathbf{X}) + P(US(T = c) \in B|C = c, T^* > c, \mathbf{X}) \times P(T^* > c|\mathbf{X}) \\ &= L(B \cap (S(T = c|\mathbf{X}), 1]) + L(B \cap [0, S(T = c|\mathbf{X})]) \\ &= L(B) \end{aligned}$$

Así queda demostrado que para  $C = c$ , a distribución das  $RSP(T^*, C, U) = RSP(T, \delta, U)$  é uniforme entre 0 e 1. Se aplicamos a regra da probabilidade total para extraer  $C$  obteremos de novo unha uniforme estándar.

As  $RSP(T, \delta, U)$  poderían usarse para comprobar se o modelo é correcto usando, por exemplo, tests de bondade de axuste respecto da uniforme estándar. Porén, os autores propoñen, en lugar de facer isto, normalizar as probabilidades de supervivencia aleatorizadas para obter os  $r^{NRSP}(T, \delta, U)$ ,

$$r^{NRSP}(T, \delta, U) = \Phi^{-1}(RSP(T, \delta, U)) \quad (5.3)$$

de xeito que as desviacións sexan máis evidentes e que poidamos usar toda a variedade de métodos de diagnóstico da estatística clásica para residuos normais, incluídos métodos gráficos. Así, por exemplo, un QQ-plot permitirá explorar a normalidade (necesaria para un bo axuste do modelo) e un gráfico dos residuos NRSP contra os valores dunha covariábel servirá para detectar posibles desviacións da linearidade.

#### 5.1.1. Test de heteroxeneidade

A primeira das aplicacións propostas é a utilización destes residuos para detectar a presenza de heteroxeneidade ao nivel do clúster e, de ser o caso, avaliar se esta provén da presenza de fraxilidade ao nivel do clúster ou doutro fenómeno, como por exemplo a presenza de datos atípicos. En certos casos podería mesmo informar sobre a forma a supór para as fraxilidades.

Supoñamos que temos unha mostra de datos de  $(\mathbf{T}, \delta)$ , e que existen unha ou varias variábeis agrupadoras que supoñemos poden ter certa influencia sobre a supervivencia, e que observamos ademais unha serie de covariábeis para cada individuo,  $\mathbf{X}$ .



Proponse axustar un modelo de supervivencia destes valores sen considerar a fraxilidade. Por exemplo, no caso de observar covariábeis e supór que o seu efecto ten a forma dun modelo de Cox, proponse estimar

$$\hat{\Lambda}(t|\mathbf{X}) = \hat{\Lambda}_0(t)e^{\hat{\beta}^t \mathbf{X}}$$

En base ao resultado da estimación obtéñense os residuos usando as ecuacións (5.1), (5.2) e (5.3). Usando estes residuos pódese contrastar a existencia de heteroxeneidade con métodos gráficos, como unha gráfica de caixas (e bigotes) sobre a variábel agrupadora, ou numéricos, como un test de igualdade de medias ANOVA. De non existir heteroxeneidade non deberíamos, na maioría dos casos, encontrar diferenzas significativas. En cambio, de existir heteroxeneidade, por exemplo unha variábel  $Z$  común a cada clúster, que ten un efecto multiplicativo sobre o risco, os clusters non provirán dunha mesma distribución, senón que os residuos serán

$$r_h^{NRSP}(T_{ij}, \delta_{ij}, U_{ij}, Z_i, \mathbf{X}_{ij}) = \Phi^{-1} \left( U_{ij}^{1-\delta_{ij}} \exp \left( -\frac{\hat{\Lambda}(T_{ij}|\mathbf{X}_{ij})}{Z_i} \right) \right) \quad (5.4)$$

cuxa media variará entre clústeres polo efecto de  $Z_i$ .

A posibilidade de usar métodos gráficos pode permitir detectar se a heteroxeneidade é, en efecto, o resultado da existencia de fraxilidade na poboación. Se non fose este o caso senón que se tratase de datos atípicos correspondendo unicamente a un clúster veríamos nun gráfico de caixas homoxeneidade entre a maioría das mesmas e unicamente unha desviándose significativamente.

Ademais, no caso de existir fraxilidade, un método gráfico pode indicar que a utilización dun modelo de fraxilidade con distribución discreta cun número reducido de puntos pode ser suficiente para axustar o modelo. Nese caso veríamos que a altura das caixas pode agruparse nun número reducido de grupos, que sería o número de puntos de discretización a utilizar. A distribución discreta das fraxilidades foi pouco tratada neste traballo, pero existe literatura ao respecto doutros autores como Wienke (2010)

Na figura 5.1 (esquerda) pode verse o resultado da aplicación do test proposto a unha mostra xerada aleatoriamente formada por 10 clusters de tamaño 80 cunha taxa de censura aproximada do 20%, logo de engadirlle un termo de fraxilidade que se distribúe de acordo cunha gamma de media 1 e varianza 0.5,  $\Gamma(1/0.5, 0.5)$ . Pode verse claramente que existe heteroxeneidade, tanto no resultado do test ANOVA, como no gráfico de caixas que, ademais, non indica que a heteroxeneidade sexa o resultado de valores atípicos nun clúster, senón que semella ser o resultado da fraxilidade subxacente.

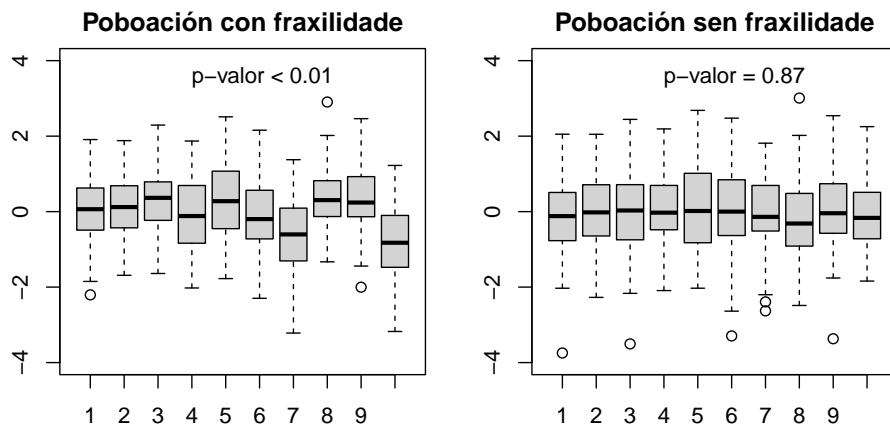


Figura 5.1: Gráfico de caixas e p-valor do ANOVA para mostras formadas por 10 clusters de tamaño 80, con taxa de censura 20% con e sen fraxilidade respectivamente

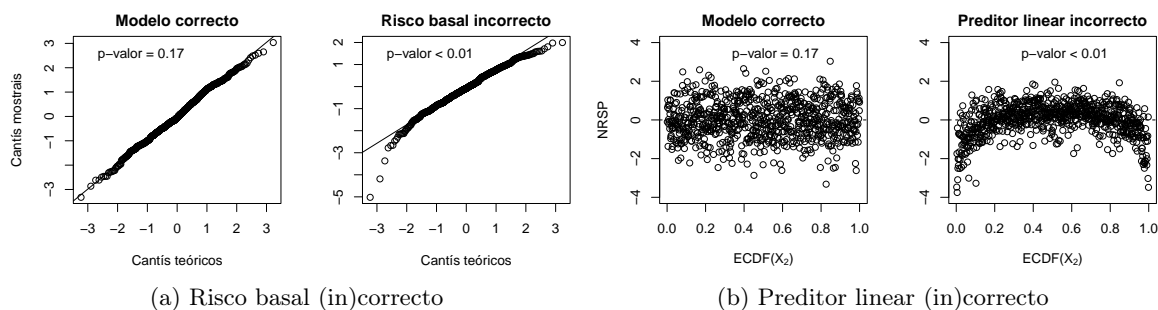


Figura 5.2: Representacións gráficas de mostrás simuladas de 10 clusters de tamaño 80 dun modelo correcto e dous modelos incorrectos

Este exemplo, así como os outros incluídos nesta sección, correspóndense coa primeira simulación dos respectivos escenarios no estudo de simulación presentado no seguinte apartado.

### 5.1.2. Test de bondade de axuste do modelo de regresión

Para analizar a bondade de axuste dun modelo de Cox con fraxilidade compartida en clústers, propóñense dous métodos de adaptar os residuos de Li, Wu e Feng ao contexto de datos en clusters.

Especificamente, propónse obter os residuos condicionais e os residuos marxinais, que como veremos, terán mellor ou peor comportamento segundo o tamaño de mostra e clusters que se consideren.

#### Residuos condicionais

O primeiro dos métodos propostos é, na miña opinión, a extensión evidente do uso dos residuos de Li, Wu e Feng para a análise da bondade de axuste aos modelos de fraxilidade. Dada a mostra  $\{(T_{ij}, \delta_{ij}, \mathbf{X}_{ij})\}$ ,  $i = 1, \dots, n, j = 1, \dots, n_i$ , axústase un modelo de regresión con fraxilidade. A falta de máis información sobre a forma da fraxilidade supónse que a súa distribución é gamma, dada a conveniencia matemática e robustez descritas anteriormente.

En base a esta estimación, os residuos para un modelo de Cox calcúlanse como

$$r_c^{NRSP}(T_{ij}, \delta_{ij}, U_{ij}, \mathbf{X}_{ij}) = \Phi^{-1} \left( U_{ij}^{1-\delta_{ij}} \exp \left( -\hat{\Lambda}_0(T_{ij}) e^{\hat{\beta}^t \mathbf{X}_{ij}} \hat{Z}_i \right) \right) \quad (5.5)$$

onde  $\hat{Z}_i$  é a predición da fraxilidade do clúster  $i$ .

Estes residuos poden usarse para analizar a bondade de axuste con métodos gráficos como un QQ-plot ou un gráfico dos residuos contra os valores observados das covariábeis, ou con métodos numéricos como o test de normalidade de Lilliefors ou de Shapiro e Wilk.

Na figura 5.2 pódense ver as gráficas, xunto ao p-valor saínte do test de Shapiro-Wilk, de mostrás simuladas de 10 clusters de tamaño 80, correctamente axustados (gráficas 1 e 3), e incorrectamente axustados por asumir que o risco basal é Gompertz en lugar de Weibull (gráfica 2) e por asumir que a  $X_2$  ten un efecto linear na probabilidade de supervivencia, cando este é en realidade cadrático (gráfica 4). Pode verse nos p-valores que o test de normalidade de Shapiro-Wilk con estes residuos é quen de detectar ámbalas dúas desviacións e que a análise gráfica dos residuos pode dar certos indicios da forma da desviación da linearidade.

#### Residuos marxinais

O segundo método segue a proposta de Kats e Gorfine (2023) que, en lugar de usar as estimacións da supervivencia condicionadas ás predicións das fraxilidades, usan a distribución marxinal, na que non

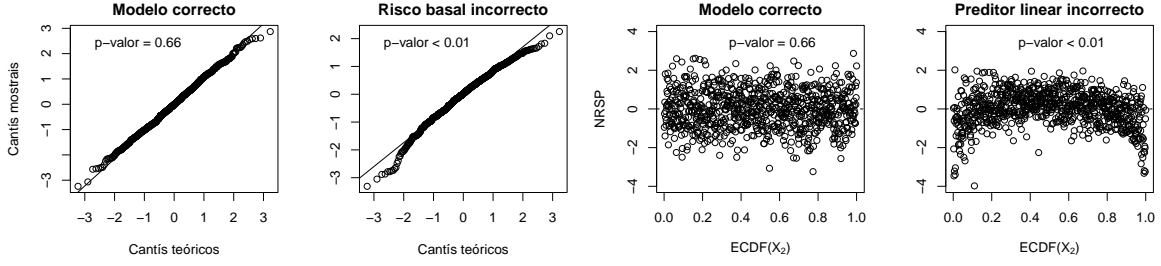


Figura 5.3: Representacións gráficas de mostrás simuladas de 80 clusters de tamaño 10 dun modelo correcto e dous modelos incorrectos

se considera a fraxilidade individual do clúster, senón que se integra respecto das fraxilidades por ser éstas descoñecidas.

A esperanza da supervivencia dun individuo respecto a  $Z$  é

$$\begin{aligned} \mathbb{E}_Z [S(t_{ij} | \mathbf{X}_{ij})] &= \int_0^\infty \exp\left(-\hat{\Lambda}_0(t_{ij}) e^{\hat{\beta}^t \mathbf{X}_{ij} z_i}\right) f(z_i) dz_i \\ &= \mathcal{L}\left(\hat{\Lambda}_0(t_{ij}) e^{\hat{\beta}^t \mathbf{X}_{ij}}\right) \end{aligned}$$

Cabe aclarar que pola dependencia entre as supervivencias marxinais dentro dun clúster, o produto destas supervivencias non é igual á función de supervivencia conxunta do clúster, e é por iso que non se pode usar á hora de estudar a verosimilitude das estimacións e se usa no seu lugar a función vista en (2.8).

Usando estes residuos, os residuos NRSP marxinais serían

$$r_m^{NRSP}(T_{ij}, \delta_{ij}, U_{ij}, \mathbf{X}_{ij}) = \Phi^{-1}\left(U_{ij}^{1-\delta_{ij}} \mathcal{L}\left(\hat{\Lambda}_0(T_{ij}) e^{\hat{\beta}^t \mathbf{X}_{ij}}\right)\right) \quad (5.6)$$

Estes residuos poderían analizarse coas mesmas ferramentas mencionadas anteriormente.

Na figura 5.3 pódense ver, de novo, as gráficas, xunto ao p-valor saínte do test de Shapiro-Wilk, neste caso de mostrás formadas por 80 clusters de tamaño 10, de modelos correctamente axustados (gráficas 1 e 3), e incorrectamente por risco basal incorrecto (gráfica 2) e por asumir efecto linear de  $X_2$  cando é cadrático (gráfica 4). De novo os resultados amosan o bo comportamento do test.

### Comparación dos métodos de diagnóstico

Destes dous métodos pódese dicir que son complementarios, porque é de esperar que o método dos residuos condicionados funcione ben cando o tamaño dos clusters é grande en relación ao tamaño da mostra e a varianza das fraxilidades é alta, mentres que o comportamento dos residuos marxinais empiora nestes casos, e viceversa para varianza da fraxilidade baixa e tamaño dos clusters pequeno.

A relación do funcionamento do método dos residuos marxinais co tamaño dos clusters e a varianza é intuitivo. A medida que aumenta o tamaño dos clusters en relación á poboación e a varianza da fraxilidade, unha maior parte da variabilidade dos tempos de evento, e con eles das probabilidades de supervivencia estimadas, dependen do valor da fraxilidade no clúster. Polo tanto, tomando o valor medio sobre estas e ignorando o seu efecto, introdúcese un nesgo nas supervivencias estimadas, que leva a unha falta de normalidade dos residuos.

O mal funcionamento do diagnóstico usando os residuos condicionais, pola súa parte, semella ser causa do nesgo na predición das fraxilidades, causada por unha excesiva concentración destas predicións

respecto da distribución real das fraxilidades, causada por non considerar a varianza na predición das fraxilidades.

Analícemos brevemente este nesgo. Como se ve na igualdade (2.12), a distribución da fraxilidade predita con distribución gamma en base ás estimacións do modelo axustado é á súa vez unha distribución gamma mais de parámetros  $1/\sigma^2 + D_i$  e  $1/(1/\sigma^2 + \sum_{j=1}^{n_i} \Lambda_{\text{clúster}}(t_{ij}|\mathbf{x}_{ij}))$  e a predición de cada unha das fraxilidades obtíñase como a media desta distribución, ignorando a varianza da mesma, o que resulta nunha excesiva concentración das predicións.

A varianza da predición da fraxilidade é

$$\begin{aligned} \text{Var}[\hat{Z}_i] &= \frac{1/\sigma^2 + D_i}{\left(1/\sigma^2 + \sum_{j=1}^{n_i} \Lambda(T_{ij}|\mathbf{X}_{ij})\right)^2} \\ &= \frac{1/\sigma^2 + D_i}{1/\sigma^4 + \sum_{j=1}^{n_i} \Lambda(T_{ij}|\mathbf{X}_{ij})^2 + 2\left(\sum_{j \neq k}^{n_i} \Lambda(T_{ik}|\mathbf{X}_{ik})\Lambda(T_{ij}|\mathbf{X}_{ij}) + 1/\sigma^2 \sum_{j=1}^{n_i} \Lambda(T_{ij}|\mathbf{X}_{ij})\right)}. \end{aligned}$$

Aínda que a esperanza desta varianza non se pode estimar sen definir a forma da función de risco, e mesmo neste caso sería difícil de estimar pola dependencia intra-clúster, si se pode intuír que, por ser a esperanza do risco acumulado nos tempos de evento igual 1, tende a 0 a medida que aumenta  $n_i$ . Co cal a medida que aumenta o tamaño do clúster reduciríase a variabilidade explicada pola varianza das predicións, e aumentaría a parte explicada pola diferenza nas esperanzas, de xeito que se reduce a excesiva concentración das predicións, e a súa distribución se achega á distribución real.

Pode verse tamén a través dunha simulación. Na figura 5.4 pode verse a través dunha simulación como a densidade das predicións das fraxilidades achégase á densidade real das fraxilidades,  $\Gamma(1/0.5, 0.5)$ , a medida que aumenta o tamaño de clúster para unha poboación formada por 5,000 clusters, con risco exponencial de parámetro 1 e sen censura. O risco acumulado non foi estimado, senón que é o risco real, que neste caso é igual ao tempo de evento. O código pode consultarse no apéndice B.1.

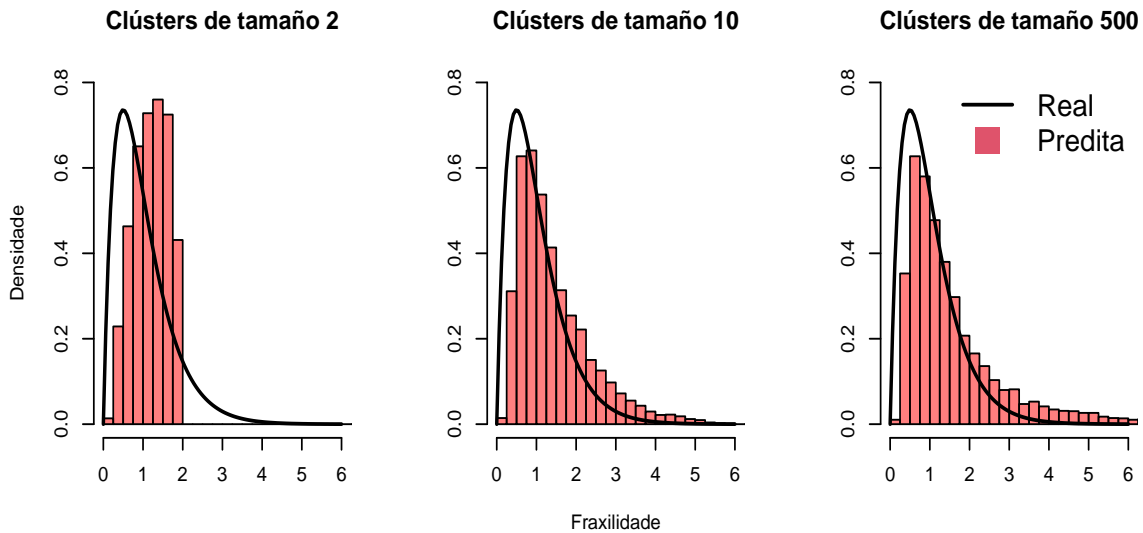


Figura 5.4: Comparación entre a distribución de fraxilidade e a predición das fraxilidades usando o valor real do risco para distintos tamaños de clúster

Non amoso esta diferenza no rendemento repetindo os tests e gráficas 5.2 e 5.3 xa que a diferenza no rendemento non é suficientemente grande para verse con poucas mostras. No seu lugar, estúdase esta diferenza no rendemento na seguinte sección a través dun estudo numérico, con 500 simulacións.

## 5.2. Estudo de simulación

### 5.2.1. Descrición do estudo

Para analizar o funcionamento do test de heteroxeneidade e dos métodos de diagnóstico a partir dos residuos condicionados e marxinais, formúlase o seguinte estudo de simulación baseado en Hirsch e Wienke (2012) e Wu, Feng e Li (2023). Simularáanse tempos de supervivencia con:

- risco basal Weibull de parámetros  $\lambda = 0.007$  e  $\alpha = 3$ ;
- 3 covariábeis,  $X_1, X_2, X_3$ , con distribución
  - uniforme estándar,  $U[0, 1]$ ,
  - normal estándar,  $N(0, 1)$ , e
  - Bernoulli de parámetro 0.25,  $Bernoulli(0.25)$ ;
- efectos fixos reais,  $\beta_1 = 1, \beta_2 = -1, \beta_3 = 0.5$ ;
- fraxilidade compartida,  $Z$ , con distribución gamma con media 1 e varianza 0.5,  $\Gamma(1/0.5, 0.5)$ ; e
- tempos de censura con distribución exponencial de parámetro escollido con simulación de Monte Carlo para obter un nivel de censura medio predefinido.

Isto leva a que o tempo de evento se poida simular como

$$T_{ij} = \left( \frac{-\log(U_{ij})}{Z_i \lambda \exp(\beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3})} \right)^{1/\alpha} \quad (5.7)$$

onde  $U_{ij}$  é un valor xerado aleatoriamente dunha distribución uniforme estándar.

Estas simulacións utilizaráanse para analizar as 3 aplicacións do método de Li, Wu e Feng anteriormente mencionadas.

#### Test de heteroxeneidade

Axústase un modelo de Cox paramétrico sen fraxilidade con distribución Weibull, obtéñense os residuos normalizados co método de Li, Wu e Feng e aplícase un test de igualdade de medias (ANOVA) respecto dos clusters. Esperamos neste caso obter unha alta porcentaxe de rexeitamentos.

Ademais, xéranse tempos  $t'$  dunha poboación homoxénea usando a expresión (5.7) sen incluír a fraxilidade,  $Z$ , e aplícaselles o mesmo test para comprobar se, baixo homoxeneidade, os p-valores seguen unha distribución uniforme estándar. Xunto a este test, tanto para a poboación heteroxénea como a homoxénea, usarase o test de heteroxeneidade de Commenges-Andersen, presentado na sección 4.2.1, usando a función `ca_test` do paquete **frailtyEM** sobre a regresión axustada usando `coxph` de **survival** a modo de comparación.

#### Test de bondade de axuste da distribución basal de risco

Axústase o modelo correcto, é dicir con fraxilidade e con risco basal Weibull, obtéñense os residuos propostos cos métodos condicional e marxinal, respectivamente, e úsase un test de Shapiro-Wilk de normalidade para amosar a distribución uniforme dos p-valores baixo a hipótese nula.

Axústase asimesmo un modelo supoñendo risco basal Gompertz e úsase o test de Shapiro-Wilk sobre os residuos para amosar a capacidade de detectar a escolla incorrecta da distribución basal de risco.

### Test de linearidade do efecto das covariábeis

Xéranse tempos con efecto non linear das covariábeis,  $t''$ , substituindo  $x_2$  en (5.7) por  $x_2^2$  e axústase o modelo sen considerar esta modificación. Sobre os residuos deste modelo úsase o test de Shapiro-Wilk para detectar a desviación de normalidade dos residuos. Non se estuda o caso baixo a hipótese nula, xa que é o mesmo que no test anterior.

En resumo, xeráanse mostras de 3 poboacións diferentes, a poboación base ou correcta, unha poboación homoxénea, e unha poboación con efecto cadrático de  $X_2$ . Ás poboacións homoxénea e base aplícaselles un test de heteroxeneidade (ANOVA), sobre a base aplícanse ademais dous tests de bondade de axuste, un co risco asumido correctamente, e outro incorrectamente, e sobre a cadrática utilízase o test de bondade de axuste. 5 tests diferentes, ademais do uso do test de Commenges-Andersen a modo de comparación.

Estes tests repetíranse sobre 500 mostras aleatorias, para obter conclusións empíricas sobre o funcionamento do test.

### Escenarios

Considéranse distintos escenarios. Respecto do tamaño da poboación e dos clusters consideraránse:

- 80 clusters de tamaño 10 - tamaño da mostra 800
- 40 clusters de tamaño 10 - tamaño da mostra 400
- 10 clusters de tamaño 10 - tamaño da mostra 100
- 10 clusters de tamaño 80 - tamaño da mostra 800
- 10 clusters de tamaño 40 - tamaño da mostra 400

Respecto do nivel de censura considerarásese un nivel de censura medio do 20 %, do 50 % e do 80 %. En total considéranse 15 escenarios diferentes.

O código utilizado para implementar este estudo en R pode consultarse no apéndice B.2.

## 5.2.2. Resultados

### Test de heteroxeneidade

Os resultados do estudo de simulación sobre o funcionamento do test de heteroxeneidade proposto poden consultarse na táboa 5.1, que amosa nas columnas 1 e 3 a porcentaxe de rexeitamentos para un nivel de significación respectivamente de 1 %, 5 % e 10 % para unha mostra homoxénea e para unha mostra heteroxénea, xerada engadindo un termo de fraxilidade multiplicativo á mostra homoxénea. Alén diso, as mostras analizadas son iguais. Os resultados amósanse para 3 niveis de censura e 5 composicións distintas da mostra. Ademais nas columnas 2 e 4 amósanse, a modo de comparación, a porcentaxe de rexeitamentos para o test de Commenges-Andersen aplicado usando a función `ca.test` do paquete **frailtyEM**.

Pode observarse que o comportamento do test proposto baixo a hipótese nula, onde se esperaría un porcentaxe de rexeitamentos similar ao nivel de significación é bastante bo e que non semella empiorar o comportamento do test de Commenges-Andersen.

Respecto do comportamento baixo a hipótese alternativa, si é evidente que existe certa perda de potencia. É dicir, rexeita a hipótese nula nun menor número de casos malia ser a hipótese nula incorrecta.

Esta perda de potencia é entre nula e mínima, segundo o tamaño da mostra, nas situacións con censura baixa, e entre mínima e moderada para o caso con censura moderada, alén do caso co tamaño de mostra pequeno, onde a perda de potencia si é importante. No caso con censura alta, si é claro que a perda de potencia en comparación co test de Commenges-Andersen é moi considerábel.

			Distribución homoxénea						Distribución heteroxénea					
			Li, Wu e Feng			Commenges-Andersen			Li, Wu e Feng			Commenges-Andersen		
cens.	$n$	$n_i \setminus \alpha$	1 %	5 %	10 %	1 %	5 %	10 %	1 %	5 %	10 %	1 %	5 %	10 %
20 %	10	10	0.8	3.5	8.6	1.2	6.0	12.0	74.3	83.9	88.2	77.0	85.0	89.8
	10	40	1.6	7.0	10.6	1.2	6.0	11.6	99.6	99.8	100	99.6	100	100
	10	80	<b>1.2</b>	<b>6.2</b>	<b>9.8</b>	<b>1.6</b>	<b>6.4</b>	<b>11.6</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
	40	10	1.0	4.6	11.0	0.4	5.6	10.8	100	100	100	100	100	100
	80	10	0.6	6.0	12.4	0.8	5.2	9.8	100	100	100	100	100	100
50 %	10	10	1.7	6.1	10.1	2.0	9.0	15.4	29.1	50.0	62.1	49.8	67.2	73.6
	10	40	1.2	4.8	9.0	0.8	4.2	10.6	92.8	97.2	98.6	97.8	99.2	99.6
	10	80	1.4	4.4	9.8	1.4	5.6	10.8	99.4	99.6	99.8	100	100	100
	40	10	1.4	6.6	10.6	1.0	6.0	10.4	86.8	95.2	96.6	99.0	99.6	100
	80	10	0.8	5	11.2	1.2	5.2	11.8	98.8	99.4	100	100	100	100
80 %	10	10	0.8	5.7	11.5	1.8	6.2	11.2	3.3	12.6	17.7	13.2	26.8	38.0
	10	40	1.4	4.9	9.1	0.6	5.6	11.2	26.3	46.4	62.7	84.2	92.0	94.0
	10	80	0.4	5.0	9.8	1.0	4.2	11.6	64.8	79.8	85.6	97.6	99.4	99.6
	40	10	0.8	4.6	11.3	1.0	5.8	11.0	10.8	26.8	37.4	57.6	74.4	81.8
	80	10	0.6	5.2	9.0	1.0	7.8	13.2	15.6	40.4	53.8	91.6	98.2	99.0

Táboa 5.1: Porcentaxe de rexeitamentos para cada test e escenario para distintos niveis de significación. **En negra o escenario co mellor comportamento**

É importante comentar que a perda de potencia nas situacións con maior censura pode ser explicada por unha maior proporción de axustes que non converxeron para maiores taxas de censura. Cabe pensar que a falta de converxencia é maior para aqueles modelos que distan máis do axuste proposto, levando a que non se consiga o axuste e que non se consideren no test, de xeito que aqueles que serían rexeitados están representados de xeito máis que proporcional.

Esta perda de potencia en comparación co test de Commenges-Andersen non é un defecto excesivamente importante se se ten en conta que non se propón usar esta aplicación dos residuos de Li, Wu e Feng como substituto ao test de Commenges-Andersen senón como un test complementario, que permita, por exemplo, estudar a natureza da heteroxeneidade, usando, como se comentou anteriormente, os resultados de tests numéricos e/ou gráficos para ver que clusters son significativamente diferentes e detectar, por exemplo, que a heteroxeneidade veña dada polos valores atípicos nun único clúster. Ademais

aínda que aquí se utiliza unicamente o test ANOVA, cabe pensar que existen unha variedade de tests baseados na distribución uniforme, normal ou calquera outra distribución con función de distribución inversa que se poderían utilizar, permitindo estudar distintas características da heteroxeneidade entre os clusters.

Ademais, pode verse na táboa sobre a coincidencia no resultado do test incluída no apéndice D.2 que a taxa de coincidencia nas decisións, nos casos nos que a taxa de rexeitamento é similar entre os tests, é moi alta. Polo tanto, propónse usar o test de Commenges-Andersen inicialmente, que está implementado en R e ten maior potencia, para detectar a presenza de heteroxeneidade e, no caso de resultar esta significativa, utilizar o test proposto neste traballo como unha ferramenta para o estudo da natureza desta heteroxeneidade.

### Tests de bondade de axuste do modelo de regresión

Os resultados, en termos de porcentaxe de rexeitamentos para distintos niveis de  $\alpha$ , para os distintos escenarios considerados dos tests utilizando tanto os residuos marxinais, como os residuos condicionais poden verse na táboa 5.2.

Respecto das primeiras columnas, que amosan o comportamento dos tests baixo a hipótese nula, é difícil dicir se os tests amosan o comportamento esperado, é dicir unha porcentaxe de rexeitamentos en liña co nivel de significación, dado que, como se explicou anteriormente non se espera un comportamento axeitado de cada un dos tests en distintos escenarios, pero tampouco sabemos con claridade en que escenarios si se debería esperar un comportamento correcto dos tests. Segundo o noso parecer, si se ve un comportamento relativamente bo dos tests condicionais cando a relación entre o tamaño do clúster e o número de clusters é favorábel ao primeiro destes. Pode dicirse o mesmo dos tests marxinais cando a relación entre os números e tamaño dos clusters se invirte. En ámbolos dous casos pódese ver que son algo *liberais* en tanto que rexeitan a hipótese nula unha proporción excesiva de veces para unha taxa de censura moderada e vólvese excesivamente conservador a medida que aumenta a taxa de censura. Por último, no escenario menos conveniente para cada test o número de rexeitamentos é excesivamente alto.

Prestando atención á capacidade de rexeitar a hipótese nula cando esta non é certa, ou potencia do test, estes dous tests teñen un rendemento claramente peor ao test de heteroxeneidade estudado anteriormente. Porén, o test segue tendo unha potencia entre moderada e moi boa para tamaños de mostra grandes, e censura baixa ou moderada, especialmente á hora de detectar a desviación polo efecto mal especificado dunha covariábel, para tamaños de mostra baixos o test perde moita potencia, levando a que só nunha minoría dos casos se detecte a desviación, mesmo cando se usa o test máis axeitado para o número e tamaño de clusters na mostra.

Aínda que é difícil facer unha comparación directa, dadas as diferenzas no estudo de simulación, todo parece indicar que este método ten mellor rendemento que o proposto por González-Manteiga et al. (2023). Poñendo o foco na hipótese alternativa de efecto cadrático dunha das covariábeis para tamaño de mostra 800, os tests propostos neste traballo rexeitan a hipótese nula un 100% das veces para  $n_i = 80$  e un 96.2% das veces para  $n_i = 10$ , con  $\alpha = 0.05$ , usando en cada caso o test máis axeitado, mentres que o de González Manteiga et al. só acada entre 41.2% e 78.4% para un tamaño de mostra maior cando estudan o modelo baixo esta hipótese alternativa.

Concentrándonos agora no caso con taxas de censura alta ámbolos dous tests amosan moi pouca potencia. Independentemente do tamaño e composición da mostra o test para detectar as desviacións por efecto non linear dunha das covariábeis na metade de casos non é quen de detectar a falta de normalidade dos residuos. É interesante ver que este test co menor tamaño de mostra leva a un número significativamente menor de rexeitamentos que baixo a hipótese nula, o que indicaría que non se debería usar para detectar esta clase de desviación nesta situación. No caso da desviación causada por un risco basal mal especificado, sen ser un test extraordinariamente bo, si que se comporta mellor.



			Modelo correcto						Risco basal incorrecto						Efecto das covariábeis incorrecto					
			Condicionais			Marxinais			Condicionais			Marxinais			Condicionais			Marxinais		
cens.	$n$	$n_i \setminus \alpha$	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
20%	10	10	0.8	6.5	13.6	1.3	6.1	10.1	10.2	25.8	36.8	32.8	54.0	65.6	39.8	53.8	63.6	42.2	60.8	68.6
	10	40	1.6	7.2	14.4	1.2	5.6	11.6	48.2	69.6	80.8	69.8	84.8	89.2	78.8	91.4	93.8	62.6	78.6	84.0
	10	80	<b>2.2</b>	<b>6.4</b>	<b>9.8</b>	8.4	21	33.2	<b>100</b>	<b>100</b>	<b>100</b>	99.4	100	100	<b>100</b>	<b>100</b>	<b>100</b>	100	100	100
	40	10	4.2	14.4	23.8	2	7.2	11.8	56.0	75.2	83.8	96.6	99.0	100	99.8	99.8	99.8	99	99.6	99.8
	80	10	10.2	26.6	38.4	<b>2.6</b>	<b>6.8</b>	<b>12</b>	90.4	96.4	97.8	<b>100</b>	<b>100</b>	<b>100</b>	100	100	100	<b>100</b>	<b>100</b>	<b>100</b>
50%	10	10	1.7	7.3	13.0	1.3	5.0	9.1	2.6	11.8	19.4	11.8	25.7	33.9	4.9	12.5	18.4	6.6	16.4	28.3
	10	40	1.6	7.2	14.4	1.2	5.6	11.6	48.2	69.6	80.8	69.8	84.2	89.2	78.8	91.4	93.8	62.6	78.6	84.0
	10	80	1.2	5.8	11.6	2.2	7.2	13.8	93.6	97.2	98.6	97.4	99.4	99.8	99.4	100	100	94.2	97.4	98.6
	40	10	3.0	9.2	15.4	0.4	4.2	9.6	20.8	40.0	52.4	74.2	87.6	93.0	34.4	56.2	65.4	47.2	66.2	75.2
	80	10	6.2	13.6	20.8	1.2	6.6	11.4	51.0	70.2	78.2	97.6	99.6	100	73.0	85.4	91.8	85.4	96.2	97.4
80%	10	10	0.8	4.3	8.1	1.3	4.1	8.7	1.0	6.6	12.4	2.2	9.6	17.9	0.5	3.5	6.8	0.3	2.4	7.6
	10	40	0.8	3.3	8.6	0.2	3.3	7.4	7.8	23.2	35.0	12.6	32.8	46.6	4.5	10.8	17.1	2.9	12.6	18.9
	10	80	2.2	6.2	10.5	0.6	5.6	8.9	30.0	53.6	64.8	45.2	70.4	81.4	11.6	27.1	38.7	6.4	19.2	29.7
	40	10	0.6	4.9	8.7	0.2	2.6	7.9	7.6	19.7	30.0	14.3	31.8	43.4	1.0	5.7	11.0	1.8	6.7	12.2
	80	10	0.6	7.0	11.0	0.4	3.4	7.6	15.2	35.2	47.8	41.0	65.0	75.2	3.0	12.0	19.3	2.2	12.0	19.7

Táboa 5.2: Porcentaxe de rexeitamentos en cada escenario cos residuos condicionais e marxinais para distintos niveis de significación.

### En negraña os escenarios co mellor comportamento

En situacións con tamaños de mostra grandes ten unha taxa de rexeitamento relativamente alta e a proporción de rexeitamentos en ningún caso é claramente inferior á observada baixo a hipótese nula.

En canto á redución da porcentaxe de rexeitamentos con taxa de censura alta, é probable que teña relación cunha maior proporción de axustes que non converxeron, levando a que non se aplique o test sobre esta mostra. Na táboa incluída no apéndice D.1 vese que esta proporción medra alí onde o nivel de rexeitamentos decrece de xeito non esperado. É plausíbel imaxinar que exista unha relación inversa entre a probabilidade de converxencia e a distancia da mostra respecto da hipótese nula, levando ás mostras que máis distan da nula a estar infrarrepresentadas entre os resultados dos tests, e a unha menor taxa de rexeitamentos. Porén, a falta de converxencia non é a única explicación, xa que a proporción de axustes que non converxen non é o suficientemente alta para explicar o mal comportamento por si sóa. Este problema non é imputable ao test proposto, senón que ten que ver coa rutina escollida para a estimación do modelo de regresión.

En canto á comparación entre os dous métodos vemos que, polo xeral, o comportamento segue o patrón esperado e descrito na sección 5.1.2. Alén diso, é interesante ver que cada un dos tests semella comportarse lixeiramente mellor para detectar respectivamente unha das desviacións estudadas, cos residuos marxinais tendo mellor rendemento á hora de detectar desviacións por risco basal mal espe-

cificado. Por pór un exemplo, vemos que a taxa de rexeitamentos no escenario máis favorábel para o test condicional ( $n = 10, n_i = 80$ ) á hora de detectar que o risco basal non é Weibull con censura 50 % é do 93.6 %, por debaixo da observada no mellor escenario para o test marxinal ( $n = 10, n_i = 80$ ) de 97.6 %, mentres que o contrario pasa no escenario con efecto non linear de  $x_2$ ; 99.4 % contra 85.4 %. Porén a relevancia deste fenómeno non pasa da observación, xa que polo xeral non é posíbel escoller o tamaño dos clusters.

O interese está en realidade en saber que test tería mellor rendemento segundo a composición da mostra. Chegar a unha conclusión sobre isto precisaría dun estudo sobre o nesgo introducido por cada un dos métodos en relación á composición e tamaño da mostra e á taxa de censura. Como única indicación, comentar que o test marxinal mellora o rendemento do test condicional nalgunhas situacións que lle son desfavorábeis, o que non pasa na dirección contraria.

Se prestamos atención á taxa de coincidencia entre o resultado do test marxinal e condicional que se pode consultar no apéndice D.2 vemos que mesmo cando as taxas de rexeitamento son similares entre os dous tests, a taxa de coincidencia entre ambos tests non é necesariamente alta. É dicir, os dous tests levan a rexeitamentos para diferentes mostras. Polo tanto, é moi importante saber que test utilizar en cada caso, xa que os tests a miúdo levan a decisións opostas.

### 5.3. Aplicación a datos

Para amosar o funcionamento dos métodos de diagnóstico propostos, aplícanse ao conxunto de datos estudado no apartado 2.3, referido ao tempo ata a perda da visión en cada ollo de persoas que sufren de retinopatía diabética, onde os datos se agrupan ao nivel do individuo, o que resulta en clusters de tamaño 2.

#### 5.3.1. Test de heteroxeneidade

Comezamos por aplicar o test de heteroxeneidade. A falta de información sobre a distribución dos tempos de evento, supúxose que seguen unha distribución Weibull pola súa flexibilidade e por ser unha das distribucións máis utilizada na práctica. O modelo axustouse coa función `parfm` do paquete homónimo. Como se propón para este test, axustouse sen fraxilidade o modelo usando como covariábeis o nivel de risco de perda de visión do individuo e unha variábel binaria indicando se foi tratado, como se propuxo anteriormente. O código, coa semente precisa para a aleatorización dos residuos, pode consultarse no apéndice B.3.1. O p-valor resultante é 0.007, levando a un rexeitamento claro da homoxeneidade. Pódese comentar que o resultado é moito maior ao p-valor visto anteriormente para o test de Commenges-Andersen de 0.0003.

A modo de comparación, xa que durante o estudo de simulación non se estudou ningún caso con tamaño de clúster tan reducido, e non sabemos moito sobre o seu funcionamento, repetín o exemplo sobre unha mostra bootstrap, na que se fixo unha remostraxe da variábel indicando o individuo. O obxectivo deste exercicio era comprobar se o test resultaría, como se espera, nunha aceptación da hipótese nula neste escenario, no que os clusters agrupan aos individuos aleatoriamente e, polo tanto, non gardan relación cos seus tempos de supervivencia. O p-valor usando esta mostra bootstrap é de 0.983, claramente en concordancia co esperado.

Unha gráfica de caixas dos residuos aleatorizados normalizados respecto do clúster pode verse na figura 5.5. Nesta gráfica pódese observar claramente a diferenza entre a distribución da varianza baixo a mostra orixinal e baixo a remostra bootstrap, coa primeira gráfica amosando que a maioría da varianza se encontra na diferenza entre os clusters, mentres que na segunda a varianza vén principalmente dada pola varianza intra-clúster.

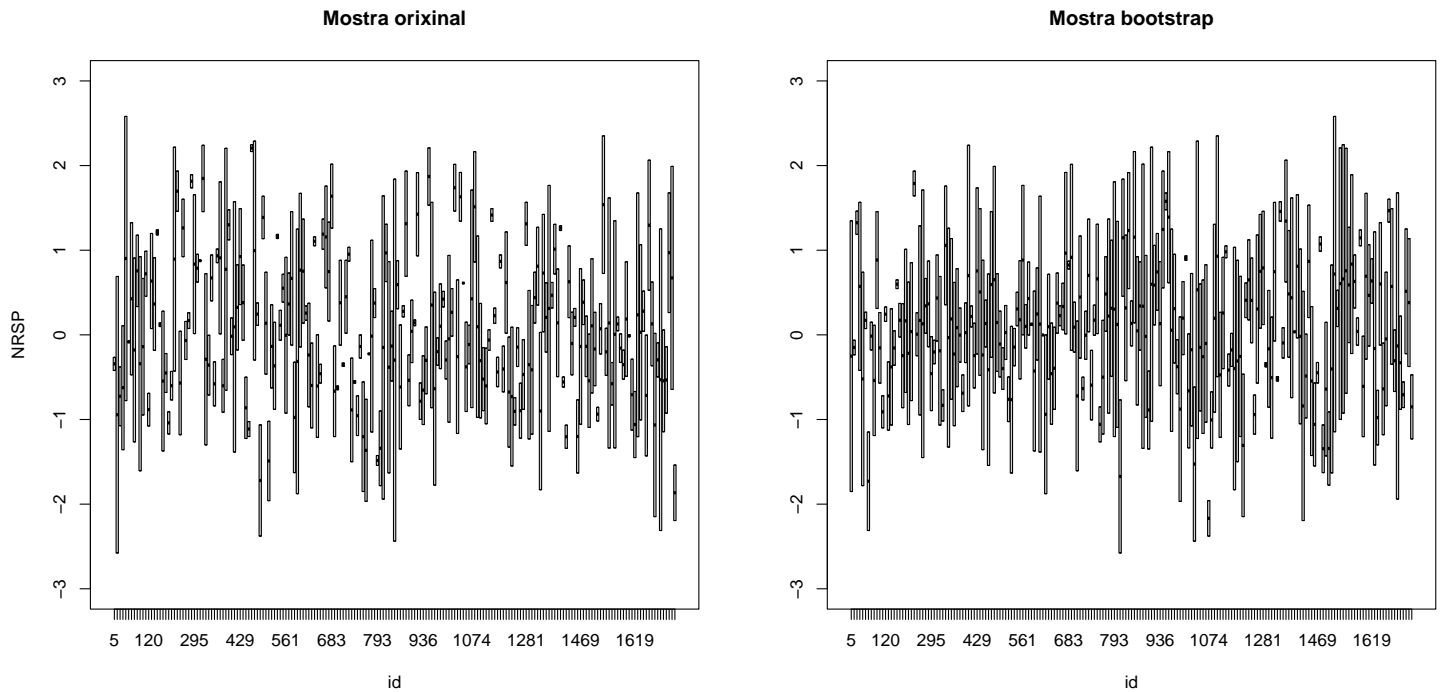


Figura 5.5: Gráfico de caixas dos NRSP respecto do individuo para a mostra orixinal e unha mostra con clusters formados aleatoriamente

### 5.3.2. Test de bondade de axuste

Para estudar a bondade de axuste do modelo de regresión de fraxilidade, axustouse usando de novo a distribución Weibull para os tempos de evento, mentres que se usou a distribución gamma para as fraxilidades. O axuste fíxose de novo usando a función `parfm`, como se pode ver no apéndice B.3.2.

Os residuos aleatorizados normalizados otivéronse utilizando a modo de comparación tanto os residuos marxinais, coma os residuos condicionais. Neste exemplo claramente os residuos marxinais son os máis axeitados, dado o baixo tamaño da mostra e o valor relativamente baixo da varianza das fraxilidades, que se pode ver na sección 2.3, e os resultados vistos no estudo de simulación. O test utilizado para comprobar a normalidade dos residuos é, igualmente que no estudo de simulación, o test de Shapiro-Wilk.

O p-valor do test usando os residuos marxinais é 0.191, levando a aceptar a hipótese nula de que o modelo de regresión é correcto, mentres que o nesgo introducido polos residuos condicionais cun tamaño de mostra tan pequeno, leva a un p-valor de 0.001, que levaría a rexeitar claramente o test.

A normalidade dos residuos pode observarse nas gráficas da parte superior da figura 5.6, aínda que semellan indicar certa desviación. É posíbel que para un tamaño maior da mostra se rexeitase a normalidade, e fose recomendábel o uso doutra distribución dos tempos de evento. Pasando aos preditores lineares, pode verse nas gráficas da parte inferior da mesma figura que non hai unha tendencia identificábel dos residuos.

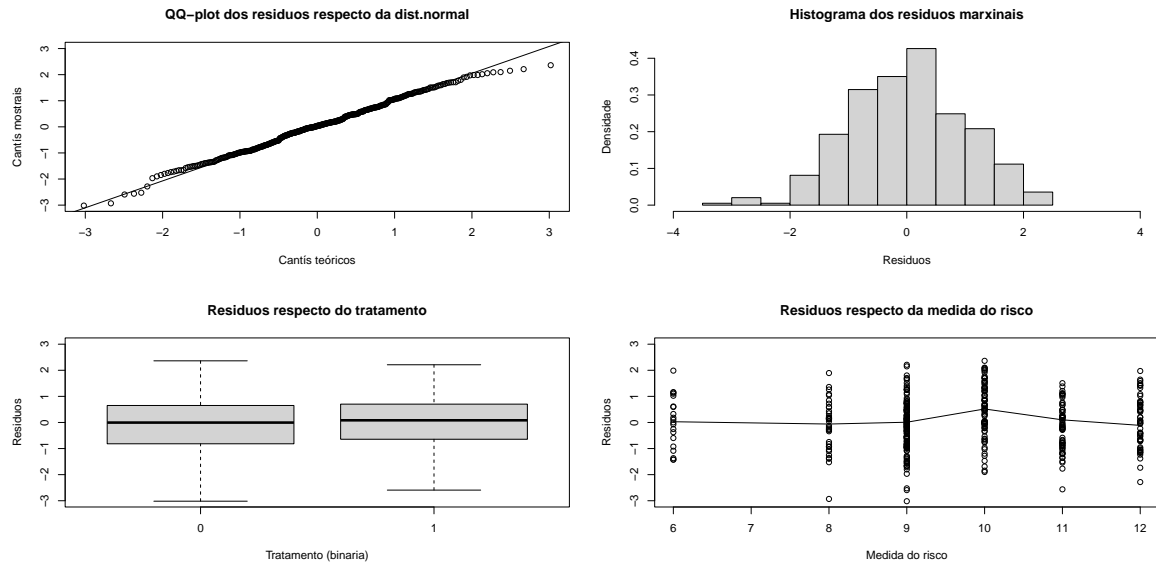


Figura 5.6: Gráficas de diagnóstico do axuste con fraxilidade dos datos sobre a perda da visión cos pacientes con diabetes retinopática baseados nos residuos marxinais

### 5.3.3. Modelo de regresión con interaccións

Ata agora viñamos usando un modelo sen interaccións porque a súa interpretación nos parecía máis sinxela. Porén, estudando un conxunto de datos tan popular, pensamos que sería interesante estudalo tamén o axuste máis popular, que é o que recolle Moore (2016) no seu manual de análise de supervivencia, que considera unha interacción entre o feito de recibir un tratamento e o grupo de idade no que se encontra o paciente (xuvénil ou adulto). Ademais, pensamos que o feito de considerar interaccións expandía a demostración da validez destes métodos de diagnóstico.

O código co que se axustou este modelo pode consultarse no apéndice B.3.3. Os resultados son de novo bastante satisfactorios, xa que, considerando neste caso unicamente os residuos marxinais, o test de heteroxeneidade resultou significativo, cun p-valor menor a 0.005, e o test de normalidade dos residuos acepta a validez do modelo cun p-valor de 0.33. Nas figuras 5.7 e 5.8 poden consultarse os resultados gráficos deste axuste, que de novo teñen un aspecto consistente coa hipótese de validez do modelo.

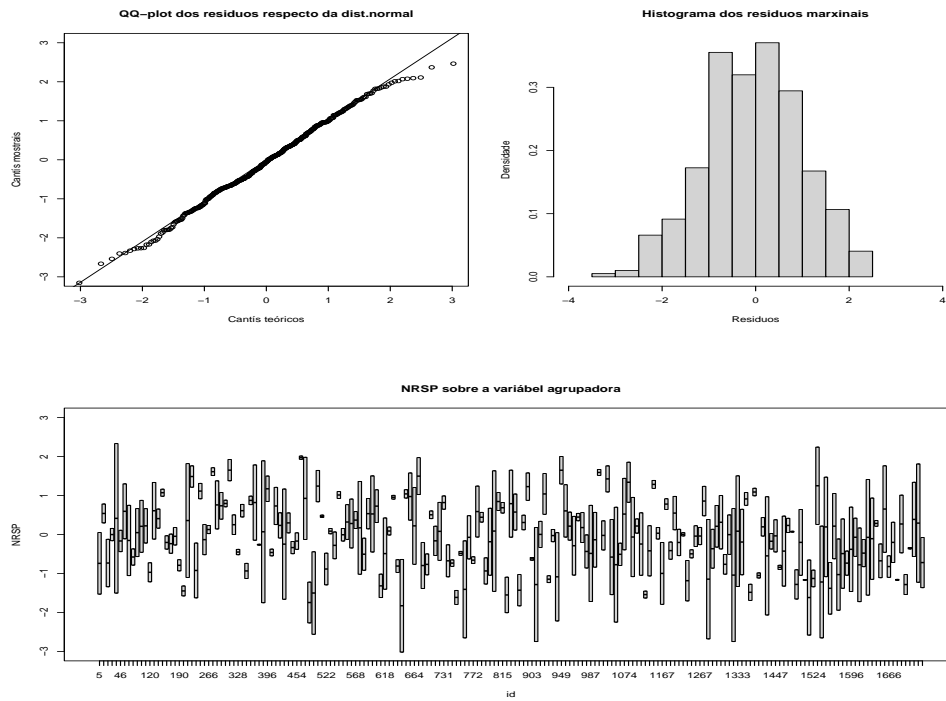


Figura 5.7: Gráficas de diagnóstico do axuste con fraxilidade e interaccións dos datos sobre a perda da visión cos pacientes con diabetes retinopática baseados nos residuos marxinais

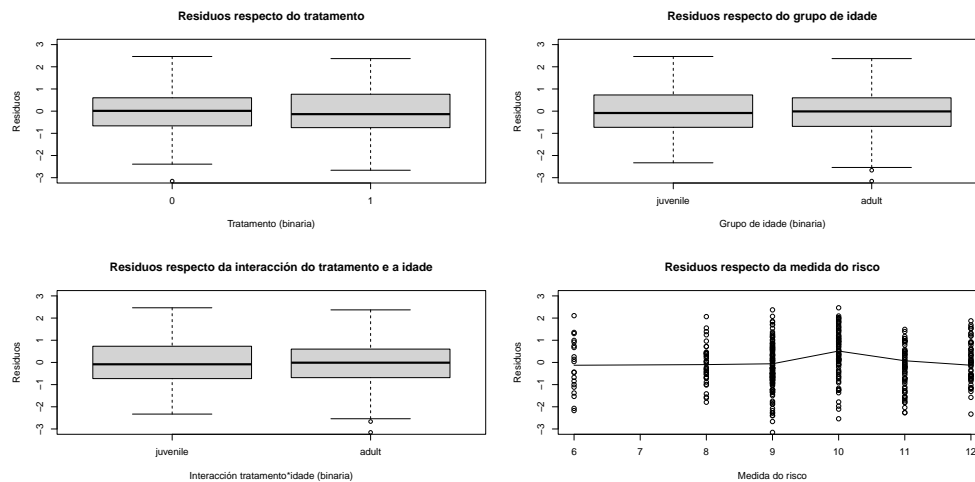


Figura 5.8: Gráficas de diagnóstico do axuste con fraxilidade e interaccións respecto das covariábeis dos datos sobre a perda da visión cos pacientes con diabetes retinopática baseados nos residuos marxinais



## Capítulo 6

# Conclusións

Ao longo deste traballo tratáronse extensamente os modelos de fragilidade compartida no ámbito dos modelos de regresión de riscos proporcionais de Cox. Púidose ver que existe un corpus teórico sólido respecto das propiedades destes modelos, que inclúe unha variedade de distribucións axeitadas para o seu estudo que permiten axustar distintos modelos de dependencia. Víuse asimesmo que estes modelos son utilizables na práctica xa que contan cunha importante variedade de métodos e implementacións para o seu estudo. A continuación percibimos que o número de métodos de diagnóstico e da implementación dos mesmos en R é claramente menor, cunha especial falta de interese nos tests de bondade de axuste dos modelos de regresión con fragilidade.

En resposta a esta carencia, propuxéronse unha serie de tests para a detección de heteroxeneidade nunha mostra e o estudo da bondade de axuste dun modelo de fragilidade baseados nun método recentemente proposto para a análise do axuste de datos con censura. A falta dunha análise máis rigurosa, déronse algunhas pinceladas da validez teórica destes métodos. A validez empírica dos mesmos estudouse a través dun estudo de simulación, cuxos resultados se amosaron neste traballo. Os resultados deste estudo dan validez ao test de heteroxeneidade proposto, cunha perda de potencia limitada respecto do establecido test de Commenges-Andersen, á par que achega ferramentas adicionais.

Os resultados sobre os tests de bondade de axuste, malia amosar problemas menores no comportamento baixo a hipótese nula, demostraron que a potencia do test podería ser considerabelmente maior á doutros tests recentemente propostos. En cambio, presentan unha desvantaxe en tanto que están baseados nun axuste paramétrico, o que limita a súa aplicabilidade. Aínda que nada indique que o test non funcione cun axuste semiparamétrico, este traballo non aporta evidencia empírica da súa validez neste escenario. A existencia de dúas versións destes tests que permiten adaptalo a distintas situacións de tamaño de clúster e obter resultados correctos é unha fortaleza deste test que trae consigo unha limitación en tanto que non é fácil escoller entre ambas opcións en situacións intermedias, o que pode levar a unha decisión incorrecta e un resultado nesgado.

Unha aplicación do test de heteroxeneidade a datos reais sobre o tempo ata a perda da visión de pacientes con diabetes retinopática semella amosar o bo funcionamento deste método, amosando claramente a presenza de heteroxeneidade na mostra. O test de bondade de axuste tamén amosa que o modelo é correcto.

Como limitación a este traballo podería apuntarse á falta de inclusión de outros modelos de fragilidade como o modelo de fragilidade univariante, onde se asigna un termo de fragilidade a cada individuo, o modelo de fragilidade correlada, onde as fragilidades nos clusters están correladas pero non son idénticas, tratado extensivamente por Wienke (2010), o modelo de tempo de fallo acelerado con fragilidade ou os modelos de riscos (semi-)competitivos con fragilidade tratados por Gorfine e Zucker (2023).

Como fortaleza do traballo destacar que se fixo un estudo cun enfoque máis específico que permitiu analizar en maior detalle as características dos modelos de fragilidade compartida e, a través deste estudo, identificar e abordar a falta de literatura en certas áreas de investigación.

De cara a continuar a investigación, podería estudarse a validez do test proposto á hora de detectar outras hipóteses alternativas (risco basal exponencial, efecto logístico das covariábeis, ...), ou cando se usa con outros modelos de fragilidade, como o modelo de riscos proporcionais semiparamétrico ou o modelo de tempo de fallo acelerado. Asimesmo sería interesante estudar se este test podería detectar a incorrecta especificación da distribución de fragilidade para tamaños de mostra suficientemente grandes, o que non se estudou neste traballo xa que resultados preliminares indicaron que non era o caso para os tamaños de mostra considerados. Por último, sería útil estudar en maior profundidade as propiedades teóricas destes tests, en concreto respecto do nesgo introducido por cada método do test (residuos marxinais ou condicionais) e en que contexto sería mellor utilizar cada un.



## Apéndice A

### Distribucións relevantes

#### A.1. Distribucións dos tempos de evento comunmente utilizadas

	<b>Exponencial</b>	<b>Weibull</b>	<b>Gompertz</b>
$\lambda(t)$	$\lambda \quad (\lambda > 0)$	$\alpha\lambda t^{\alpha-1} \quad (\lambda, \alpha > 0)$	$\lambda e^{\varphi t} \quad (\lambda, \varphi > 0)$
$\Lambda(t)$	$\lambda t$	$\lambda t^\alpha$	$\frac{\lambda(e^{\varphi t}-1)}{\varphi}$
$f(t)$	$\lambda e^{-\lambda t}$	$\alpha\lambda t^{\alpha-1} e^{-\lambda t^\alpha}$	$\lambda e^{\varphi t} e^{-\frac{\lambda(e^{\varphi t}-1)}{\varphi}}$
$F(t)$	$1 - e^{-\lambda t}$	$1 - e^{-\lambda t^\alpha}$	$1 - \exp\left(-\frac{\lambda(e^{\varphi t}-1)}{\varphi}\right)$
$S(t)$	$e^{-\lambda t}$	$e^{-\lambda t^\alpha}$	$\exp\left(-\frac{\lambda(e^{\varphi t}-1)}{\varphi}\right)$
$\mathbb{E}[T]$	$\frac{1}{\lambda}$	$\lambda^{-\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right)$	$e^{m+s^2/2}$
$\text{Var}[T]$	$\frac{1}{\lambda^2}$	$\lambda^{-\frac{2}{\alpha}} \left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2\right)$	$e^{2m+s^2} (e^{s^2} - 1)$
	<b>Log-normal</b>	<b>Log-loxística</b>	
$\lambda(t)$	$\frac{(1/(st))\phi((\log(t)-m)/s)}{1-\Phi((\log(t)-m)/s)} \quad (s > 0)$	$\frac{\nu\kappa(vt)^{\kappa-1}}{1+(vt)^\kappa} \quad (v, \kappa > 0)$	
$\Lambda(t)$		$\log(1 + (vt)^\kappa)$	
$f(t)$	$\frac{1}{\sqrt{2\pi}st} e^{-(\log(t)-m)^2/(2s^2)}$	$\frac{\nu\kappa(vt)^{\kappa-1}}{(1+(vt)^\kappa)^2}$	
$F(t)$	$\Phi\left(\frac{\log(t)-m}{s}\right)$	$1 - \frac{1}{1+(vt)^\kappa}$	
$S(t)$	$1 - \Phi\left(\frac{\log(t)-m}{s}\right)$	$\frac{1}{1+(vt)^\kappa}$	

## A.2. Distribuciones das fraxilidades comunmente utilizadas

	Distribución gamma	Distribución gaussiana inversa
Notación	$\Gamma(\theta, k)$	$IG(\mu, \alpha)$
$f(z)$	$\frac{1}{\Gamma(k)\theta^k} z^{k-1} e^{-z/\theta}$	$\sqrt{\frac{\alpha}{2\pi z^3}} \exp\left(-\frac{\alpha(z-\mu)^2}{2z\mu^2}\right)$
Medidas	$\mathbb{E}[Z] = \theta k, \quad \text{Var}[Z] = \theta^2 k$	$\mathbb{E}[Z] = \mu, \quad \text{Var}[Z] = \mu^3/\alpha$
$f(z \mathbb{E}[Z] = 1)$	$\frac{1}{\Gamma(1/\sigma^2)\sigma^{2/\sigma^2}} z^{(1/\sigma^2)-1} \exp\left(-\frac{z}{\sigma^2}\right)$	$\sqrt{\frac{1}{2\pi\sigma^2 z^3}} \exp\left(-\frac{(z-1)^2}{\sigma^2 2z}\right)$
$\mathcal{L}(u)$	$\frac{1}{(1+\sigma^2 u)^{1/\sigma^2}}$	$\exp\left(\frac{(1-\sqrt{1+2\sigma^2 u})}{\sigma^2}\right)$
$\mathcal{L}^{(d)}(u)$	$\frac{(-1)^d \prod_{l=0}^{d-1} [1+l\sigma^2]}{(1+\sigma^2 u)^{1/\sigma^2+d}}$	Non ten expresión pechada
	Distribución positiva estábel	Distribución potencia-varianza
Notación	$PS(\theta, k)$	$PVF(\theta, \mu, \nu)$
$f(z)$	$-\frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{\Gamma(k\theta+1)}{k!} \left(-\frac{1}{z\theta}\right)^k \text{sen}(\theta k\pi)$	$\exp\left(-\frac{\nu}{\theta} \left(\frac{z}{\mu} + \frac{1}{\nu-1}\right)\right) \frac{1}{\pi z} \times$ $\sum_{k=1}^{\infty} \frac{(\nu/\theta)^{k\nu} (z/\mu)^{k(\nu-1)} \Gamma(1-k(\nu-1)) \text{sen}(\pi k(\nu-1))}{k!(\nu-1)^k}$
Medidas	Non existen	$\mathbb{E}[Z] = \mu, \quad \text{Var}[Z] = \theta\mu^2$
$f(z \mathbb{E}[Z] = 1)$	A media non existe	$\exp\left(-\frac{\nu}{\theta} \left(z + \frac{1}{\nu-1}\right)\right) \frac{1}{\pi z} \times$ $\sum_{k=1}^{\infty} \frac{(\nu/\theta)^{k\nu} (z)^{k(\nu-1)} \Gamma(1-k(\nu-1)) \text{sen}(\pi k(\nu-1))}{k!(\nu-1)^k}$
$\mathcal{L}(u)$	$\exp(-u^\theta)$	$\exp\left(\frac{\nu}{\theta(1-\nu)} \left(1 - \left(1 + \frac{\theta\mu u}{\nu}\right)^{1-\nu}\right)\right)$
Log-normal		
Notación	$LN(\mu, \gamma)$	
Medidas	$\mathbb{E}[Z] = e^{\gamma/2}, \quad \text{Var}[Z] = e^{2\gamma} - e^\gamma$	
$f(z \mathbb{E}[\exp(z)] = 0)$	$f(Z) = \frac{-1}{z\sqrt{2\pi\gamma}} \exp\left(-\frac{(\log(z))^2}{2\gamma}\right)$	
$\mathcal{L}(u)$	Non existe	

# Apéndice B

## Código de R

### B.1. Nesgo da distribución das fraxilidades preditas

```
windows()
par(mfrow = c(1, 3))
n = 5 * 10^3

theta = 1/0.5

ni = 2
t <- rexp(n*ni) * rep(rgamma(n, theta, theta), each = ni)
t.cluster <- aggregate(t, by = list(rep(1:n, each = ni)), FUN = sum)$x
hist((theta + ni)/(theta + t.cluster), freq = F, col = rgb(1, 0, 0, .5),
     xlim = c(0, 6), ylim = c(0, .8), ylab = "Densidade", xlab = "",
     main = paste0("clusters de tamaño ", ni), breaks = seq(0, 12, .25))
curve(dgamma(x, theta, theta), lwd = 2, add = T)

ni = 10
t <- rexp(n*ni) * rep(rgamma(n, theta, theta), each = ni)
t.cluster <- aggregate(t, by = list(rep(1:n, each = ni)), FUN = sum)$x
hist((theta + ni)/(theta + t.cluster), freq = F, col = rgb(1, 0, 0, .5),
     xlim = c(0, 6), ylim = c(0, .8), xlab = "Fraxilidade", ylab = "",
     main = paste0("clusters de tamaño ", ni), breaks = seq(0, 60, .25))
curve(dgamma(x, theta, theta), lwd = 2, add = T)

ni = 500
t <- rexp(n*ni) * rep(rgamma(n, theta, theta), each = ni)
t.cluster <- aggregate(t, by = list(rep(1:n, each = ni)), FUN = sum)$x
hist((theta + ni)/(theta + t.cluster), freq = F, col = rgb(1, 0, 0, .5),
     xlim = c(0, 6), ylim = c(0, .8), xlab = "", ylab = "",
     main = paste0("clusters de tamaño ", ni), breaks = seq(0, 120, .25))
curve(dgamma(x, theta, theta), lwd = 2, add = T)

legend("topright", legend = c("Real", "Predita"), bty = "n", col = c(1, 2),
      lwd = c(2, 0), lty = c(1, 0), pch = c(NA, 15), cex = 1.5,
      pt.cex = c(0, 3))
```

### B.2. Estudo de simulación

```
#####
##### ESTUDO DE SIMULACIÓN #####
#####
library(frailtyEM)
library(parfm)

rm(list = ls())

#####
#### Parámetros da simulación
## Número de simulacións
```

```

Nsim = 500
## Tamaño da mostra
# nni = 800
# nni = 400
nni = 100
## Número de clusters na mostra
# n = 80
# n = 40
n = 10
## Número de individuos en cada clúster
ni = nni/n

## Parámetros de regresión
beta1 = 1
beta2 = -1
beta3 = 0.5

## Parámetros da distribución da supervivencia
alpha = 3
lambda = 0.007

## Parámetro da distribución de fraxilidade
theta = 1/0.5

## Parámetro da distribución de censura
taxa.cens = 0.2
# taxa.cens = 0.5
# taxa.cens = 0.8

## Sementes
# set.seed(46730134) # n = 80, ni = 10
# set.seed(19428631) # n = 40, ni = 10
set.seed(19982642) # n = 10, ni = 10
# set.seed(29582775) # n = 10, ni = 80
# set.seed(42873721) # n = 10, ni = 40

Nsim.cens = 10^6

u <- runif(Nsim.cens)
x1 <- runif(Nsim.cens)
x2 <- rnorm(Nsim.cens)
x3 <- rbinom(Nsim.cens, 1, 0.25)

## Valores das fraxilidades e clusters
z <- rgamma(Nsim.cens/ni, theta, theta)
z.rep <- rep(z, each = ni)

## Tempos de supervivencia
t.bien <- (-log(u)/(lambda * exp(beta1 * x1 + beta2 * x2 + beta3 * x3) * z.rep))^(1/alpha)
t.homo <- (-log(u)/(lambda * exp(beta1 * x1 + beta2 * x2 + beta3 * x3)))^(1/alpha)
t.cuad <- (-log(u)/(lambda * exp(beta1 * x1 + beta2 * x2^2 + beta3 * x3) * z.rep))^(1/alpha)

rate.func <- function(rate, t){
  (mean(t <= qexp(seq(0 + 1/(Nsim.cens + 1), 1 - 1/(Nsim.cens + 1),
    length.out = Nsim.cens), rate)) - (1 - taxa.cens))^2}

rate.bien <- try(optimize(rate.func, interval = c(0, 4), t = t.bien)$minimum,
  silent = T)
rate.homo <- try(optimize(rate.func, interval = c(0, 4), t = t.homo)$minimum,
  silent = T)
rate.cuad <- try(optimize(rate.func, interval = c(0, 4), t = t.cuad)$minimum,
  silent = T)

##### Simulacións
## Variábeis para almacenar os resultados
crono.het <- numeric(0)
p.val.het <- numeric(0)

crono.homo <- numeric(0)
p.val.homo <- numeric(0)

crono.CA <- numeric(0)
p.val.CA <- numeric(0)

crono.CA.homo <- numeric(0)
p.val.CA.homo <- numeric(0)

```

```

p.val.bien <- numeric(0)
p.val.bien.m <- numeric(0)

p.val.gomp <- numeric(0)
p.val.gomp.m <- numeric(0)

p.val.cuad <- numeric(0)
p.val.cuad.m <- numeric(0)

for(isim in 1:Nsim){
  ##### Simulación dos valores
  ## Valor da supervivencia basal
  u <- runif(nni)

  ## Valor das covariábeis
  x1 <- runif(nni)
  x2 <- rnorm(nni)
  x3 <- rbinom(nni, 1, 0.25)

  ## Valores das fraxilidades e clusters
  z <- rgamma(n, theta, theta)
  z.rep <- rep(z, each = ni)
  id <- rep(1:n, each = ni)

  ## Tempos de supervivencia
  t.bien <- (-log(u)/(lambda * exp(beta1 * x1 + beta2 * x2 + beta3 * x3) * z.rep))^(1/alpha)
  t.homo <- (-log(u)/(lambda * exp(beta1 * x1 + beta2 * x2 + beta3 * x3)))^(1/alpha)
  t.cuad <- (-log(u)/(lambda * exp(beta1 * x1 + beta2 * x2^2 + beta3 * x3) * z.rep))^(1/alpha)

  ## Tempos de censura
  c.bien <- rexp(nni, rate = rate.bien)
  c.homo <- rexp(nni, rate = rate.homo)
  c.cuad <- rexp(nni, rate = rate.cuad)

  ## Tempo observado e delta
  y.bien <- apply(cbind(t.bien, c.bien), 1, min)
  delta.bien <- t.bien <= c.bien
  delta.bien = delta.bien * 1

  y.homo <- apply(cbind(t.homo, c.homo), 1, min)
  delta.homo <- t.homo <= c.homo
  delta.homo = delta.homo * 1

  y.cuad <- apply(cbind(t.cuad, c.cuad), 1, min)
  delta.cuad <- t.cuad <= c.cuad
  delta.cuad = delta.cuad * 1

  dat.bien <- data.frame(y = y.bien, delta = delta.bien, x1, x2, x3, id, z.rep)
  dat.homo <- data.frame(y = y.homo, delta = delta.homo, x1, x2, x3, id, z.rep)
  dat.cuad <- data.frame(y = y.cuad, delta = delta.cuad, x1, x2, x3, id, z.rep)

  ##### Test de heteroxeneidade - H1
  ## Mido o tempo de computación
  crono <- Sys.time()
  mod.het <- try(parfm(Surv(y, delta) ~ x1 + x2 + x3, dist = "weibull",
                     data = dat.bien), silent = T)

  ## Comprobo que o modelo converxeu e analizo os resultados
  if(mean(class(mod.het) != "try-error") == 1){
    ## Risco basal cumulativo
    alpha.est.het <- mod.het[1, 1]
    lambda.est.het <- mod.het[2, 1]
    Lambda0t.est.het <- lambda.est.het * dat.bien$y^alpha.est.het

    ## Efecto das covariábeis
    beta.est.het <- mod.het[3:5, 1]
    lin.pred.het <- as.numeric(beta.est.het %*% t(dat.bien[, 3:5]))

    ## Probabilidades de supervivencia
    P.surv.het <- as.numeric(exp(-Lambda0t.est.het * exp(lin.pred.het)))

    ## Probabilidades de supervivencia aleatorizadas
    RSP.het <- P.surv.het
    RSP.het[dat.bien$delta == 0] = runif(nni - sum(dat.bien$delta)) * RSP.het[dat.bien$delta == 0]

    ## Probabilidades de supervivencia aleatorizadas normalizadas
    NRSP.het <- qnorm(RSP.het)

```

```

## Test de igualdade de medias
lm.het <- lm(NRSP.het ~ as.factor(dat.bien$id))
test.het <- anova(lm.het)
p.val.het <- c(p.val.het, test.het$`Pr(>F)`[1])

## Tempo de computación
crono.het <- c(crono.het, Sys.time() - crono)
} else {
p.val.het <- c(p.val.het, NA)
crono.het <- c(crono.het, NA)
}

##### Test de heteroxeneidade - H0
crono <- Sys.time()
mod.homo <- try(parfm(Surv(y, delta) ~ x1 + x2 + x3, dist = "weibull",
                    data = dat.homo), silent = T)

## Comprobo que o modelo converxeu e analizo os resultados
if(mean(class(mod.homo) != "try-error") == 1){
## Risco basal cumulativo
alpha.est.homo <- mod.homo[1, 1]
lambda.est.homo <- mod.homo[2, 1]
Lambda0t.est.homo <- lambda.est.homo * dat.homo$y^alpha.est.homo

## Efecto das covariábeis
beta.est.homo <- mod.homo[3:5, 1]
lin.pred.homo <- as.numeric(beta.est.homo %*% t(dat.homo[, 3:5]))

## Probabilidades de supervivencia
P.surv.homo <- as.numeric(exp(-Lambda0t.est.homo * exp(lin.pred.homo)))

## Probabilidades de supervivencia aleatorizadas
RSP.homo <- P.surv.homo
RSP.homo[dat.homo$delta == 0] = runif(nni - sum(dat.homo$delta)) * RSP.homo[dat.homo$delta == 0]

## Probabilidades de supervivencia aleatorizadas normalizadas
NRSP.homo <- qnorm(RSP.homo)

## Test de igualdade de medias
lm.homo <- lm(NRSP.homo ~ as.factor(dat.homo$id))
test.homo <- anova(lm.homo)
p.val.homo <- c(p.val.homo, test.homo$`Pr(>F)`[1])

## Tempo de computación
crono.homo <- c(crono.homo, Sys.time() - crono)
} else {
p.val.homo <- c(p.val.homo, NA)
crono.homo <- c(crono.homo, NA)
}

##### Test de heteroxeneidade de Commenges-Andersen - H1
crono <- Sys.time()
mod.CA <- try(coxph(Surv(y, delta) ~ x1 + x2 + x3, data = dat.bien,
                  model = TRUE, x = TRUE), silent = T)

if(mean(class(mod.CA) != "try-error") == 1){
test.CA <- ca_test(mod.CA, id = dat.bien$id)
p.val.CA = c(p.val.CA, test.CA[3])
crono.CA <- c(crono.CA, Sys.time() - crono)
} else {
p.val.CA = c(p.val.CA, NA)
}

##### Test de heteroxeneidade de Commenges-Andersen - H0
crono <- Sys.time()
mod.CA.homo <- try(coxph(Surv(y, delta) ~ x1 + x2 + x3, data = dat.homo,
                      model = TRUE, x = TRUE), silent = T)

if(mean(class(mod.CA.homo) != "try-error") == 1){
test.CA.homo <- ca_test(mod.CA.homo, id = dat.homo$id)
p.val.CA.homo = c(p.val.CA.homo, test.CA.homo[3])
crono.CA.homo <- c(crono.CA.homo, Sys.time() - crono)
} else {

```

```

p.val.CA.homo = c(p.val.CA.homo, NA)
}

#### Test de bondade de axuste - Distribución correcta
mod.bien <- try(parfm(Surv(y, delta) ~ x1 + x2 + x3, cluster = "id", data = dat.bien,
  dist = "weibull", frailty = "gamma"), silent = T)

if(mean(class(mod.bien) != "try-error") == 1){
  ## Risco marxinal estimado
  lambda.est.bien <- mod.bien[3, 1]
  alpha.est.bien <- mod.bien[2, 1]
  Lambda0t.est.bien <- lambda.est.bien * dat.bien$y^alpha.est.bien

  ## Efecto das covariábeis estimado
  beta.est.bien <- mod.bien[4:6, 1]
  lin.pred.bien <- as.numeric(beta.est.bien %*% t(dat.bien[, 3:5]))

  ## Fraxilidades preditas
  z.pred.bien <- predict(mod.bien)

  ## Probabilidades de supervivencia estimadas
  ## Condicionada
  P.surv.bien <- as.numeric(exp(-Lambda0t.est.bien * exp(lin.pred.bien) *
    rep(z.pred.bien, each = ni)))

  ## Marxinal
  theta.est.bien <- mod.bien[1, 1]
  P.surv.bien.m <- (1 + theta.est.bien * Lambda0t.est.bien * exp(lin.pred.bien))^(1/theta.est.bien)

  ## Probabilidades de supervivencia aleatorizadas
  RSP.bien <- P.surv.bien
  RSP.bien[dat.bien$delta == 0] = runif(nni - sum(dat.bien$delta)) *
  RSP.bien[dat.bien$delta == 0]
  RSP.bien.m <- P.surv.bien.m
  RSP.bien.m[dat.bien$delta == 0] = runif(nni - sum(dat.bien$delta)) *
  RSP.bien.m[dat.bien$delta == 0]

  ## Probabilidades de supervivencia aleatorizadas normalizadas
  NRSP.bien = qnorm(RSP.bien)
  NRSP.bien.m = qnorm(RSP.bien.m)

  p.val.bien = c(p.val.bien, shapiro.test(NRSP.bien)$p.value)
  p.val.bien.m = c(p.val.bien.m, shapiro.test(NRSP.bien.m)$p.value)
} else {
  p.val.bien = c(p.val.bien, NA)
  p.val.bien.m = c(p.val.bien.m, NA)
}

#### Test de bondade de axuste - Distribución incorrecta
mod.gomp <- try(parfm(Surv(y, delta) ~ x1 + x2 + x3, cluster = "id", data = dat.bien,
  dist = "gompertz", frailty = "gamma"), silent = T)

if(mean(class(mod.gomp) != "try-error") == 1){
  ## Risco basal estimado
  lambda.est.gomp <- mod.gomp[3, 1]
  gamma.est.gomp <- mod.gomp[2, 1]
  Lambda0t.est.gomp <- (lambda.est.gomp/gamma.est.gomp) * (exp(dat.bien$y * gamma.est.gomp)-1)

  ## Efecto das covariábeis estimado
  beta.est.gomp <- mod.gomp[4:6, 1]
  lin.pred.gomp <- as.numeric(beta.est.gomp %*% t(dat.bien[, 3:5]))

  ## Fraxilidades preditas
  z.pred.gomp <- predict(mod.gomp)

  ## Probabilidades de supervivencia estimadas
  P.surv.gomp <- as.numeric(exp(-Lambda0t.est.gomp * exp(lin.pred.gomp) *
    rep(z.pred.gomp, each = ni)))

  ## Marxinal
  theta.est.gomp <- mod.gomp[1, 1]
  P.surv.gomp.m <- (1 + theta.est.gomp * Lambda0t.est.gomp * exp(lin.pred.gomp))^(1/theta.est.gomp)

  ## Probabilidades de supervivencia aleatorizadas
  RSP.gomp <- P.surv.gomp
  RSP.gomp[dat.bien$delta == 0] = runif(nni - sum(dat.bien$delta)) *
  RSP.gomp[dat.bien$delta == 0]
}

```

```

RSP.gomp.m <- P.surv.gomp.m
RSP.gomp.m[dat.bien$delta == 0] = runif(nni - sum(dat.bien$delta)) *
  RSP.gomp.m[dat.bien$delta == 0]

## Probabilidades de supervivencia aleatorizadas normalizadas
NRSP.gomp = qnorm(RSP.gomp)
NRSP.gomp.m = qnorm(RSP.gomp.m)

p.val.gomp = c(p.val.gomp, shapiro.test(NRSP.gomp)$p.value)
p.val.gomp.m = c(p.val.gomp.m, shapiro.test(NRSP.gomp.m)$p.value)
} else {
  p.val.gomp = c(p.val.gomp, NA)
  p.val.gomp.m = c(p.val.gomp.m, NA)
}

### Test de bondade de axuste - Efecto non linear das covariábeis
mod.cuad <- try(parfm(Surv(y, delta) ~ x1 + x2 + x3, cluster = "id", data = dat.cuad,
  dist = "weibull", frailty = "gamma"), silent = T)

if(mean(class(mod.cuad) != "try-error") == 1){
  ## Risco basal estimado
  lambda.est.cuad <- mod.cuad[3, 1]
  alpha.est.cuad <- mod.cuad[2, 1]
  Lambda0t.est.cuad <- lambda.est.cuad * dat.cuad$y^alpha.est.cuad

  ## Efecto das covariábeis estimado
  beta.est.cuad <- mod.cuad[4:6, 1]
  lin.pred.cuad <- as.numeric(beta.est.cuad %*% t(dat.cuad[, 3:5]))

  ## Fraxilidades preditas
  z.pred.cuad <- predict(mod.cuad)

  ## Probabilidades de supervivencia estimadas
  ## Condicionais
  P.surv.cuad <- as.numeric(exp(-Lambda0t.est.cuad * exp(lin.pred.cuad)) *
    rep(z.pred.cuad, each = ni))

  ## Marxinais
  theta.est.cuad <- mod.cuad[1, 1]
  P.surv.cuad.m <- (1 + theta.est.cuad * Lambda0t.est.cuad * exp(lin.pred.cuad))^(1/theta.est.cuad)

  ## Probabilidades de supervivencia aleatorizadas
  RSP.cuad <- P.surv.cuad
  RSP.cuad[dat.cuad$delta == 0] = runif(nni - sum(dat.cuad$delta)) *
    RSP.cuad[dat.cuad$delta == 0]
  RSP.cuad.m <- P.surv.cuad.m
  RSP.cuad.m[dat.cuad$delta == 0] = runif(nni - sum(dat.cuad$delta)) *
    RSP.cuad.m[dat.cuad$delta == 0]

  ## Probabilidades de supervivencia aleatorizadas normalizadas
  NRSP.cuad = qnorm(RSP.cuad)
  NRSP.cuad.m = qnorm(RSP.cuad.m)

  p.val.cuad = c(p.val.cuad, shapiro.test(NRSP.cuad)$p.value)
  p.val.cuad.m = c(p.val.cuad.m, shapiro.test(NRSP.cuad.m)$p.value)
} else {
  p.val.cuad = c(p.val.cuad, NA)
  p.val.cuad.m = c(p.val.cuad.m, NA)
}
}

save.image(file = paste0("cens", taxa.cens, "n", n, "ni", ni, ".Rdata"))

```

## B.3. Aplicación dos métodos de diagnóstico propostos

### B.3.1. Test de heteroxeneidade

```

library(survival)
library(parfm)

n = nrow(retinopathy)

### Test de heteroxeneidade
mod.homo <- parfm(Surv(futime, status) ~ trt + risk, dist = "weibull",

```



```

data = retinopathy)

alpha.est.homo <- mod.homo[1, 1]
lambda.est.homo <- mod.homo[2, 1]
Lambda0t.est.homo <- lambda.est.homo * retinopathy$futime^alpha.est.homo

## Efecto das covariáveis
beta.est.homo <- mod.homo[3:4, 1]
lin.pred.homo <- as.numeric(beta.est.homo %% t(cbind(retinopathy$trt, retinopathy$risk)))

## Probabilidades de supervivencia
P.surv.homo <- as.numeric(exp(-Lambda0t.est.homo * exp(lin.pred.homo)))

## Probabilidades de supervivencia aleatorizadas
RSP.homo <- P.surv.homo
RSP.homo[retinopathy$status == 0] = runif(n - sum(retinopathy$status)) *
  RSP.homo[retinopathy$status == 0]

## Probabilidades de supervivencia aleatorizadas normalizadas
NRSP.homo <- qnorm(RSP.homo)

## Test de igualdade de medias
set.seed(29452207)
lm.homo <- lm(NRSP.homo ~ as.factor(retinopathy$id))
test.homo <- anova(lm.homo)
p.val.homo <- test.homo$`Pr(>F)`[1]

id.mal = sample(retinopathy$id)
lm.homo.mal <- lm(NRSP.homo ~ as.factor(id.mal))
test.homo.mal <- anova(lm.homo.mal)
p.val.homo.mal <- test.homo.mal$`Pr(>F)`[1]

## Representación gráfica
windows()
par(mfrow = c(1, 2))
boxplot(NRSP.homo ~ retinopathy$id, ylim = c(-3, 3), main = "Mostra orixinal",
  xlab = "id", ylab = "NRSP")
boxplot(NRSP.homo ~ id.mal, ylim = c(-3, 3), main = "Mostra bootstrap",
  xlab = "id", ylab = "")

```

### B.3.2. Test de bondade de axuste do modelo de regresión

```

mod <- parfm(Surv(futime, status) ~ trt + risk, cluster = "id", dist = "weibull",
  frailty = "gamma", data = retinopathy)

lambda.est <- mod[3, 1]
alpha.est <- mod[2, 1]
Lambda0t.est <- lambda.est * retinopathy$futime^alpha.est

## Efecto das covariáveis estimado
beta.est <- mod[4:5, 1]
lin.pred <- as.numeric(beta.est %% t(cbind(retinopathy$trt, retinopathy$risk)))

## Probabilidades de supervivencia
z.pred <- rep(predict(mod), 2)
theta.est <- mod[1, 1]

P.surv <- exp(-Lambda0t.est * lin.pred * z.pred)
P.surv.m <- (1 + theta.est * Lambda0t.est * exp(lin.pred))^(1/theta.est)

## Probabilidades de supervivencia aleatorizadas
RSP <- P.surv
RSP.m <- P.surv.m
set.seed(70225492)
RSP[retinopathy$status == 0] = runif(n - sum(retinopathy$status)) *
  RSP[retinopathy$status == 0]
RSP.m[retinopathy$status == 0] = runif(n - sum(retinopathy$status)) *
  RSP.m[retinopathy$status == 0]

## Probabilidades de supervivencia aleatorizadas normalizadas
NRSP = qnorm(RSP)
NRSP.m = qnorm(RSP.m)

p.val = shapiro.test(NRSP)$p.value
p.val.m = shapiro.test(NRSP.m)$p.value

```

```

## Representación gráfica
windows()
par(mfrow = c(2, 2))
qqnorm(NRSP.m, main = "QQ-plot dos residuos respecto da dist.normal", ylab = "Cantís mostrais",
       xlab = "Cantís teóricos", ylim = c(-3, 3))
qqline(NRSP.m)
hist(NRSP.m, main = "Histograma dos residuos marxinais", freq = F, ylab = "Densidade",
     xlab = "Residuos", xlim = c(-4, 4))
boxplot(NRSP.m ~ retinopathy$str, ylim = c(-3, 3), xlab = "Tratamento (binaria)",
       ylab = "Residuos", main = "Residuos respecto do tratamento")
plot(NRSP.m ~ retinopathy$risk, ylim = c(-3, 3), xlab = "Medida do risco",
     ylab = "Residuos", main = "Residuos respecto da medida do risco")
lines(lowess(NRSP.m ~ retinopathy$risk))

```

### B.3.3. Modelo de regresión con interaccións

```

library(survival)
library(parfm)

n = nrow(retinopathy)

#####
##### Test de heteroxeneidade #####
#####
mod.homo.2 <- parfm(Surv(futime, status) ~ trt*type + risk, dist = "weibull",
                  data = retinopathy)

alpha.est.homo.2 <- mod.homo.2[1, 1]
lambda.est.homo.2 <- mod.homo.2[2, 1]
Lambda0t.est.homo.2 <- lambda.est.homo.2 * retinopathy$futime^alpha.est.homo.2

## Efecto das covariábeis
beta.est.homo.2 <- mod.homo.2[4:7, 1]
matriz.data <- model.matrix(~ trt*type + risk, data = retinopathy)[, -1]
lin.pred.homo.2 <- as.numeric(beta.est.homo.2 %*% t(matriz.data))

## Probabilidades de supervivencia
P.surv.homo.2 <- as.numeric(exp(-Lambda0t.est.homo.2 * exp(lin.pred.homo.2)))

## Probabilidades de supervivencia aleatorizadas
set.seed(4276504)
RSP.homo.2 <- P.surv.homo.2
RSP.homo.2[retinopathy$status == 0] = runif(n - sum(retinopathy$status)) *
  RSP.homo.2[retinopathy$status == 0]

## Probabilidades de supervivencia aleatorizadas normalizadas
NRSP.homo.2 <- qnorm(RSP.homo.2)

## Test de igualdade de medias
lm.homo.2 <- lm(NRSP.homo.2 ~ as.factor(retinopathy$id))
test.homo.2 <- anova(lm.homo.2)
p.val.homo.2 <- test.homo.2$Pr(>F)[1]

#####
##### Test de bondade de axuste #####
#####
mod.2 <- parfm(Surv(futime, status) ~ trt*type + risk, cluster = "id", dist = "weibull",
              frailty = "gamma", data = retinopathy)

lambda.est.2 <- mod.2[3, 1]
alpha.est.2 <- mod.2[2, 1]
Lambda0t.est.2 <- lambda.est.2 * retinopathy$futime^alpha.est.2

## Efecto das covariábeis estimado
beta.est.2 <- mod.2[4:7, 1]
lin.pred.2 <- as.numeric(beta.est.2 %*% t(matriz.data))

## Probabilidades de supervivencia
theta.est.2 <- mod.2[1, 1]

P.surv.m.2 <- (1 + theta.est.2 * Lambda0t.est.2 * exp(lin.pred.2))^(1/theta.est.2)

## Probabilidades de supervivencia aleatorizadas
RSP.m.2 <- P.surv.m.2
set.seed(4056724)
RSP.m.2[retinopathy$status == 0] = runif(n - sum(retinopathy$status)) *

```

```

RSP.m.2[retinopathy$status == 0]

## Probabilidades de supervivencia aleatorizadas normalizadas
NRSP.m.2 = qnorm(RSP.m.2)
p.val.m.2 = shapiro.test(NRSP.m.2)$p.value

## Representación gráfica
windows()
par(mfrow = c(1, 2))
qqnorm(NRSP.m.2, main = "QQ-plot dos residuos respecto da dist.normal", ylab = "Cantís mostrais",
       xlab = "Cantís teóricos", ylim = c(-3, 3))
qqline(NRSP.m.2)
hist(NRSP.m.2, main = "Histograma dos residuos marxinais", freq = F, ylab = "Densidade",
     xlab = "Residuos", xlim = c(-4, 4))
par(mfrow = c(1, 1))
boxplot(NRSP.homo.2 ~ retinopathy$id, ylim = c(-3, 3), xlab = "id", ylab = "NRSP",
       main = "NRSP sobre a variábel agrupadora",)
par(mfrow = c(2, 2))
boxplot(NRSP.m.2 ~ retinopathy$trt, ylim = c(-3, 3), xlab = "Tratamento (binaria)",
       ylab = "Residuos", main = "Residuos respecto do tratamento")
boxplot(NRSP.m.2 ~ retinopathy$type, ylim = c(-3, 3), xlab = "Grupo de idade (binaria)",
       ylab = "Residuos", main = "Residuos respecto do grupo de idade")
boxplot(NRSP.m.2 ~ retinopathy$type, ylim = c(-3, 3), xlab = "Interacción tratamento*idade (binaria)",
       ylab = "Residuos", main = "Residuos respecto da interacción do tratamento e a idade")
plot(NRSP.m.2 ~ retinopathy$risk, ylim = c(-3, 3), xlab = "Medida do risco",
     ylab = "Residuos", main = "Residuos respecto da medida do risco")
lines(lowess(NRSP.m.2~retinopathy$risk))

```



## Apéndice C

# Derivación da densidade posterior das fragilidades gamma

Comezamos por construír a fórmula usando (2.5, 2.10 e 2.11) coa forma de (2.9) e obtemos

$$f(z_i | \mathbf{t}_i, \boldsymbol{\delta}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\lambda}, \mathbf{x}_i) = \frac{\prod_{j=1}^{n_i} \left[ \left( \hat{\lambda}_0(t_{ij}) e^{\hat{\boldsymbol{\beta}}^t \mathbf{x}_{ij}} \right)^{\delta_{ij}} \right] z_i^{d_i} \exp(-\hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i) z_i) \frac{1}{\Gamma(1/\hat{\sigma}^2) \hat{\sigma}^{2/\hat{\sigma}^2}} z_i^{1/\hat{\sigma}^2 - 1} \exp(-z_i/\hat{\sigma}^2)}{\prod_{j=1}^{n_i} \left[ \left( \hat{\lambda}_0(t_{ij}) e^{\hat{\boldsymbol{\beta}}^t \mathbf{x}_{ij}} \right)^{\delta_{ij}} \right] \prod_{l=0}^{d_i-1} [1 + l\hat{\sigma}^2] \left( \frac{1}{1 + \hat{\sigma}^2 \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i)} \right)^{1/\hat{\sigma}^2 + d_i}}.$$

Os produtos das funcións de risco dos membros do clúster no dividendo e no divisor canceláanse. Ademais, xuntamos as  $z_i$  no dividendo baixo un único exponente, e xuntamos as funcións exponenciais extraendo o factor común  $-z_i$ . Por último,  $\Gamma(1/\hat{\sigma}^2) \hat{\sigma}^{2/\hat{\sigma}^2}$  que está dividindo no dividendo pásase ao divisor, o que resulta en

$$f(z_i | \mathbf{t}_i, \boldsymbol{\delta}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\lambda}, \mathbf{x}_i) = \frac{z_i^{1/\hat{\sigma}^2 + d_i - 1} \exp\left(-z_i(1/\hat{\sigma}^2 + \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i) z_i)\right)}{\Gamma(1/\hat{\sigma}^2) \hat{\sigma}^{2/\hat{\sigma}^2} \prod_{l=0}^{d_i-1} [1 + l\hat{\sigma}^2] \left( \frac{1}{1 + \hat{\sigma}^2 \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i)} \right)^{1/\hat{\sigma}^2 + d_i}}.$$

Se agora extraemos  $\hat{\sigma}^{2/\hat{\sigma}^2 + 2d_i}$  do último termo do divisor de xeito que teñamos

$$\left( \frac{1}{1 + \hat{\sigma}^2 \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i)} \right)^{1/\hat{\sigma}^2 + d_i} = \left( \frac{1}{1/\hat{\sigma}^2 + \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i)} \right)^{1/\hat{\sigma}^2 + d_i} \frac{1}{\hat{\sigma}^{2/\hat{\sigma}^2}} \frac{1}{\hat{\sigma}^{2d_i}}$$

podemos eliminar as  $\hat{\sigma}^{2/\hat{\sigma}^2}$  do divisor que respectivamente multiplican e dividen e resulta

$$f(z_i | \mathbf{t}_i, \boldsymbol{\delta}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\lambda}, \mathbf{x}_i) = \frac{z_i^{1/\hat{\sigma}^2 + d_i - 1} \exp\left(-z_i(1/\hat{\sigma}^2 + \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i) z_i)\right)}{\frac{\Gamma(1/\hat{\sigma}^2) \prod_{l=0}^{d_i-1} [1 + l\hat{\sigma}^2]}{\hat{\sigma}^{2d_i}} \left( \frac{1}{1/\hat{\sigma}^2 + \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i)} \right)^{1/\hat{\sigma}^2 + d_i}}.$$

Tense que  $\Gamma(1/\hat{\sigma}^2 + d_i) = \frac{\Gamma(1/\hat{\sigma}^2) \prod_{l=0}^{d_i-1} [1 + l\hat{\sigma}^2]}{\hat{\sigma}^{2d_i}}$ , e polo tanto

$$f(z_i | \mathbf{t}_i, \boldsymbol{\delta}_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \hat{\lambda}, \mathbf{x}_i) = \frac{z_i^{1/\hat{\sigma}^2 + d_i - 1} \exp\left(-z_i(1/\hat{\sigma}^2 + \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i) z_i)\right)}{\Gamma(1/\hat{\sigma}^2 + d_i) \left( \frac{1}{1/\hat{\sigma}^2 + \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i | \mathbf{x}_i)} \right)^{1/\hat{\sigma}^2 + d_i}},$$

e con iso temos a función que buscábamos con

$$\hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i|\mathbf{x}_i) = \sum_{j=1}^{n_i} \hat{\Lambda}_0(t_{ij}) e^{\hat{\theta}^t \mathbf{x}_{ij}}$$

ao igual que en (2.4).

Dado que a densidade dunha variábel con distribución gamma é

$$f(z) = \frac{1}{\Gamma(k)\theta^k} z^{k-1} e^{-z/\theta},$$

temos que a anterior é a densidade dunha distribución gamma de parámetros

$$k = 1/\hat{\sigma}^2 + d_i,$$

$$\theta = \frac{1}{1/\hat{\sigma}^2 + \hat{\Lambda}_{\text{clúster}}(\mathbf{t}_i|\mathbf{x}_i)}.$$

## Apéndice D

### Táboas adicionais sobre o estudo de simulación

#### D.1. Proporción de axustes que non converxeron

cens.		20 %					50 %					80 %				
$n$		10	10	10	40	80	10	10	10	40	80	10	10	10	40	80
Poboación	Axuste \ $n_i$	10	40	80	10	10	10	40	80	10	10	10	40	80	10	10
Homoxénea	Homoxéneo	4.2	0	0	0	0	5.2	0	0	0	0	23.4	1.6	0	0.6	0
Base	Homoxéneo	0.4	0	0.2	0	0	1.2	0	0	0	0	14.0	0.4	0	0	0
	Base	4.6	0.2	0	0	0	7.2	0	0	0	0	21.6	2.6	0.6	1.2	0.2
	Gompertz	0	0	0	0	0	0.2	0.2	0	0	0	18.2	0	0	0.6	0
caadrática	Base	0	0	0	0	0	2.4	0	0	0	0	26.0	1.6	0.2	1.6	0.4

Táboa D.1: Proporción de axustes que non converxeron para as distintas poboacións e tests utilizados nas simulacións

## D.2. Coincidencia entre tests

Test			Homoxeneidade						Bondade de axuste								
Poboación			Homoxénea			Base			Base			Gompertz			caadrática		
cens.	$n$	$n_i \setminus \alpha$	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
20%	10	10	98.3	91.9	81.8	85.5	89.8	93.8	98.7	91.2	82.6	70.6	60.2	59.6	78.8	79.0	79.4
	10	40	97.6	89.8	83.4	100	99.8	100	95.6	83.8	76.6	87.4	94.2	96.8	98.6	99.4	99.6
	10	80	98.0	89.4	83.0	100	100	100	90.6	76.2	66.2	99.4	100	100	100	100	100
	40	10	99.0	91.8	83.0	100	100	100	94.2	82.4	70.8	57.4	75.4	83.8	97.8	99.4	99.6
	80	10	98.6	92.0	85.0	100	100	100	88.8	71.0	59.2	90.4	96.4	97.8	100	100	100
50%	10	10	96.4	87.8	80.2	68.0	72.3	75.7	97.4	90.3	83.2	86.8	71.7	67.5	91.8	81.8	75.4
	10	40	98.0	91.4	81.6	93.8	97.6	98.6	97.6	88.8	80.4	58.8	68.8	77.2	73.0	80.8	85.0
	10	80	97.2	91.2	83.4	99.4	99.6	99.8	96.6	89.8	79.8	91.0	96.6	98.4	94.4	97.4	98.6
	40	10	97.6	89.4	81.8	87.8	95.6	96.6	96.6	88.6	80.6	39.0	44.8	53.8	67.2	67.6	69.8
	80	10	98.8	91.4	81.4	98.8	99.4	100	93.8	83.0	76.6	51.8	69.8	78.2	76.8	86.0	91.2
80%	10	10	97.7	88.8	79.4	84.9	72.6	65.1	98.0	93.1	85.2	96.8	87.3	76.5	99.2	94.6	87.8
	10	40	98.0	90.0	81.5	40.6	52.8	67.1	99.0	93.8	86.9	83.6	62.0	54.8	93.7	80.1	72.0
	10	80	99.0	91.6	81.8	67.2	80.0	85.6	97.2	90.1	83.9	62.0	59.6	63.8	86.4	72.1	65.3
	40	10	98.2	89.9	80.5	49.2	44.4	47.2	99.2	92.9	84.2	80.9	65.0	58.4	97.2	88.8	82.5
	80	10	98.4	87.4	79.8	23.6	42.2	54.8	99.0	90.8	84.6	59.4	53.0	57.4	95.2	80.7	73.5

Táboa D.2: Porcentaxe de rexeitamentos para cada test e escenario para distintos niveis de significación



# Bibliografía

- Bagdonavičius, V., & Nikulin, M. (2011). Chi-squared tests for general composite hypotheses from censored samples. *Comptes Rendus Mathématique*, *349*(3-4), 219–223. <https://doi.org/10.1016/j.crma.2011.01.007>.
- Balan, T. A., & Putter, H. (2019). frailtyEM: An R package for estimating semiparametric shared frailty models. *Journal of Statistical Software*, *90*(7), 1–29. <https://doi.org/10.18637/jss.v090.i07>.
- Balan, T. A., & Putter, H. (2020). A tutorial on frailty models. *Statistical methods in medical research*, *29*(11), 3424–3454. <https://doi.org/10.1177/0962280220921889>.
- Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part II: multivariate data analysis—an introduction to concepts and methods. *British journal of cancer*, *89*(3), 431–436. <https://doi.org/10.1038/sj.bjc.6601119>.
- Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. *British journal of cancer*, *89*(2), 232–238. <https://doi.org/10.1038/sj.bjc.6601118>.
- Commenges, D., & Andersen, P. K. (1995). Score test of homogeneity for survival data. *Lifetime data analysis*, *1*, 145–156. <https://doi.org/10.1007/BF00985764>.
- Comprehensive R Archive Network. (2020, outubro 4). *CRAN - Package phmm*. <https://cran.r-project.org/web/packages/phmm/index.html>.
- Comprehensive R Archive Network. (2024, apr 13). *CRAN - Package frailtyPack*. <https://cran.r-project.org/web/packages/frailtypack/index.html>.
- Cui, S., & Sun, Y. (2004). Checking for the gamma frailty distribution under the marginal proportional hazards frailty model. *Statistica Sinica*, *14*(1), 249–267.
- Duchateau, L., & Janssen, P. (2008). *The frailty model* (1st ed.). Springer. <https://doi.org/10.1007/978-0-387-72835-3>.
- Geerdens, C., Claeskens, G., & Janssen, P. (2013). Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty. *Biostatistics*, *14*(3), 433–446. <https://doi.org/10.1093/biostatistics/kxs053>.
- Glidden, D. V. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika*, *86*(2), 381–393. <https://doi.org/10.1093/biomet/86.2.381>.
- Glidden, D. V., & Vittinghoff, E. (2004). Modelling clustered survival data from multicentre clinical trials. *Statistics in medicine*, *23*(3), 369–388. <https://doi.org/10.1002/sim.1599>.
- González-Manteiga, W., Martínez-Miranda, M. D., & Van Keilegom, I. (2023). Goodness-of-fit tests in proportional hazards models with random effects. *Biometrical Journal*, *65*(1), 2000353. <https://doi.org/10.1002/bimj.202000353>.
- Gorfine, M., De-Picciotto, R., & Hsu, L. (2012). Conditional and marginal estimates in case-control family data—extensions and sensitivity analyses. *Journal of statistical computation and simulation*, *82*(10), 1449–1470. <https://doi.org/10.1080/00949655.2011.581669>.
- Gorfine, M., & Zucker, D. M. (2023). Shared frailty methods for complex survival data: a review of recent advances. *Annual Review of Statistics and Its Application*, *82*(10), 51–73. <https://doi.org/10.1146/annurev-statistics-032921-021310>.

- Gorfine, M., Zucker, D. M., & Hsu, L. (2006). Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach. *Biometrika*, *93*(3), 735–741. <https://doi.org/10.1093/biomet/93.3.735>.
- Govindarajulu, U. S., & D’Agostino, R. B. (2020). Review of current advances in survival analysis and frailty models. *Wiley Interdisciplinary Reviews: Computational Statistics*, *12*(6), e1504. <https://doi.org/10.1002/wics.1504>.
- Gu, C. (2014). Smoothing spline anova models: R package gss. *Journal of Statistical Software*, *58*(5), 1–25. <https://doi.org/10.18637/jss.v058.i05>.
- Ha, I. D., & Lee, Y. (2003). Estimating frailty models via poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*, *12*(3), 663–681. <https://doi.org/10.1198/1061860032256>.
- Ha, I. D., Lee, Y., & Song, J.-k. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, *88*(1), 233–233. <https://doi.org/10.1093/biomet/88.1.233>.
- Ha, I. D., Noh, J. K., & Lee, Y. (2019). frailtyhl: A package for fitting frailty models with h-likelihood. *R package version*.
- Hirsch, K., & Wienke, A. (2012). Software for semiparametric shared gamma and log-normal frailty models: an overview. *Computer methods and programs in biomedicine*, *107*(3), 582–597. <https://doi.org/10.1016/j.cmpb.2011.05.004>.
- Hsu, L., Gorfine, M., & Malone, K. (2007). On robustness of marginal regression coefficient estimates and hazard functions in multivariate survival analysis of family data when the frailty distribution is mis-specified. *Statistics in Medicine*, *26*(25), 4657–4678. <https://doi.org/10.1002/sim.2870>.
- Kats, L., & Gorfine, M. (2023). An accelerated failure time regression model for illness–death data: A frailty approach. *Biometrics*, *79*(4), 3066–3081. <https://doi.org/10.1111/biom.13880>.
- Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data* (2nd ed., Vol. 1230). Springer. <https://doi.org/10.1007/b97377>.
- Li, L., Wu, T., & Feng, C. (2021). Model diagnostics for censored regression via randomized survival probabilities. *Statistics in medicine*, *40*(6), 1482–1497. <https://doi.org/10.1002/sim.8852>.
- McGilchrist, C. A. (1993). ReML estimation for survival models with frailty. *Biometrics*, 221–225. <https://doi.org/10.2307/2532615>.
- Monaco, J. V., Gorfine, M., & Hsu, L. (2018). General semiparametric shared frailty model: estimation and simulation with frailtysurv. *Journal of statistical software*, *86*(4). <https://doi.org/10.18637/jss.v086.i04>.
- Moore, D. F. (2016). *Applied survival analysis using R* (1st ed., Vol. 473). Springer. <https://doi.org/10.1007/978-3-319-31245-3>.
- Munda, M., & Legrand, C. (2013). *A diagnostic plot for guiding the choice of the frailty distribution in clustered survival data* (Tech. Rep.). Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles.
- Munda, M., Rotolo, F., & Legrand, C. (2012). parfm: Parametric frailty models in r. *Journal of Statistical Software*, *51*(11), 1–20. <https://doi.org/10.18637/jss.v051.i11>.
- Rotolo, F., Munda, M., & Callegaro, A. (2023). Package ‘parfm’. *R package version*. <https://doi.org/10.32614/CRAN.package.parfm>.
- Shih, J. H., & Louis, T. A. (1995). Assessing gamma frailty models for clustered failure time data. *Lifetime Data Analysis*, *1*, 205–220. <https://doi.org/10.1007/BF00985771>.
- Spiegel, M. R. (1965). *Laplace transforms* (1st ed.). McGraw-Hill New York.
- Therneau, T. M. (2024). Package ‘coxme’. *R package version*, *2*(2-20). <https://doi.org/10.32614/CRAN.package.coxme>.
- Therneau, T. M., & Lumley, T. (2024). R survival package. *R Core Team*. <https://doi.org/10.32614/CRAN.package.survival>.
- Vonta, F. (2020). Goodness-of-fit tests based on divergence measures for frailty models. *Statistical Topics and Stochastic Models for Dependent Data with Applications*, 235–251. <https://doi.org/10.1002/9781119779421.ch11>.

- Vonta, F., & Karagrigoriou, A. (2014). Goodness-of-fit tests via  $\phi$ -measures of divergence for censored data. *Journal of Statistical Computation and Simulation*, 84(5), 946–963. <https://doi.org/10.1080/00949655.2012.733396>.
- Wienke, A. (2003). Frailty models. *Rostock, Germany Max Planck institute for demographic research*.
- Wienke, A. (2007). *Frailty models in survival analysis* (Habilitation zur Erlangung des akademischen Grades). Universität Halle-Wittenberg.
- Wienke, A. (2010). *Frailty models in survival analysis*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781420073911>.
- Wu, T., Feng, C., & Li, L. (2023). *A comparison of parameter estimation methods for shared frailty models* (Tech. Rep.). arXiv preprint. <https://arxiv.org/pdf/2311.11543>.