

UNIVERSIDAD DE GRANADA



UNIVERSIDAD  
DE GRANADA

DOCTORADO EN ESTADÍSTICA MATEMÁTICA Y APLICADA

---

**Aplicación de los Métodos Analíticos Matriciales  
en el estudio de la supervivencia al cáncer de  
vejiga**

---

**TESIS DOCTORAL**

Presentada por: Alicia Pereira das Neves Yedig

Dirigida por: Dra. Delia Montoro Cazorla

Bajo la tutoría de: Dr. Rafael Pérez Ocón y Dra. María del Mar Rueda García

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Editor: Universidad de Granada. Tesis Doctorales  
Autor: Alicia Pereira das Neves Yedig  
ISBN: 978-84-1306-933-3  
URI: <http://hdl.handle.net/10481/69650>



---

# Índice general

---

<b>Índice general</b>	<b>v</b>
<b>Introducción</b>	<b>xiv</b>
<b>1 Preliminares</b>	<b>1</b>
1.1 Modelos Markovianos multidimensionales y distribuciones tipo fase . . . . .	1
1.1.1 Procesos de Markov . . . . .	1
1.1.2 Modelos Markovianos multivariantes con uno o más de un estado absorbente . . . . .	8
1.2 Modelo de Cox de riesgos proporcionales . . . . .	14
1.3 Procesos de Renovación Markovianos y Procesos Semi-Markovianos . . . . .	17
1.4 Metodología . . . . .	20
<b>2 Un modelo multidimensional con tiempos no exponenciales en los estados</b>	<b>35</b>
2.1 Análisis de la información recogida en la base de datos . . . . .	35
2.2 Ajuste para la distribución de los tiempos de permanencia en cada estado . . . . .	40
2.3 Generador del Sistema y las principales medidas de la supervivencia a la enfermedad . . . . .	65
2.4 Sistema con las transiciones de los pacientes que derivan en progresión . . . . .	84
2.5 Sistema con las transiciones de los pacientes que salen del sistema sin progresión	105
2.6 Costes medios en función del tiempo . . . . .	131
<b>3 La supervivencia a la enfermedad en distintos grupos de riesgo</b>	<b>133</b>
3.1 El modelo espacio - estados . . . . .	133
3.2 Determinación de los distintos grupos de riesgo . . . . .	136
3.3 Análisis de los tiempos de transición entre estados . . . . .	142
3.4 Ajuste máximo verosímil del generador del sistema a partir del ingreso de covariables . . . . .	146
<b>4 Estimaciones con procesos semi-Markovinos</b>	<b>155</b>

---

4.1	Construcción de los estimadores de las principales medidas de supervivencia . . .	156
4.2	Estimaciones a partir de la base de datos . . . . .	162
4.3	Análisis de la consistencia de los estimadores propuestos . . . . .	166
	<b>Conclusiones</b>	<b>173</b>
	<b>Apéndice I</b>	<b>179</b>
	<b>Apéndice II</b>	<b>183</b>
	<b>Apéndice III</b>	<b>189</b>
	<b>Apéndice IV</b>	<b>197</b>
	<b>Apéndice V</b>	<b>201</b>
	<b>Índice de figuras</b>	<b>207</b>
	<b>Índice de tablas</b>	<b>217</b>
	<b>Bibliografía</b>	<b>219</b>

---

## Agradecimientos

---

A la directora de esta tesis Dra. Delia Montoro Cazorla, a quien ha desempeñado la labor de tutor hasta retirarse, Dr. Rafael Pérez Ocón y a la Dra. María del Mar Rueda García que la ha continuado, por haberme permitido ingresar en esta línea de investigación de extensa trayectoria y de aportes sustantivos a esta y otras áreas de conocimiento.

Al Dr. José Miguel Angulo Ibáñez y al Dr. Francisco de Asís Torres Ruiz, coordinadores del programa de Doctorado *Estadística Matemática y Aplicada* de la Universidad de Granada en el que se ha desarrollado este proyecto.

Destaco la colaboración de Nikolaos Limnios, quien con gran dedicación ha dirigido la extensión del proyecto inicial de trabajo, recibíendome en el Laboratoire de Mathématiques Appliquées de Compiègne (LMAC) y promoviendo el desarrollo de modelos alternativos para el abordaje del problema de investigación.

El desarrollo de este proyecto ha sido posible gracias a la movilidad que se me ha permitido realizar como funcionaria del *Consejo de Formación en Educación* de Uruguay, mi país de origen, ejerciendo mi trabajo como docente en forma telemática. Debo agradecer a quienes han colaborado de una u otra forma en la realización del mismo: a los funcionarios del *Hospital Universitario y Politécnico de la Fe de Valencia* y a los expertos que han realizado el registro y gestión de las bases de datos que hemos utilizado, pero además, a los investigadores de la *Fundación Canaria de Investigación y Salud* (FUNCIS) vinculada al Servicio de Evaluación del Servicio Canario de Salud (SESCS), por los aportes que nos han brindado y que responden a una extensa experiencia en la gestión de bases de datos vinculadas con este tema.

Agradezco especialmente a mis dos hijos Sofía y Mateo que han comprendido la importancia de esta tarea en mi desarrollo profesional, asumiendo cambios importantes a nivel familiar. A Juande, mi apoyo incondicional para concluir el proyecto. Es a ellos a quienes va dedicada mi tesis.



---

# Resumen

---

## Abstract

This thesis presents two state-space models for analyzing the evolution of the bladder cancer in two groups of patients. This cancer is one of the most extended in the population in many countries, it is highly aggressive, the treatment and following up of the patients is long and, consequently, the associated costs are very high in comparison with other types of cancer. The two state-space models applied to study the disease are Markovian processes: a multidimensional Markov process, based on the matrix-analytical methods and a semi-Markov process. Both procedures have different methodologies and allow to obtain information about the disease in two different and complementary ways.

The matrix-analytic methods are applied to a dataset. Based on the data and the characteristics of the disease, the states are defined and the staying empirical times in states are calculated. Phase-type distributions are fitted to these times with an adequate goodness of fit. Under these conditions, the Markovian structure is introduced to analyze the model, though the staying times did not follow exponential distributions. This shows the versatility and power of these methods. A multidimensional Markov process is constructed based on the phase-type distributions. The performance measures are calculated by using matrix calculations and presented in a algorithmic form that can be performed using computational algorithm.

In the analysis of the data, different groups of patients are distinguished in terms of the disease. The analysis and comparison of these groups is carried out introducing covariates in the model. This is a structure that has proved to be very useful in survival, and it has been applied when a Markov process governs the evolution of the system. We introduce them in the multidimensional case, extending the Cox model. In the multidimensional case, the covariates are affected by matrix coefficients. These parameters have been estimated using an adaptation of the Nelder-Mead algorithm to optimize the partial likelihood, which has allowed the entry of censored data. The procedure used allows to carry out later extensions of the study, with the estimation of the parameters from the optimization of more complex matrix functions. The analysis of the disease is completed and the groups are compared. The model is applicable even in the case of few data. The results contribute to a better knowledge of the disease.

The semi-Markov model uses a different methodology. Based on the data, the states are defined and the transition probabilities among states are estimated. In this case, the future evolution of the system depends on the occupied state at a certain instant and on the time since the last transition. The sojourn times in states can have arbitrary distributions. The model presented in an adequate goodness of fit to the real process, as a consequence of the strong consistency properties of the used estimators. The estimations have involved procedures that have been generalized and can be easily adapted to be used in other applications.

The proposed models and the estimation of the main survival measures, stand out relevant achievements and provide useful information to health care experts who will be able to plan preventive actions, specially when the survival time is large as we observe in this disease. They are constructed starting from a dataset, that is analyzed in detail, and then, two ways are applied to contribute to a better knowledge of the disease. In each case, the

algorithms are easily adaptable in other domains of applications, such as reliability, ensuring their applicability even despite the limitations of the available information.

## Resumen

Esta tesis presenta dos modelos multi-estados para analizar la evolución del cáncer de vejiga en dos grupos de pacientes. Este cáncer es uno de los más extendidos en la población en muchos países, es altamente agresivo, el tratamiento y seguimiento de los pacientes es largo y, en consecuencia, los costes asociados son muy elevados en comparación con otros tipos de cáncer. Los dos modelos espacio-estados aplicados al estudio de la enfermedad se basan en los procesos Markovianos: un proceso de Markov multidimensional con soporte en los métodos analíticos matriciales y un proceso semi-Markoviano. Ambos procedimientos tienen diferentes metodologías y permiten obtener información sobre la enfermedad siguiendo caminos diferentes.

Los métodos analíticos matriciales se aplican a un conjunto de datos. A partir de los datos y las características de la enfermedad se definen los estados y se calculan los tiempos empíricos de permanencia en ellos. Las distribuciones tipo fase se ajustan a los tiempos de permanencia en cada estado con una bondad de ajuste adecuada. En estas condiciones, la estructura de Markov se introduce para analizar el modelo, aunque los tiempos de permanencia no siguen distribuciones exponenciales. Esto muestra la versatilidad y el poder de estos métodos. Se construye un proceso multidimensional de Markov basado en las distribuciones tipo de fase. Las medidas de supervivencia son estimadas utilizando cálculos matriciales. Estos son presentados a través de algoritmos ejecutables computacionalmente.

En el análisis de los datos, se distinguen diferentes grupos de pacientes en función de la enfermedad. El análisis y la comparación de estos grupos se llevan a cabo introduciendo covariables en el modelo. Es una estructura que ha demostrado ser muy útil en la

supervivencia y se ha aplicado cuando un proceso de Markov gobierna la evolución del sistema. Los introducimos en el caso multidimensional, ampliando así el modelo de Cox. En el caso multidimensional, las covariables se ven afectadas por coeficientes que se registran como componentes de una matriz. Estos parámetros se han estimado mediante una adaptación del algoritmo de Nelder-Mead para optimizar la verosimilitud parcial, lo que ha permitido la entrada de datos censurados. El procedimiento utilizado permite realizar posteriores ampliaciones del estudio, con la estimación de los parámetros a partir de la optimización de funciones matriciales más complejas. El modelo es aplicable incluso en el caso de pocos datos. El análisis de la enfermedad correspondiente y la comparación de los grupos podría ampliarse, contribuyendo de esta forma a un mejor conocimiento de la misma.

El modelo basado en procesos semi-Markovianos utiliza una metodología diferente. A partir de una base de datos se definen los estados y se estiman las probabilidades de transición entre ellos. En este caso, la evolución futura del sistema depende del estado de ocupación en un determinado instante y del tiempo transcurrido desde la última transición. Los tiempos de estancia en los estados pueden tener distribuciones arbitrarias. El modelo presenta adecuada bondad de ajuste al proceso real de la enfermedad, como consecuencia de las fuertes propiedades de consistencia de los estimadores utilizados. Las estimaciones han involucrado procedimientos que han sido generalizados y pueden ser fácilmente adaptables para ser utilizados en otras aplicaciones.

Los modelos propuestos y la estimación de las principales medidas de supervivencia, destacan logros relevantes y aportan información útil a los expertos en salud que podrán llevar a cabo acciones preventivas, especialmente cuando el tiempo de supervivencia sea considerablemente extenso como en esta enfermedad. Se construyen a partir de un conjunto de datos que se analiza en detalle, luego se aborda el análisis de dos formas distintas para contribuir a un mejor conocimiento de la enfermedad. En cada caso, los algoritmos son

fácilmente adaptables en otros dominios como la fiabilidad, asegurando su aplicabilidad incluso a pesar de las limitaciones de la información disponible.



---

# Introducción

---

## Antecedentes del problema abordado

El cáncer de vejiga es un grave problema de salud en España, determinado por su alta incidencia y prevalencia de casos. La mayor incidencia del tumor vesical sin invasión del músculo en los países de Europa se produce en España, duplicando el número de casos respecto a los otros países. A su vez, es el tumor que genera mayores costes económicos de todas las formas en que el cáncer se presenta, debido a la baja mortalidad y a los largos períodos de seguimiento [10].

En el año 2012 el médico urólogo Saturnino Luján Marco [28] del Departamento de Cirugía de la Facultad de Medicina y Odontología de la Universidad de Valencia, presentó su tesis doctoral *Modelización matemática de la multirrecidiva y heterogeneidad individual para el cálculo del riesgo biológico de recidiva y progresión del tumor vesical no músculo invasivo*. En el campo de investigación interdisciplinaria, la *Biología de sistemas* tiene en cuenta las interacciones de los elementos internos y externos que influyen en el desarrollo de los procesos. Desde un enfoque holístico, se procuró comprender el funcionamiento de los sistemas biológicos (procesos) para entender cómo las interacciones internas y con otros sistemas conllevan la aparición de nuevas propiedades. El trabajo fue justificado ampliamente debido a la importancia de la toma de decisiones diagnósticas y terapéuticas. Eso motivó el desarrollo de tal investigación aunque se destaca el objetivo de lograr modelos

matemáticos que se ajusten a la naturaleza de los datos, teniendo en cuenta que suponer hipótesis no validadas adecuadamente generaría resultados alejados de la realidad clínica de los pacientes. El trabajo se desarrolló a partir de la misma base de datos que se usará en el presente estudio. Al reconocer las variables biológicas implicadas como continuas se propuso los puntos de corte más adecuados para establecer grupos de riesgo. Posteriormente se aplicó el *modelo de regresión de Cox* para predecir la variable dependiente *tiempo de primera recidiva tumoral* y en otro modelo el *tiempo de progresión*, en función de un conjunto de variables independientes elegidas a partir del método *stepwise*. De las siete variables clínicas y patológicas incluidas en la base de datos, cinco resultaron significativas en el modelo creado para la variable *tiempo de primera recidiva tumoral*: *edad*, *grado*, *número*, *tamaño* y *tratamiento*. Para el modelo de regresión cuya variable dependiente es el *tiempo de progresión* fueron significativas las variables *edad* y *grado*. El modelo propuesto en ese trabajo no contempló las repeticiones del suceso de interés (multirrecidiva tumoral) dado que, como lo señala el autor, puede existir correlación entre los tiempos de recurrencia.

Otro análisis del proceso enfermedad-muerte de los pacientes a partir de la misma base de datos [41] se basó en la creación de un modelo de sistema multiestado. Se manifestó en ese trabajo el interés por modelar el riesgo general de la progresión de la enfermedad que invade los músculos de la vejiga, brindando una distribución de probabilidades del tiempo que se tarda en alcanzarla. En el sistema, el estado 0 corresponde al momento en el que el paciente está libre de tumor, luego de su primera resección. El estado 1 corresponde a la primera recurrencia, en la que el paciente es diagnosticado con recidiva y el estado 2 representa la progresión de la enfermedad. El proceso de la enfermedad hasta su primera recidiva ó progresión se analiza en un modelo global, teniendo en cuenta en forma simultánea las posibles transiciones hacia uno u otro estado. En este caso se utiliza una mezcla finita de distribuciones de *Erlang*, incorporando covariables. Esta mezcla finita se relaciona finalmente con una distribución *tipo fase*, resultado que surge de considerar

la familia de estas distribuciones como una clase densa en el conjunto más amplio de distribuciones definidas en la recta real positiva.

El siguiente paso sería ampliar el modelo creado, teniendo en cuenta múltiples recurrencias o recidivas, sin suponer a priori que exista independencia entre las variables aleatorias que cuantifican el tiempo de permanencia en cada estado del sistema, o buenos ajustes con las distribuciones más usadas. En el presente trabajo de investigación se estudiará la evolución del cáncer de vejiga con tumor vesical no invasivo del músculo de la vejiga, en el mismo grupo de pacientes. Se utilizarán los *métodos analíticos matriciales* propuestos por Neuts [30] construyendo un modelo Markoviano multidimensional. Esto permitirá conocer con detalle la supervivencia a la enfermedad, caracterizando las distintas etapas, teniendo en cuenta no solamente la primera recidiva sino también las posteriores y la progresión.

En el Departamento de Estadística e Investigación Operativa de la Universidad de Granada, dentro de la línea de investigación *modelos Markovianos y fiabilidad de sistemas*, se han desarrollado distintos trabajos sobre la supervivencia a distintos tipos de cáncer:

- *A multivariate model to measure the effect of treatments in survival to breast cancer* (Pérez-Ocón, R. et. al. ; 1998). En este artículo, los autores presentan un proceso de Markov homogéneo en tiempo continuo con dos estados transitorios (recaída y no recaída) y uno absorbente (muerte) para analizar la supervivencia al cáncer de mama. El estudio se efectuó en una cohorte de 300 pacientes, categorizados en distintos grupos según tres tratamientos y las combinaciones entre ellos introducidos en el modelo como covariables dicotómicas [33].
- En *Markov models with lognormal transition rates in the analysis of survival times* (Pérez-Ocón, R. et. al. ; 2000), los autores presentan para la misma cohorte de 300 pacientes mujeres con cáncer de mama, un proceso de Markov no homogéneo con

iguales estados que los considerados en el estudio anterior y las mismas covariables, estimando los parámetros por regresión a partir de una distribución lognormal [34].

- En *Non-homogeneous Markov models in the analysis of survival after breast cancer* (Pérez-Ocón, R. et. al. ; 2001) se considera un análisis similar al anterior pero las estimaciones de los parámetros se realizan por regresión, a partir de una distribución de Weibull [35].
- En otro trabajo, *A piecewise Markov process for analysing survival from breast cancer in different risk groups* (Pérez-Ocón, R. et. al. ; 2001), los autores hacen un planteo similar a los anteriores, con regresión de Cox [36].
- En *A multiple-absorbent Markov process in survival studies: application to breast cancer* (Pérez-Ocón, R. y Ruiz-Castro, J. ; 2003) se crea un modelo con dos estados absorbentes con la aplicación de una distribución *tipo fase* para los tiempos de absorción a partir de una aproximación en la clase de distribuciones de este tipo [37].

En el estudio de la evolución de las enfermedades, el modelo de Cox [12] ha jugado un rol importante. En particular, el modelo de Cox de riesgos proporcionales [6, 7] ha sido aplicado en estudios en los que se consideran los efectos de ciertos factores de riesgo en el tiempo de supervivencia. La literatura sobre estos modelos o sus extensiones es amplia [1, 3, 4, 39], sin embargo, para las enfermedades con sucesivas recurrencias no son directamente aplicables. La identificación de diferentes niveles de la enfermedad a lo largo del tiempo, permite construir sistemas multi-estados que resultan más dinámicos, a partir de los cuales los procesos de Markov juegan un papel importante debido a sus propiedades analíticas y probabilísticas en conjunto con la posibilidad que brindan para el tratamiento computacional de las medidas de supervivencia. En el modelo de Cox las covariables pueden ser incorporadas en las razones de transición entre los estados, con coeficientes que pueden ser estimados por máxima verosimilitud. Algunos ejemplos de aplicación de estos

modelos pueden encontrarse en [20, 34–36, 44] y, a partir de procesos de Markov con varios estados absorbentes, que constituyen una antesala del presente trabajo [37]. En [42] se han utilizado otros modelos de estados múltiples bajo procesos semi-Markovianos, discutiendo la idoneidad y adecuación del método para el abordaje del problema específico.

## Sobre el presente trabajo

Esta tesis recoge los principales resultados que se han obtenido en el estudio de la evolución del cáncer de vejiga en distintos grupos de pacientes. Para su realización, se ha dispuesto de una base de datos principal, que ha sido elaborada a partir de los registros médicos correspondientes a los pacientes que han sido atendidos en el Hospital Universitario y Politécnico de La Fe de Valencia, durante el seguimiento de la enfermedad en sus distintas etapas. El período de registro estuvo comprendido entre el año 1995 y 2010.

Los modelos multidimensionales que se han aplicado han sido diseñados a partir de la caracterización clínica de las etapas de la enfermedad. Se ha podido constatar que las distribuciones usuales no presentan ajustes satisfactorios para los tiempos de permanencia en los estados de cada sistema, limitación que ha sido superada utilizando distribuciones tipo-fase (PH) a partir de las propiedades que las caracterizan.

A través de los *métodos analíticos matriciales* (MAM) de Neuts [30], la información censurada se incorporó en primer lugar de una manera versátil, agregando al estado absorbente que representa la progresión de la enfermedad, otro, que representa la salida del sistema pero por otras razones. La generalidad de este primer método da cuenta de su importancia en tanto que es aplicable a muchas otras enfermedades y otros procesos estocásticos.

Para completar el estudio de la enfermedad se ha generalizado el modelo a partir de una base de datos que, a pesar de su menor tamaño, cuenta con información adicional

sobre distintas covariables que aluden al diagnóstico inicial de los pacientes. Para ello, fue necesario adecuar el MAM con el objetivo de discutir los efectos de las covariables en el generador del modelo y, por lo tanto, en las principales medidas de supervivencia, estableciendo comparaciones entre distintos grupos de riesgo a partir de la construcción de un modelo global que depende de tales covariables. Según se ha constatado, estos efectos suelen estudiarse en forma aislada dentro de los distintos estratos, lo que limita la comparación entre los colectivos en cuestión.

Dadas las características específicas de la enfermedad, se consideró adecuado realizar un abordaje del problema con *procesos semi-Markovianos* (SMP) sobre un modelo multiestado con un único estado absorbente que representa la progresión de la misma, incorporando directamente la información censurada. Existe una amplia trayectoria de investigación acerca de las propiedades teóricas generales de este tipo de modelos, sin embargo, su implementación en el estudio de este tipo de enfermedades a partir de una base de datos específica y un modelo multi-estado concreto, ha generado la necesidad de construir los estimadores apropiados de las medidas de interés; lo que constituye un antecedente innovador. A su vez, el estudio de la consistencia de los estimadores empleados ofrece garantías de la bondad de cada ajuste.

En síntesis, se desarrollan y generalizan modelos aplicables a esta y a otras enfermedades para casos en los que el proceso se ajusta específicamente a un modelo SMP, según sus condiciones de hipótesis que no son demasiado restrictivas y se presentan además modelos que a partir de las propiedades de las distribuciones tipo PH, serán aplicables en el estudio de cualquier otra enfermedad, permitiendo el análisis de los efectos de covariables en la supervivencia de los pacientes. En cualquier caso, de existir información censurada, se ofrece alternativas para los estimadores propuestos con excelentes propiedades de consistencia y por tanto de ajuste a los procesos reales.

## Objetivos

El objetivo general de este trabajo de investigación es aplicar los métodos analíticos matriciales sobre modelos multidimensionales que caracterizan la evolución de una enfermedad. A partir de estos modelos, se estimarán las principales medidas de supervivencia de esta, brindando a los expertos en el cuidado de la salud la posibilidad de establecer actuaciones preventivas.

En forma específica, nos planteamos :

- Aplicar los MAM en el estudio longitudinal del cáncer de vejiga a partir de dos bases de datos que incluyen información relevante del seguimiento realizado a pacientes con esta enfermedad.
- Discutir la aplicabilidad de los modelos paramétricos más usados en supervivencia y proponer alternativas a partir de las propiedades destacadas de las distribuciones PH.
- Presentar modelos alternativos a partir de SMP, con la consecuente construcción de los estimadores adecuados al sistema multi-estado específico y ulterior discusión de propiedades.
- Incorporar toda la información disponible con datos censurados para optimizar los ajustes logrados en cada modelo.
- Identificar covariables que puedan tener efectos sobre las medidas de supervivencia en el contexto del problema abordado.
- Adecuar los MAM para establecer comparaciones en distintos grupos de riesgo, ingresando las covariables seleccionadas en un modelo generalizado.
- Expresar los procedimientos involucrados en forma algorítmica para su implementación computacional en el estudio de la supervivencia a esta u otras enfermedades a partir de bases de datos alternativas.

## **Organización**

La tesis se organiza de la siguiente manera: el Capítulo 1 presenta los preliminares necesarios para el desarrollo posterior y la metodología de trabajo. Los principales resultados de la investigación se exponen en los Capítulos 2, 3 y 4. Las conclusiones y exposición de líneas de ampliación de la investigación, se incluyen al final del trabajo. Los apéndices muestran todos los procedimientos algorítmicos que se han aplicado para obtener los resultados de cada capítulo.

# Preliminares



---

# Preliminares

---

## 1.1. Modelos Markovianos multidimensionales y distribuciones tipo fase

### 1.1.1. Procesos de Markov

Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad.  $X(t) : \Omega \rightarrow \mathbb{R}$  una variable aleatoria definida en el instante  $t$  cuyo valor nos indicará el estado del sistema en ese momento. En nuestro caso, el conjunto de índices  $I$  es  $[0, \infty]$ , dado que se trata de un proceso estocástico en tiempo continuo.

Sea  $E$  un conjunto numerable. Un proceso estocástico  $(X(t), t \geq 0)$  con valores en  $E$  se dice que es una cadena de Markov en tiempo continuo si para todos los valores  $s, t \geq 0$  e  $i, j, i_u \in E$  se verifica:

$$P\{X(s+t) = j \mid X(s) = i, X(u) = i_u, 0 \leq u < s\} = P\{X(s+t) = j \mid X(s) = i\}$$

El conjunto  $E$  se denomina *espacio de estados de la cadena*. En el caso en el que  $E$  es finito, la cadena se dice *cadena finita*.

Al trabajar en tiempo continuo se exige condiciones de continuidad [31], como por ejemplo:

$$\forall \omega \in \Omega \text{ y } t \geq 0 : (\exists \varepsilon > 0 \mid (X(s)(\omega) = X(t)(\omega), \forall s \mid t \leq s \leq t + \varepsilon))$$

La *cadena de Markov* se dice *homogénea*, si además para todos los valores  $s, r, t \geq 0$  e  $i, j \in E$  se verifica:

$$P\{X((s+r)+t) = j \mid X(s+r) = i\} = P\{X(s+t) = j \mid X(s) = i\}$$

Esta propiedad permite caracterizar en una *cadena de Markov homogénea* la probabilidad de pasar del estado  $i$  al estado  $j$  habiendo transcurrido un tiempo  $t$ , independiente del instante en que ocurre la transición. De esta forma la probabilidad de transición del estado  $i$  al  $j$  en un tiempo  $t$  es:

$$p_{ij}(t) = P\{X(s+t) = j \mid X(s) = i\} = P\{X(t) = j \mid X(0) = i\}, \forall t, s \geq 0, \forall i, j \in E$$

Queda así definida la matriz cuadrada de probabilidades de transición entre estados, en un tiempo  $t$ :

$$P(t) = (p_{ij}(t)), i, j \in S$$

En otras palabras, una *cadena de Markov en tiempo continuo* es un *proceso estocástico en tiempo continuo* que satisface la propiedad de Markov, esto es, que la distribución condicional de un estado en un tiempo  $s+t$ , dado el estado presente en  $s$  y todos los estados

pasados, depende solo del estado presente y es independiente del pasado. Si además, la probabilidad  $P\{X(s+t) = j | X(s) = i\}$  es independiente del instante de evaluación  $s$ , se dice que la *cadena es homogénea*.

Otra forma de caracterizar una *cadena de Markov* es a partir de  $T_i$  la variable aleatoria que hace alusión al tiempo que la cadena permanece en el estado  $i$ , antes de efectuar una transición a otro estado. Por la propiedad de Markov, la probabilidad de que la cadena permanezca en  $i$  durante el lapso  $[s, s+t]$  sabiendo que en  $s$  estaba en el estado  $i$ , será igual a la probabilidad (no condicional) de permanecer en el estado  $i$  un tiempo mayor a  $t$ , sin memoria de lo que ocurrió antes:

$$P\{T_i > s+t | T_i > s\} = P\{T_i > t\}, \forall t, s \geq 0$$

A partir de lo anterior, aplicando la definición de probabilidad condicional y de la función de distribución para una variable aleatoria se tiene:

$$\frac{P\{T_i > s+t, T_i > s\}}{P\{T_i > s\}} = P\{T_i > t\}, \forall t, s \geq 0$$

$$\frac{P\{T_i > s+t\}}{P\{T_i > s\}} = P\{T_i > t\}, \forall t, s \geq 0$$

$$\frac{1 - F_{T_i}(s+t)}{1 - F_{T_i}(s)} = 1 - F_{T_i}(t), \forall t, s \geq 0$$

Definiendo la función  $\bar{F}_{T_i} = 1 - F_{T_i}$ , se tiene:

$$\frac{\bar{F}_{T_i}(s+t)}{\bar{F}_{T_i}(s)} = \bar{F}_{T_i}(t), \forall t, s \geq 0$$

$$\bar{F}_{T_i}(s+t) = \bar{F}_{T_i}(t)\bar{F}_{T_i}(s), \forall t, s \geq 0$$

La solución, que además es única, de esa ecuación funcional corresponde a la función  $F$  que es la de distribución de una variable aleatoria exponencial. Así se tiene otra forma de caracterizar una *cadena de Markov*:

1. El tiempo que permanece en ese estado antes de hacer una transición a otro estado distinto es exponencial con tasa  $q_i$ , que luego interpretaremos.
2. Cuando el proceso abandona el estado  $i$ , al cabo de un tiempo  $t$  entra en otro estado  $j$  con una probabilidad  $p_{ij}(t)$  tal que  $\sum_{j \neq i, j \in E} p_{ij}(t) = 1$

Una *cadena de Markov en tiempo continuo* es un *proceso estocástico* que se mueve entre los estados de un conjunto  $E$ , de tal forma que antes de moverse del estado  $i$  a otro estado  $j$  del conjunto de estados  $E$ , permanece en  $i$  un tiempo con distribución exponencial y, por supuesto, el tiempo que permanece en el estado siguiente  $j$  queda descrito por una variable aleatoria independiente de la anterior, que también tiene distribución exponencial.

Dado que:

$$p_{ij}(t) = P\{X(s+t) = j | X(s) = i\} = P\{X(t) = j | X(0) = i\}, \forall t, s \geq 0, \forall i, j \in E$$

Es posible definir la *matriz de probabilidades de transición* en un tiempo  $t$ :

$$P(t) = (p_{ij}(t))$$

Estas probabilidades de transición, cumplen las siguientes propiedades:

- $p_{ij}(t) \geq 0, \forall t > 0$

- $\sum_{j \neq i, j \in E} p_{ij}(t) = 1, \forall t > 0$
- Ecuaciones de Chapman-Kolmogorov: para pasar del estado  $i$  al estado  $j$  en un tiempo  $w$  se puede hacerlo a través de cualquier estado  $k$  en un tiempo intermedio. En efecto, para cualquier descomposición de  $w = s + t$  se tiene en virtud de la independencia entre unas y otras transiciones consecutivas y por aditividad de la función probabilidad que:

$$p_{ij}(s+t) = \sum_{k \in E} p_{ik}(s)p_{kj}(t), \forall t, s > 0$$

- Como hipótesis de continuidad, asumiremos que:

$$\lim_{t \rightarrow 0^+} p_{ij}(t) = \delta_{ij} = \begin{cases} 1 & , si i = j \\ 0 & , si i \neq j \end{cases}$$

Asumir el último enunciado como condición de continuidad presupone que en  $t = 0$  los valores de las probabilidades de transición entre estados son los de la matriz identidad:  $P(0) = I$ . De esta forma, a partir de la condición anterior se entiende que los límites de las funciones  $p_{ij}(t)$  cuando  $t$  tiende a 0, existen y son los propios valores  $p_{ij}(0)$ :

$$p_{ij}(0) = \begin{cases} 1 & , i = j \\ 0 & , i \neq j \end{cases}$$

A partir de las propiedades anteriores es posible demostrar los siguientes teoremas que garantizan la diferenciabilidad de las probabilidades de transición:

1. Para cada  $i \in E$ , el siguiente límite existe, aunque puede ser infinito:

$$\lim_{t \rightarrow 0} \frac{1-p_{ii}(t)}{t} = \lim_{t \rightarrow 0} \frac{p_{ii}(0)-p_{ii}(t)}{t-0} = -p'_{ii}(0) = q_i > 0$$

En virtud del cociente que ha definido a  $q_i$  es posible interpretarlo como la tasa de transiciones del estado  $i$  a cualquier otro estado, por unidad de tiempo. Téngase en cuenta que la probabilidad de un suceso, en este caso el de transitar de  $i$  a cualquier otro estado diferente en  $E$ , conceptualmente está vinculada a la tasa (tanto por uno)

con que ocurre dicho suceso.

A  $q_i$  también se lo conoce como *razón de escape del estado  $i$* .

2. Para cada par de estados  $i, j \in E$ ,  $i \neq j$ , el siguiente límite existe y es finito:

$$\lim_{t \rightarrow 0} \frac{p_{ij}(t)}{t} = \lim_{t \rightarrow 0} \frac{p_{ij}(t) - p_{ij}(0)}{t - 0} = p'_{ij}(0) = q_{ij} \geq 0$$

$q_{ij}$ , en relación a cómo ha quedado definido, representa la razón de transición del estado  $i$  al  $j$  por unidad de tiempo. Desde el punto de vista de probabilidades, hace alusión a la tasa (tanto por uno) de transición de  $i$  a  $j$ , por unidad de tiempo.

El *generador asociado al proceso* es la matriz  $Q$  cuyos elementos se obtienen a partir de la *matriz de probabilidades de transición* derivando las correspondientes funciones  $p_{ij}(t)$ , para cada par  $(i, j)$  en  $t = 0$ :

$$Q = (q_{ij}), \text{ donde } q_{ij} = \begin{cases} -q_i = p'_{ii}(0) & , i \in E, i = j \\ q_{ij} = p'_{ij}(0) & , i, j \in E, i \neq j \end{cases}$$

Para cada  $i \in E$  se cumple que:

$$\begin{aligned} \sum_{j \in E, j \neq i} q_{ij} &= \sum_{j \in E, j \neq i} \lim_{t \rightarrow 0} \frac{p_{ij}(t)}{t} \\ &= \lim_{t \rightarrow 0} \sum_{j \in E, j \neq i} \frac{p_{ij}(t)}{t} \\ &= \lim_{t \rightarrow 0} \frac{1}{t} \sum_{j \in E, j \neq i} p_{ij}(t) \\ &= \lim_{t \rightarrow 0} \frac{1}{t} (1 - p_{ii}(t)) \\ &= q_i \end{aligned}$$

Téngase en cuenta que límite y sumatoria se pueden intercambiar en virtud del último teorema en el que se enuncia la existencia de cada límite y que son finitos.

El resultado anterior garantiza que la suma de las filas de la matriz generador  $Q$  es 0.

La cadena se denomina conservativa si se cumple:

$$\sum_{j \in E, j \neq i} q_{ij} = q_i < \infty, \forall i \in E$$

La condición de conservatividad expresa que la razón de escape o salida del estado  $i$  es igual a la suma de todas las razones de llegadas desde  $i$  a cualquier otro estado.

Las cadenas conservativas, cumplen las *ecuaciones diferenciales hacia atrás de Kolmogorov* y también las *ecuaciones diferenciales hacia adelante*, fácilmente demostrables a partir de las definiciones que hemos considerado antes y de la propiedad propia de una *cadena conservativa* esto es:

1. Ecuaciones *Backward de Kolmogorov*:

$$p'_{ij}(t) = \sum_{k \neq i} q_{ik} p_{kj}(t) - q_i p_{ij}(t), \forall i, j \in E$$

2. Ecuaciones *Forward de Kolmogorov*:

$$p'_{ij}(t) = \sum_{k \neq j} p_{ik}(t) q_{kj} - p_{ij}(t) q_j, \forall i, j \in E$$

Se puede verificar que las ecuaciones anteriores pueden sintetizarse de la siguiente forma:

$$\begin{cases} P'(t) = QP(t) & \text{Ecuaciones Backward de Kolmogorov} \\ P'(t) = P(t)Q & \text{Ecuaciones Forward de Kolmogorov} \end{cases}$$

La solución de esas ecuaciones diferenciales matriciales nos permitirá determinar los valores de probabilidades de transición al cabo de un tiempo  $t$ , que conforman la matriz  $P$ . Si  $P(t)$  es una función del tiempo y  $Q$  es una matriz constante, la solución es:

$$P(t) = e^{Qt} = \sum_{j=0}^{\infty} \frac{(Qt)^j}{j!}, \text{ donde } Q^0 \text{ se define como la matriz identidad } I$$

En resumen, la solución de las ecuaciones proporciona las probabilidades  $P(t)$  de la cadena y si los  $q_i$  son acotados o el espacio de estados es finito las probabilidades  $P(t)$  se podrán obtener a través de la expresión que hemos planteado antes.

### 1.1.2. Modelos Markovianos multivariantes con uno o más de un estado absorbente

En la teoría de Supervivencia la distribución exponencial surge muchas veces como el modelo adecuado para los tiempos aleatorios de permanencia en ciertos estados. El estudio de los procesos de Markov es esencial, ya que una de las características de estos modelos es el buen ajuste a este tipo de distribuciones. Sin embargo, muchas situaciones donde interesa el análisis de la supervivencia no pueden ser representadas con un ajuste satisfactorio a partir de ellos.

Una primera generalización surge de considerar etapas o fases exponencialmente distribuidas en el tiempo. Si las fases se organizan en serie, se tiene la distribución de Erlang, convolución de las distribuciones exponenciales de las etapas, independientes e idénticas. Cuando las etapas se organizan en paralelo, surge la distribución hiperexponencial, que es una mezcla finita de distribuciones exponenciales. Cuando la organización de las fases no es de uno ni otro tipo, es posible proponer un modelo de distribución PH, o aproximar la distribución tanto como se quiera, con una distribución de este tipo [30].

- Procesos de Markov con un estado absorbente:

Sea  $\{X(t), t \geq 0\}$  un proceso de Markov con espacio de estados  $S = \{1, \dots, m, m+1\}$ , continuo a la derecha, donde los primeros  $m$  estados son transitorios y el último estado  $m+1$  es absorbente.

El generador infinitesimal del proceso que representa las intensidades o razones de transición entre estados, puede ser escrito de la forma:

$$Q = \begin{bmatrix} A & A^0 \\ 0 & 0 \end{bmatrix}$$

La matriz  $A = (a_{ij})$  que contiene las intensidades o razones de transición entre los estados transitorios, es cuadrada de orden  $m$  y sus elementos cumplen:

$$\begin{cases} a_{ii} < 0, \forall i \in \{1, \dots, m\} \\ a_{ij} > 0, (\forall i, j \in \{1, \dots, m\} \wedge i \neq j) \end{cases}$$

A su vez la matriz  $A^0$  que contiene las razones de absorción por parte del estado absorbente  $m+1$ , desde cada uno de los estados transitorios verifica:

$$Ae + A^0 = \theta$$

ya que es conservativa, siendo  $e$  una matriz columna de  $m$  elementos iguales a 1 y  $\theta$  otra matriz columna de  $m$  elementos iguales a 0.

El hecho de que los  $m$  primeros estados sean transitorios, supone en términos prácticos que la absorción por parte del estado  $(m+1)$  es cierta desde cualquier estado inicial. Esto se satisface sí y solo si  $A$  es no singular.

Sea  $(\alpha, \alpha_{m+1})$  el vector que contiene las probabilidades de inicio del proceso en cada uno de los estados, con  $\alpha$  un vector de  $m$  componentes y  $\alpha_{m+1}$  un real tales que:

$$\alpha_i = p(X(0) = i), i \in \{1, \dots, m\} \text{ y}$$

$$\alpha e + \alpha_{m+1} = 1$$

En general podremos suponer que  $\alpha_{m+1} = 0$ , dado que la probabilidad de iniciar el proceso en el estado absorbente suele ser 0.

Si  $\tau$  representa el tiempo de absorción del proceso por el estado absorbente:

$$\tau = \inf_t \{t \geq 0 : X(t) = m + 1\}$$

Puede demostrarse que la distribución de probabilidades para dicha variable continua es:

$$F(t) = 1 - \alpha \exp(At) e, \quad t \geq 0,$$

donde  $A(t) = \exp(At)$  es la matriz de probabilidades de transición entre los  $m$  estados transitorios que se obtiene como solución de las ecuaciones adelantadas de Kolmogorov,  $\alpha \exp(At) e$  es la probabilidad de permanecer en el conjunto de estados transitorios, teniendo en cuenta las probabilidades de inicio del proceso contenidas en  $\alpha$ . Es así que con tal expresión,  $F(t)$  representa la probabilidad de absorción del proceso.

La probabilidad de que el proceso no haya sido absorbido por el estado  $m + 1$  en el instante  $t$  tiene especial importancia cuando se modela el proceso de una

enfermedad, en cuyo caso el estado absorbente representa la muerte del paciente. Dicha probabilidad recibe el nombre de *supervivencia* en el instante  $t$ :

$$S(t) = \alpha \exp(At) e, \quad t \geq 0$$

Una variable aleatoria  $\tau$  con distribución de probabilidad dada por  $F$  se dice que es *tipo fase* y queda caracterizada por  $\alpha$  y  $A$ . La denotaremos entonces  $PH(\alpha, A)$

La densidad del tiempo de absorción se obtiene derivando la expresión matricial de la función  $F(t)$ :

$$f(t) = \alpha \exp(At) A^0, \quad t \geq 0$$

La razón de fallo es:

$$\begin{aligned} h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{\alpha \exp(At) A^0}{\alpha \exp(At) e}, \quad t \geq 0 \end{aligned}$$

- Procesos de Markov con más de un estado absorbente:

Sea  $\{X(t), t \geq 0\}$  un proceso de Markov con espacio de estados  $S = W \cup F$ , con  $W = \{1, \dots, m\}$  y  $F = \{m+1, \dots, m+k\}$ , donde  $W$  y  $F$  son los conjuntos de estados transitorios y absorbentes respectivamente.

Con los estados ordenados apropiadamente se tiene:

$$Q = \begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix}$$

Donde  $A$  denota la matriz cuadrada de orden  $m$  que contiene las intensidades o razones de transición entre los estados transitorios. Sin embargo,  $B$  es una matriz de orden  $m \times k$  que contiene las razones de absorción por parte de los  $k$  estados absorbentes, desde los  $m$  estados transitorios.

También en este caso la matriz  $Q$  es no singular y conservativa, por lo que la suma de las entradas de cada fila, es cero.

Si bien el tiempo de absorción por algún estado absorbente como variable aleatoria no presenta distribución del tipo PH, puede aproximarse con una distribución PH.

La solución a la ecuación matricial siguiente (donde  $Q$  es la matriz que se ha presentado antes),

$$P(t) = \exp \{Qt\},$$

tiene la forma:

$$P(t) = \begin{bmatrix} A(t) & B(t) \\ 0 & I \end{bmatrix} = \exp \{Qt\} = \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n$$

$A(t)$  contiene las probabilidades de transición entre los estados transitorios en un tiempo  $t$  y es por lo tanto solución de la ecuación matricial, en la que  $A$  tiene por entradas las razones de transición entre los estados transitorios:

$$A(t) = \exp \{At\}$$

Luego se tiene:

$$A(t) = \exp\{At\} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n$$

La entrada  $b_{ij}$  de la matriz  $B(t)$ ,  $\forall i \in \{1, \dots, m\}$ ,  $\forall j \in \{1, \dots, k\}$ , representa la función de distribución para la variable aleatoria tiempo de absorción por parte del estado  $j$ , habiendo iniciado el proceso en el estado transitorio  $i$  y puede obtenerse a partir de la siguiente expresión:

$$B(t) = \exp\{Bt\} = \sum_{n=1}^{\infty} \frac{t^n}{n!} A^{n-1} B = \left( \int_0^t A(u) du \right) B$$

Resolviendo la integral anterior se tiene:

$$B(t) = A^{-1} (\exp(At) - I) B, t \geq 0$$

Si  $\alpha$  es la matriz fila de  $m$  entradas que contiene las probabilidades de iniciar el proceso en cada uno de los estados transitorios, la siguiente matriz fila de orden  $1 \times k$  contiene las probabilidades de absorción por parte de cada uno de los  $k$  estados absorbentes en un tiempo  $t$ :

$$\alpha B(t), t \geq 0$$

Como la entrada  $h$  ( $h \in \{1, \dots, k\}$ ) de la matriz anterior  $A_h(t)$  contiene la probabilidad de absorción por parte del estado  $m+h$  en un tiempo  $t$ , se tiene que la siguiente matriz contiene las probabilidades de que no ocurran tales absorciones:

$$e - \alpha B(t), t \geq 0$$

donde  $e$  es una matriz fila con entradas 1 y dimensión  $1 \times k$ .

De la teoría general de procesos de Markov, es conocido que el tiempo medio  $\tau_{ij}(t)$  en el que el proceso permanece en un estado  $j$  partiendo de un estado  $i$  en un intervalo  $(0,t)$  viene dado por:

$$\tau_{ij}(t) = \int_0^t p_{ij}(u) du$$

Es posible definir una matriz, que contiene estos tiempos medios, para cada par de estados transitorios  $(i, j)$ :

$$TM(t) = \int_0^t P(u) du = \sum_{n=0}^{\infty} \frac{t^{n+1}}{(n+1)!} Q^n = \begin{bmatrix} N(t) & M(t) \\ 0 & tI \end{bmatrix}$$

Puede demostrarse que las matrices  $N(t)$  y  $M(t)$  tienen la forma:

$$N(t) = A^{-1} (A(t) - I) \quad \text{y} \quad M(t) = A^{-1} (B(t) - tB)$$

El elemento  $(i, j)$  de  $N(t)$ , es el tiempo esperado en el que el proceso pasa en el estado transitorio  $j$  hasta el tiempo  $t$ , partiendo del estado transitorio  $i$  en el tiempo 0.

El elemento  $(i, j)$  de  $M(t)$ , es el tiempo esperado en el que el proceso pasa en el estado absorbente  $j$  hasta el tiempo  $t$ , partiendo del estado transitorio  $i$  en el tiempo 0.

## 1.2. Modelo de Cox de riesgos proporcionales

Cuando se accede a información de supervivencia a una enfermedad, en general se accede a información adicional de cada individuo, así como la edad, sexo, valores de variables fisiológicas, si recibe otro u otros tratamientos, etc. Todos estos factores

pueden influir en la supervivencia del paciente, por lo cual tiene sentido incluir esta información en el modelo para el análisis de supervivencia, estos datos se incluyen como variables explicativas, llamadas covariables. Es posible usar el *Modelo de Cox de riesgos proporcionales*. Para ello se considera la función de riesgo  $\lambda(\cdot|z)$  en un tiempo  $t$  y para un individuo con vector de covariables  $z$ , con  $t \in \mathbb{R}^+$  y  $z \in \mathbb{R}^d$ :

$$\lambda(t|z) = \lambda_0(t)\phi(\beta, z)$$

En la expresión anterior  $\lambda_0(t)$  es una función arbitraria no paramétrica (aunque se puede suponer paramétrica como Weibull o Exponencial) y consideraremos la función nexo de las variables explicativas  $\phi(\beta, z) = \exp(\beta' z)$  siendo  $\beta$  el vector de parámetros del modelo, obteniendo:

$$\lambda(t|z) = \lambda_0(t) \exp(\beta' z)$$

Es posible observar que si se calcula la razón de riesgos para distintos valores de las covariables  $z_i, z_j$  se tiene:

$$\begin{aligned} \rho(t|\beta, z_i, z_j) &= \frac{\lambda(t|z_i)}{\lambda(t|z_j)} \\ &= \frac{\lambda_0(t) \exp(\beta' z_i)}{\lambda_0(t) \exp(\beta' z_j)} \\ &= \frac{\exp(\beta' z_i)}{\exp(\beta' z_j)} \\ &= \exp(\beta' (z_i - z_j)) \end{aligned}$$

La última expresión corresponde al *riesgo relativo* de un individuo con covariables  $z_i$  respecto a otro con covariables  $z_j$ .

### Función de verosimilitud parcial

Sea un conjunto de observaciones de valores de supervivencia en una muestra de tamaño  $n$  donde puede existir censura por la derecha, esto es, se tiene una muestra iid de vectores  $(X_i, \delta_i, Z_i)$  donde  $X_i = \min(T_i, C_i)$ ,  $\delta_i = \mathbf{1}_{X_i=C_i}$ , con  $T_i$  y  $C_i$  el tiempo de supervivencia o censura respectivamente y  $\delta_i$  la variable indicatriz que expresa si lo que se observa es un tiempo de censura o de supervivencia. También se observa el vector de covariables  $Z_i$  para cada individuo  $i$  con  $i \in \{1, \dots, n\}$ .

Supondremos que hay  $r$  tiempos de supervivencia diferentes y  $n - r$  tiempos censurados por derecha, con  $t_1 < t_2 < \dots < t_r$ . Para  $t_j$  con  $j \in \{1, \dots, r\}$ , sea  $R(t_j)$  el conjunto de individuos en riesgo de morir al tiempo  $t_j$ , formado por los individuos con vida y sin censura antes de dicho tiempo. Nótese que conforme incrementamos el valor de  $t_j$ , dicho conjunto será menor, en el sentido de la inclusión.

La función de *verosimilitud parcial* queda determinada por los productos de las probabilidades puntuales para los tiempos de supervivencia, dado que se supondrá que la función de riesgo es cero en los instantes que no se corresponden con éstos:

$$\begin{aligned} L(\beta) &= \prod_{j=1}^{j=r} \left\{ \exp(\beta' z_j) / \left[ \sum_{l \in R(t_j)} \exp(\beta' z_l) \right] \right\} \\ &= \prod_{i=1}^{i=n} \left\{ \exp(\beta' z_i) / \left[ \sum_{l \in R(x_i)} \exp(\beta' z_l) \right] \right\}^{1-\delta_i} \end{aligned}$$

La función *log-verosimilitud parcial* viene dada por:

$$\log(L(\beta)) = \sum_{i=1}^{i=n} (1 - \delta_i) \left[ \beta' z_i - \log \sum_{l \in R(x_i)} \exp(\beta' z_l) \right]$$

Cuando se dispone de una base de datos de supervivencia con algunos tiempos censurados y tiempos de supervivencia repetidos, será necesario ajustar la verosimilitud para  $\beta$  y en consecuencia la log-verosimilitud. Si el número de individuos que mueren  $d_i$  es pequeño comparado con el número de individuos en riesgo, para cada tiempo  $t_i$  de supervivencia, con  $i \in \{1, \dots, k\}$ , es posible aplicar el ajuste de la verosimilitud parcial propuesto por Breslow [11]:

$$\prod_{i=1}^k \frac{\exp(\beta' S_i)}{[\sum_{l \in R(t_i)} \exp(\beta' Z_l)]^{d_i}}$$

Siendo  $S_i$  la suma de los vectores de covariables para los  $d_i$  individuos que mueren en  $t_i$ , quienes generan los empates en dicho instante.

### 1.3. Procesos de Renovación Markovianos y Procesos Semi-Markovianos

Sea el conjunto de estados finito  $E = \{1, \dots, s\}$  y un proceso estocástico de valores en  $E$ :  $Z = (Z_t, t \in R^+)$ , de forma tal que:

- Los tiempos de saltos de  $Z$  están dados por  $S_0 = 0 \leq S_1 \leq \dots \leq S_n \leq S_{n+1} \dots$
- Los estados visitados en  $Z$  son respectivamente  $J_0, J_1, \dots, J_n, J_{n+1}, \dots$
- Los tiempos de permanencia en cada estado, entre saltos, son  $X_0 = 0$  y en general:  
 $X_n = S_n - S_{n-1}$ , con  $n \geq 1$ .

El proceso estocástico  $(S_n, J_n)$ , se dice proceso de renovación de Markov con espacio de estados  $E$  si la siguiente relación se verifica *casi seguramente*:

$$\begin{aligned} p(J_{n+1} = j, S_{n+1} - S_n \leq t \mid J_0, \dots, J_n, S_1, \dots, S_n) = \\ p(J_{n+1} = j, S_{n+1} - S_n \leq t \mid J_n), \forall j \in E, \forall t \in \mathbb{R}^+ \end{aligned}$$

El proceso  $Z = (Z_t, t \in \mathbb{R}^+)$  en tal caso se denomina *proceso estocástico Semi-Markoviano* (ver [14], [22] y [23]).

El proceso  $(J_n, S_n)$  es una cadena de Markov, con espacio de estados  $E \times \mathbb{R}^+$  y *Núcleo de transición* denominado *Núcleo de transición de Markov*  $Q_{ij}(t)$  que se define de la siguiente forma:

$$Q_{ij}(t) = p(J_{n+1} = j, S_{n+1} - S_n \leq t \mid J_n = i), \forall i, j \in E, \forall t \geq 0$$

$(J_n)$  es una cadena de Markov con espacio de estados  $E$  homogénea en el tiempo, a partir de lo cual se pueden definir las probabilidades de transición  $p_{ij} = Q_{ij}(+\infty) = p(J_{n+1} = j \mid J_n = i)$ . Diremos que  $(J_n)$  es una *cadena de Markov embebida en Z*. Siendo  $F_{ij}(t) = p(X_{n+1} \leq t \mid J_{n+1} = j, J_n = i)$  la *función de distribución para el tiempo entre saltos*, se tiene por propiedades de probabilidad condicional:

$$\begin{aligned} F_{ij}(t) &= \frac{p(X_{n+1} \leq t, J_{n+1} = j \mid J_n = i)}{p(J_{n+1} = j \mid J_n = i)} = \frac{Q_{ij}(t)}{p_{ij}} \\ Q_{ij}(t) &= F_{ij}(t)p_{ij}, \forall i, j \in E, \forall t \in \mathbb{R}^+ \end{aligned}$$

Es posible definir la *función de distribución para el tiempo de permanencia* en el estado  $i \in E$  de la siguiente forma:

$$H_i(t) = \sum_{j \in E} p(J_{n+1} = j, S_{n+1} - S_n \leq t \mid J_n = i) = \sum_{j \in E} Q_{ij}(t)$$

El valor medio  $m_i = E(H_i(t))$  se interpreta entonces como el valor medio o valor esperado de permanencia en el estado  $i$ .

Para un proceso Markoviano con generador  $A = (a_{ij})$  cuyos elementos representan las razones de transición entre estados, el *Núcleo* correspondiente se obtiene como:

$$Q_{ij}(t) = \frac{a_{ij}}{a_i} (1 - e^{-a_i t}), \text{ con } a_i = -a_{ii}, \forall i \in E$$

El proceso  $(N(t), t \in \mathbb{R}^+)$  define el número de saltos de  $Z$  en el intervalo  $(0, t]$ :

$$N(t) = \sup \{n \in \mathbb{N} : S_n \leq t\} = \# \{n \in \mathbb{N}^* : S_n \leq t\}$$

A partir de dicho proceso se definen respectivamente el *número de visitas a un estado  $i$*  de  $E$ , hasta el instante  $t$ ,  $N_i(t)$  y el *número de saltos entre los estados  $i, j$*  hasta el tiempo  $t$ ,  $N_{ij}(t)$ :

$$N_i(t) = \sum_{h=1}^{h=N(t)} \mathbb{1}_{\{J_h=i\}} = \sum_{h=1}^{h=+\infty} \mathbb{1}_{\{J_h=i, S_h \leq t\}}$$

$$N_{ij}(t) = \sum_{h=1}^{h=N(t)} \mathbb{1}_{\{J_{h-1}=i, J_h=j\}} = \sum_{h=1}^{h=+\infty} \mathbb{1}_{\{J_{h-1}=i, J_h=j, S_h \leq t\}}$$

Si  $\phi(i, t)$ , con  $i \in E$  y  $t \in \mathbb{R}$ , es una función real medible, definimos la *convolución* de  $\phi$  y  $Q$  de la siguiente forma:

$$Q * \phi(i, t) = \sum_{k \in E} \int_0^t Q_{ik}(ds) \phi(k, t-s) = \sum_{k \in E} \int_0^t Q_{ik}(s) \phi(k, t-s) ds$$

Si consideramos la  $n$ -convolución de  $Q$  consigo mismo para cada par  $i, j \in E$ , se tiene:

$$Q_{ij}^{*(n)}(t) = \begin{cases} \sum_{k \in E} \int_0^t Q_{ik}(ds) Q_{kj}^{*(n-1)}(t-s) & , n \geq 2 \\ Q_{ij}(t) & , n = 1 \\ \mathbf{1}(t) = \begin{cases} 1, & \text{if } t \geq 0 \\ 0, & \text{if } t < 0 \end{cases} & , n = 0 \end{cases}$$

Es posible probar que:

$$Q_{ij}^{*(n)}(t) = p_i(J_n = j, S_n \leq t)$$

Una función de significativa utilidad es la *función de distribución marginal condicional del proceso*:

$$p_{ij}(t) = p(Z_t = j \mid Z_0 = i), \forall i, j \in E, \forall t \in \mathbb{R}^+$$

## 1.4. Metodología

El desarrollo del primer modelo que se ha expuesto en el Capítulo 2, se ha realizado en referencia a una base de datos que ha sido elaborada a partir de los registros que han efectuado médicos especialistas al realizar el seguimiento de la enfermedad en las distintas etapas y a partir de las consultas de los pacientes que han sido atendidos en el *Hospital Universitario y Politécnico de La Fe* de Valencia; en el período comprendido entre el año 1995 y 2010.

El conjunto de datos corresponde a 960 pacientes a los que se les ha extirpado un tumor con células cancerígenas en la vejiga. Luego de haber realizado la intervención quirúrgica en la que se ha removido el tumor, éste puede reaparecer en un nivel similar de malignidad

(recidiva) o de progreso en un nivel superior invadiendo los músculos de la vejiga, en cuyo caso el órgano debe ser extirpado. El proceso de la enfermedad se ha analizado hasta que finaliza el seguimiento del paciente, ya sea por detectarse invasión del músculo de la vejiga, porque en cierta consulta el paciente no presenta recidiva y ya no se realizan sobre él nuevos registros, porque se le diagnostica recidiva y luego no se realizan sobre él nuevos registros, o por haber finalizado el período de observación.

Surgen del registro médico diferentes variables que describiremos a continuación con la notación del registro original:

- *Id*: el identificador numérico que no recoge información personal de los pacientes. Permite interpretar el proceso que ha seguido cada uno, en relación a las demás variables.
- *Fecha.rtu*: corresponde a la fecha en que se realiza la primera intervención quirúrgica para remover el tumor maligno que caracteriza a la enfermedad. Es cuando situamos al paciente en el estado 0 del modelo.
- *Enum*: es el número de consulta a la que asiste el paciente y en el que se recoge y registra la información para su seguimiento. La primera revisión posterior a la primera intervención quirúrgica se registra con valor 1 para dicha variable y así sucesivamente.
- *Time1*: contiene la información del instante en el que el paciente realizó la consulta anterior. Se fija el valor 0 para el instante en el que cada paciente es intervenido quirúrgicamente por primera vez.
- *Time2*: contiene la información del instante en el que el paciente realiza la consulta actual.
- *Time.GT*: presenta la información del tiempo transcurrido entre la consulta próxima anterior ( $Enum=(n-1)$ ) y la presente ( $Enum=n$ ), es decir,  $Time2 - Time1$ .

- *Status*: es una variable dicotómica, indicador de presencia de recidiva del tumor (1) o ausencia de recidiva (0).
- *Status.prog*: es otra variable dicotómica que describe de alguna forma el grado de malignidad del tumor. Si el nivel de avance de la enfermedad es tal que no es posible mantener al paciente con la vejiga, ésta tendrá que extirparse. En esta última situación la variable *Status.prog* toma valor 1. Si la recidiva es de tales características que requiere intervención quirúrgica para remover el tumor, pero no es necesario extirpar la vejiga, toma valor 0.

Se ha tenido en cuenta que cada paciente acude a los distintos controles posteriores a su primer intervención quirúrgica y en ellos puede presentar o no recidiva, siendo o no necesaria una nueva intervención quirúrgica. A su vez, dependiendo del nivel de malignidad del tumor, se extirpa el tumor o la vejiga. El registro para el seguimiento de la trayectoria de la enfermedad continúa mientras no se haya extirpado la vejiga (progresión) y continúen los controles médicos, dentro del período del estudio.

Cabe destacar que no se ha registrado información luego de cierta consulta para aquellos pacientes que están en alguna de las siguientes situaciones:

- Presenta progresión de la enfermedad (cuando se debe extirpar la vejiga).
- Se le diagnostica la ausencia de recidiva.
- Se le diagnostica recidiva, es intervenido y posteriormente el paciente no asiste a nuevas consultas de control en ese centro de atención médica.

El hecho de que no se cuente con nuevos registros en la base de datos para aquellos pacientes que en una consulta no presentan recidiva, no determina en forma tácita que el paciente ha dejado de controlarse. Podría tratarse de pacientes que trasladan sus consultas

a otro centro de atención médica. Sin embargo, se ha considerado que en el instante en que ocurre la última consulta, el paciente sale del sistema a partir del que se modela la enfermedad. Esto ha generado en una primera instancia la necesidad de definir dos estados absorbentes con características diferenciadas:

- El estado que hemos denotado como  $P$  y que indica la progresión de la enfermedad.
- El estado que hemos denotado como  $NP$  y que indica que el paciente sin haber registrado progresión de la enfermedad, ha dejado de realizar controles médicos (con o sin diagnóstico de recidiva en la última consulta) o bien, que el período de estudio ha concluído.

Si luego de un diagnóstico de recidiva e ingreso a algún estado definido como transitorio el paciente ha salido del sistema sin nuevas consultas, el tiempo de permanencia en dicho estado transitorio es nulo. Es por esta razón que se ha desarrollado el modelo de ajuste para 847 casos, ya que del total de 960 pacientes hay 113 casos con tiempos nulos en alguno de los estados transitorios debido a esta razón: 62 casos con tiempo de permanencia nulo en el estado 1, 30 casos en el estado 2 y 21 casos en el estado 3.

El sistema multi-estado para el que se ha desarrollado un modelo de ajuste, es el que se presenta en la Figura 1.1.

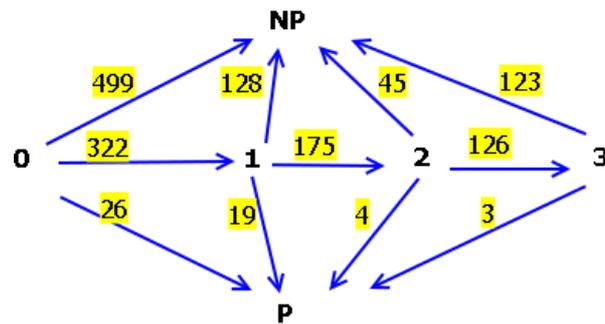


Figura 1.1. Diagrama del modelo que representa en forma completa las etapas de la enfermedad observadas en  $n=847$  pacientes, con dos estados absorbentes.

La interpretación de cada estado es la que se presenta a continuación, teniendo en cuenta lo anteriormente expuesto:

- Estado 0: el paciente ha sido intervenido quirúrgicamente y se le ha extirpado por primera vez tumor/es en la vejiga.
- Estados  $k = 1$  y  $k = 2$ : el paciente es sometido a la intervención quirúrgica número  $k$ , posterior a la primera intervención, por presentar recidiva sin progresión músculo invasiva.
- Estado 3: El paciente se encuentra en este estado en un instante  $t$  desde que se le aplica la intervención número 3 y hasta que sale del sistema. Permanece en este estado con 3 o más recidivas antes de salir del sistema. Muy pocos pacientes que han llegado a 3 presentan progresión a partir de allí: uno de ellos en la consulta número 4, otro en la 5 y un tercer paciente en la consulta 9, de un total de 147 personas que han alcanzado dicho estado. Crear nuevos estados a partir de 3 significa admitir mayores márgenes de error en el ajuste logrado a partir de modelos teóricos, lo que distorsionaría el estudio. Es posible asumir en términos generales que los pacientes que han alcanzado el estado 3 se mantienen en estado de supervivencia, mientras no presenten progresión, dado que una amplia mayoría sobrevive sin progresión.

- Estado  $P$ : este estado absorbente es alcanzado cuando la enfermedad ha progresado diagnosticándose un tumor músculo invasivo, siendo necesario extirpar la vejiga.
- Estado  $NP$ : es el estado absorbente que indica que el paciente ha salido del sistema sin progresión de la enfermedad. En todos los casos esto ocurre cuando en cierta consulta no se ha diagnosticado recidiva, o porque el período de estudio ha concluído.

Ocurre un *suceso* (estado transitorio) cada vez que el paciente es sometido a una intervención quirúrgica para extirpar el tumor, preservando la vejiga. En el modelo esto se visualiza a través de la transición de un estado transitorio  $k$  al siguiente  $k + 1$ , con  $k = 0, 1, 2$ .

El procedimiento que se ha seguido incluye una discusión de posibles ajustes para los tiempos de permanencia en cada estado transitorio del sistema y en el propio sistema, con las distribuciones más conocidas por sus aplicaciones en supervivencia, a través del paquete `riskDistributions` de *R*. Posteriormente se han obtenido distribuciones tipo fase con ajuste satisfactorio para los tiempos de permanencia en cada estado transitorio, a partir del algoritmo (EM) de Asmussen, Nerman y Olsson [5]. Cada ajuste ha sido validado a través de contrastes de hipótesis de Kolmogorov-Smirnov. Finalmente se han estimado las medidas de interés en relación a la supervivencia de la enfermedad.

Los resultados del Capítulo 3, corresponden a un estudio más profundo de la misma enfermedad a partir de una base de datos de menor tamaño que la anterior pero que cuenta con información relevante de cada paciente observado, su diagnóstico inicial y el proceso particular de su enfermedad. El conjunto de datos corresponde a 122 pacientes a los que se les ha extirpado un tumor con células cancerígenas en la vejiga. Luego de haber sido realizada la primera intervención quirúrgica el tumor puede reaparecer con una o más recidivas locales, con o sin metástasis, hasta un último registro que corresponde al fallecimiento del

paciente. En muchos casos la finalización del seguimiento se debe a la culminación del tiempo de estudio. Cuando esto ocurre, el tiempo de seguimiento corresponde a un tiempo de vida censurado aleatoriamente por derecha.

Surgen del registro médico diferentes variables que describiremos a continuación con la notación del registro original. Las variables relacionadas con el aspecto y descripción del tumor, se han valorado en el inicio del seguimiento y otras como la *Situación*, *Metástasis*, *Número de resecciones trans-uretrales* (RTU) se han valorado al final del seguimiento:

- Id: el identificador numérico que no recoge información personal de los pacientes. Permite interpretar el proceso que ha seguido cada uno, en relación a las demás variables.
- Edad del paciente (en años).
- Sexo: 0 si es varón y 1 si es mujer.
- Estadío: comprende el tamaño, extensión o diseminación y el grado. Se ha asignado 0 al estadío  $T_2$ , 1 al  $T_{2-3}$ , 2 al  $T_3$  y 3 al  $T_4$ .
- Grado: 0 si el grado es *I* ó *II* y 1 al *III* ó *IV*.
- Multiplicidad: 0 si el tumor es único o 1 si son varios.
- Tamaño: 0 hasta 5 centímetros y 1 para mayor tamaño.
- Aspecto: 0 si es papilar y 1 si es sólido.
- Pielografía: 0 si es negativa y 1 si es positiva.
- Número de RTU: 0 si es una resección y 1 si se han realizado dos o más.
- Tumor antes del tratamiento con radioterapia (RT): 0 si no presentó tumor y 1 de haber presentado tumor en forma previa al tratamiento.
- Recidiva local (RL): 0 si no la presenta y 1 en caso contrario.

- Metástasis: 0 si se trata de un paciente que no ha presentado durante el seguimiento metástasis y 1 en caso contrario.
- Situación: 0 si el paciente finaliza el seguimiento vivo y 1 en caso de fallecimiento.
- Supervivencia: tiempo (en meses) transcurrido durante el seguimiento hasta el último registro.
- Tiempo de recidiva local: tiempo (en meses) hasta la primera recidiva local.

El sistema multiestado que representa las distintas etapas de la enfermedad se presenta en la Figura 1.2 y muestra todas las transiciones que han realizado los  $n = 122$  pacientes que se han observado, con los correspondientes recuentos. El estado 0 representa el ingreso al sistema de control y seguimiento de la enfermedad. En el instante  $t = 0$  se ha fijado el comienzo de dicho seguimiento individual y en todos los casos corresponde a la primera intervención quirúrgica del paciente. Los estados 1 y 2 corresponden a la primera y segunda recidiva local (RL) sin metástasis. Un paciente ingresa al estado 3 si presenta metástasis pero sin RL e ingresa a 4 si presenta ambas cosas. Los estados  $S$  y  $M$  representan la salida del sistema con vida y muerte respectivamente.

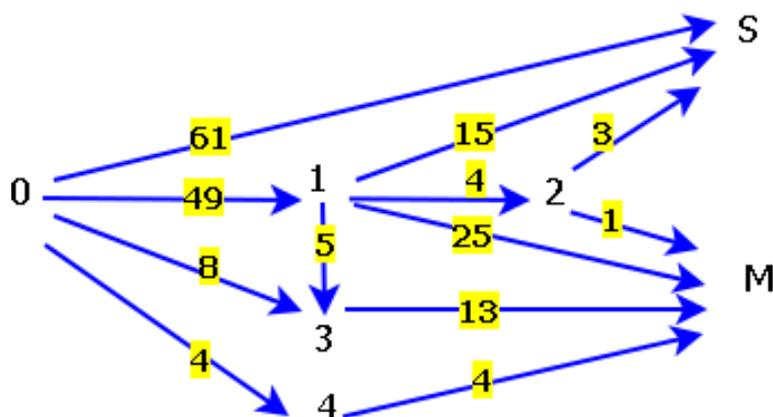


Figura 1.2. Sistema multi-estado que representa las distintas etapas de la enfermedad de los  $n = 122$  pacientes.

Para el abordaje del problema ha sido necesario realizar algunas simplificaciones del modelo anterior, teniendo además en cuenta la gran cantidad de observaciones censuradas con las que se cuenta en esta base de datos.

Una forma de utilizar la información de las covariables de interés para el modelo, es a partir del modelo de riesgos proporcionales de Cox que se ha presentado previamente. Para ello se ha considerado la función de riesgo  $\lambda(\cdot|z)$  en un tiempo  $t$  y para un individuo con vector de covariables  $Z = z$ , con  $t \in \mathbb{R}^+$  y  $Z \in \mathbb{R}^d$ . La información disponible ha permitido incorporar una única covariable como se argumentará en la presentación de resultados, así es que  $d = 1$ .

$$\begin{aligned}\lambda(t|z) &= \lambda_0(t)\phi(\beta, z) \\ &= \lambda_0(t)\exp(\beta'z)\end{aligned}$$

En este caso,  $\lambda(t|z)$  representa la matriz de riesgos o razones de transición  $Q(z)$ , la función  $\lambda_0(t)$  ha sido considerada como constante e igual a la matriz de riesgos o razones de transición en el conjunto de pacientes de menor riesgo (en relación al vector de covariables  $Z$ ). Dicho estrato ha sido determinado a partir del análisis de las observaciones en relación a las variables de interés y es donde se han calculado los bloques matriciales que conforman la matriz  $Q_0$  a partir del algoritmo EM para distribuciones tipo fase, logrando ajustes adecuados. Una vez se ha determinado  $Q(Z)$ , que depende de los valores de las covariables de  $Z$ , ha sido posible calcular las medidas de interés en supervivencia para los distintos grupos y establecer comparaciones.

La estimación de parámetros ha requerido la definición de la función *verosimilitud parcial* porque se ha dispuesto de una base de datos de supervivencia con observaciones

censuradas. Como además existen repeticiones de algunos tiempos de supervivencia y en cada caso  $d_i$ , que corresponde al número de fallecimientos para cierto instante  $t_i$  con  $i \in \{1, \dots, k\}$ , es pequeño comparado con el número de individuos en riesgo justo antes de dicho tiempo, se ha realizado un buen ajuste de la verosimilitud parcial a partir de la función que ha sido propuesta por Breslow y expuesta en los preliminares de este trabajo. Siendo  $S_i$  la suma de los vectores de covariables para los  $d_i$  individuos que fallecen en  $t_i$ , quienes generan las repeticiones de los tiempos de supervivencia:

$$\prod_{i=1}^k \frac{\exp(\beta' S_i)}{[\sum_{l \in R(t_i)} \exp(\beta' Z_l)]^{d_i}}$$

Para la estimación de parámetros, se ha optimizado el logaritmo de la función anterior, maximizándola, a partir del algoritmo de Nelder-Mead adaptado a la función matricial implicada, utilizando el lenguaje Wolfram en Mathematica y ejecutándolo sobre las funciones particulares para las distintas transiciones entre estados. Puede verse las referencias a los procedimientos implicados en el apéndice IV.

En el Capítulo 4 hemos considerado procesos semi-Markovianos para modelar el proceso del cáncer de vejiga a través de diferentes estados que representan las distintas etapas de la enfermedad a través del tiempo. Se ha construido un modelo multi-estado a partir de la misma base de datos inicial, constituida por las observaciones realizadas durante un período fijo de quince años en el que se recoge información sobre 960 pacientes que padecen esta enfermedad. Este nuevo enfoque ha tenido por objetivo optimizar el uso de la información de la que se dispone, de forma tal que el modelo construido contenga precisamente todo lo que sabemos de cada paciente.

Como ya se ha explicado, luego de remover el tumor, el carcinoma puede reaparecer con distintos niveles de malignidad. Si luego de algunas recurrencias la enfermedad progresa con la presencia de un tumor que invade los músculos del órgano, es la vejiga que debe ser removida y finaliza el registro de información relativo al seguimiento del paciente. Por

otro lado el período de observación puede concluir debido a la finalización del estudio, en cuyo caso el tiempo hasta la progresión es censurado por derecha. Es así que los estados transitorios del modelo quedan determinados por las sucesivas recidivas sin progresión de la enfermedad, posteriores a la primera cirugía que da inicio a cada proceso individual. En este modelo se ha incorporado un único estado absorbente que representa la ocurrencia de la progresión de la enfermedad, dado que las salidas del sistema por las causas antes descritas se ha considerado que generan tiempos de supervivencia censurados por derecha.

En la Figura 1.3, se presenta el diagrama de transición entre los estados de  $E = E_0 \cup D$ , donde  $E_0 = \{0, 1, 2, 3\}$  es el conjunto de los estados transitorios y  $D = \{4\}$  es el conjunto unitario con un estado absorbente que representa la progresión de la enfermedad ( $P$ ). La salida del sistema sin progresión nuevamente se ha representado con  $NP$ . Se ha considerado que el tiempo de permanencia en el último estado transitorio de  $E_0$  antes de que finalice el período de estudio y por lo tanto el seguimiento del paciente, es un dato censurado porque ya que no se cuenta con ninguna información posterior en relación a su supervivencia. De lo anterior se deduce la razón por la que  $NP$  no ha sido considerado un estado absorbente.

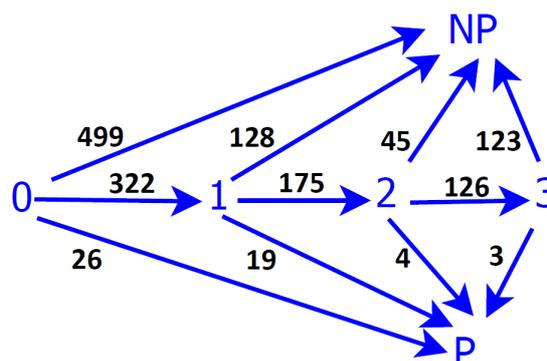


Figura 1.3. Diagrama de transiciones entre los estados del sistema.

$E$  es el conjunto finito de estados del modelo y se ha considerado procesos estocásticos semi-Markovianos para representar las diferentes trayectorias de los pacientes en el sistema multi-estados,  $Z = (Z_t, t \in R^+)$ , de manera que:

- Los tiempos de saltos de  $Z$  son  $S_0 = 0 \leq S_1 \leq \dots \leq S_n \leq S_{n+1} \dots$
- Las sucesivas visitas a los estados son  $J_0, J_1, \dots, J_n, J_{n+1}, \dots$
- Los tiempos de permanencia en los estados transitorios entre saltos son  $X_0 = 0$  y  $X_n = S_n - S_{n-1}$ , con  $n \geq 1$ .

Las bases de datos en Biomedicina usualmente contienen gran cantidad de observaciones censuradas. A pesar de ello se ha obtenido buen ajuste de las principales medidas de supervivencia gracias a las propiedades de consistencia de los estimadores construidos, que serán expuestas en el Capítulo 4. Los procedimientos algorítmicos empleados se han registrado en el Apéndice V.



## **Principales resultados**



---

## **Un modelo multidimensional con tiempos no exponenciales en los estados**

---

El modelo multidimensional que se construye en este capítulo, ha sido diseñado a partir de la caracterización clínica de las etapas de la enfermedad a partir de la base de datos que se ha mencionado en la descripción de la metodología de trabajo (Capítulo 1). Se ha podido constatar que las distribuciones usuales no presentan ajustes satisfactorios para los tiempos de permanencia en los estados de cada sistema, limitación que ha sido superada utilizando distribuciones tipo-fase (PH) a partir de las propiedades que las caracterizan.

### **2.1. Análisis de la información recogida en la base de datos**

El sistema para el que realizaremos el modelo de ajuste, es el que se presenta en la Figura [2.1](#).

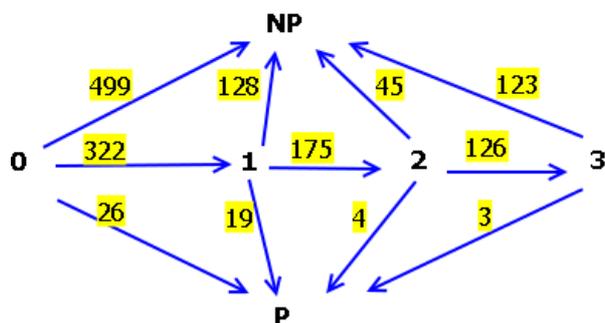


Figura 2.1. Diagrama del modelo que representa en forma completa las etapas de la enfermedad observadas en  $n=847$  pacientes, con dos estados absorbentes.

En el Tabla 2.1 se realizan los recuentos correspondientes a las transiciones desde cada estado, al resto de los estados del modelo:

	0	1	2	3	P	NP
0	0	322	0	0	26	499
1	0	0	175	0	19	128
2	0	0	0	126	4	45
3	0	0	0	0	3	123

Tabla 2.1. Transiciones entre los estados del modelo

Para obtener los tiempos de permanencia en 0 se realiza una selección de casos en la base original, dando valores a las variables *Estado ant*= 0 y *Estado act*= 1, P ó NP. Análogamente para obtener los tiempos de permanencia en 1 y 2 se puede hacer una selección de casos dando valores a la variable *Estado ant*= 1 y *Estado act*= 2 respectivamente.

Una vez guardada la selección de casos en los nuevos ficheros, el tiempo de permanencia en cada estado se identifica con el valor de la variable *Time.GT* de la base de datos, que según se explicó anteriormente, presenta la información del tiempo transcurrido entre la

consulta próxima anterior ( $Enum = (n - 1)$ ) y la presente ( $Enum = n$ ), es decir:  $Time_2 - Time_1$ .

Para los pacientes que luego de una consulta no concurren a un nuevo control en el centro hospitalario luego de alguna de las ocurrencias del suceso *intervención quirúrgica* 1 o 2, no se tiene información alguna sobre el tiempo de ocurrencia del siguiente suceso. Hay 62 casos de transición desde 1 hasta *NP* explicados por la inasistencia de los pacientes a una nueva consulta luego de la intervención quirúrgica correspondiente. Por igual causa, hay 30 casos de transiciones desde 2 a *NP* y 21 de 3 a *NP*. En todos esos casos asignamos valor 0 a los tiempos de permanencia en el último estado transitorio antes de salir del sistema. Esos casos son los que se han desestimado del sistema, según ya se ha reseñado en la metodología. Esto es diferente para el estado transitorio 0, en el que se observan 499 transiciones a *NP* por ser que en la primera consulta el paciente no presenta recidiva. En estos casos los tiempos de permanencia en 0 corresponden al tiempo transcurrido hasta la primera consulta, luego de la cual el paciente sale del sistema. Estos tiempos de permanencia en el sistema, son no nulos.

El resumen de los tiempos de permanencia en cada estado transitorio, para el total de 847 pacientes, se presenta en la Tabla 2.2.

	n	Rango	Mínimo	Máximo	Mediana	Media	Desviación Típica
0	847	4248	85	4333	656,00	971,02	880,659
1	322	3895	87	3982	399,50	679,55	671,478
2	175	4009	83	4092	372,00	628,73	660,283
3	126	3192	82	3274	891,50	1101,56	800,25

Tabla 2.2. Descripción estadística de los tiempos en cada estado transitorio.

Es claro que para el estado transitorio 3, a pesar de que con él se representan 3 o más recidivas, el rango es notablemente menor respecto a los demás casos, pero sin embargo el promedio de los tiempos de permanencia en el mismo es superior a los demás, con un desvío estándar importante y que supera al desvío estándar de los estados 1 y 2.

En términos generales, los pacientes permanecen en algún estado transitorio antes de salir del sistema sin recidiva, durante tiempos promedialmente elevados, lo que generará como se verá posteriormente incremento en los costes de seguimiento.

Puede además observarse que la mediana en cada caso es notoriamente inferior a la media, lo que hace referencias a un sesgo izquierdo en la distribución de los valores respecto a esta última. Se presentan a continuación los histogramas correspondientes a cada caso. Véase las Figuras 2.2, 2.3, 2.4 y 2.5.

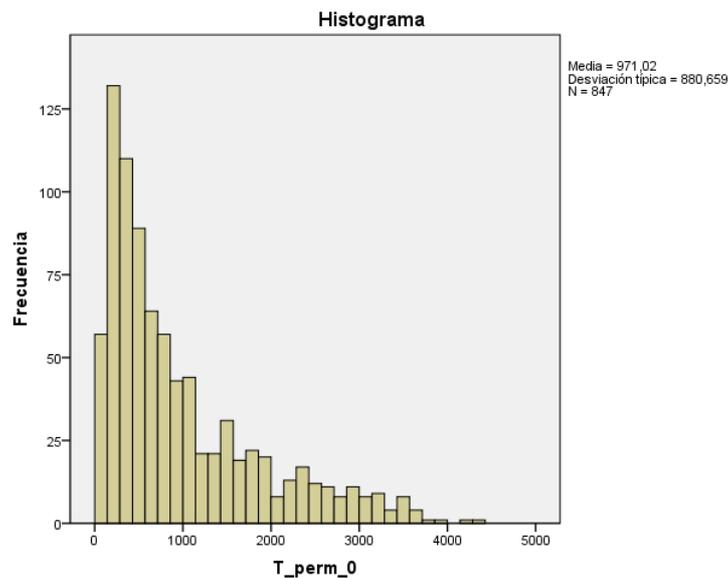


Figura 2.2. Histograma correspondiente a la variable *Tiempo de permanencia en 0*, a partir de 847 casos.

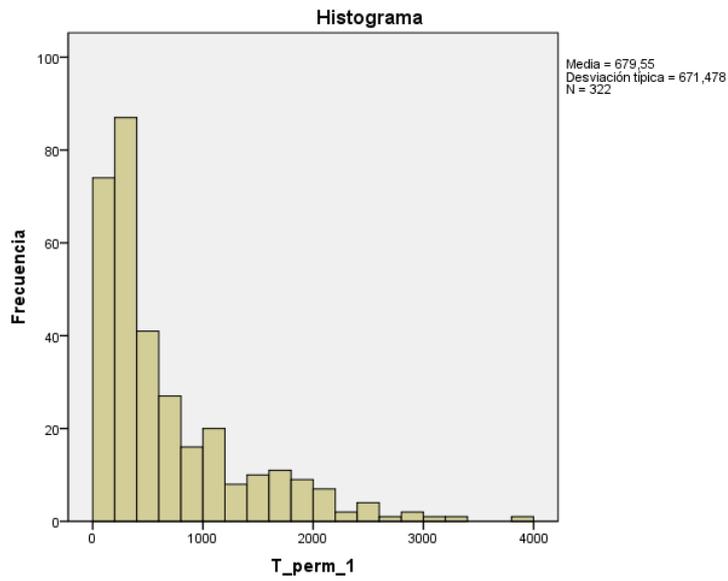


Figura 2.3. Histograma correspondiente a la variable *Tiempo de permanencia en 1*, a partir de 322 casos.

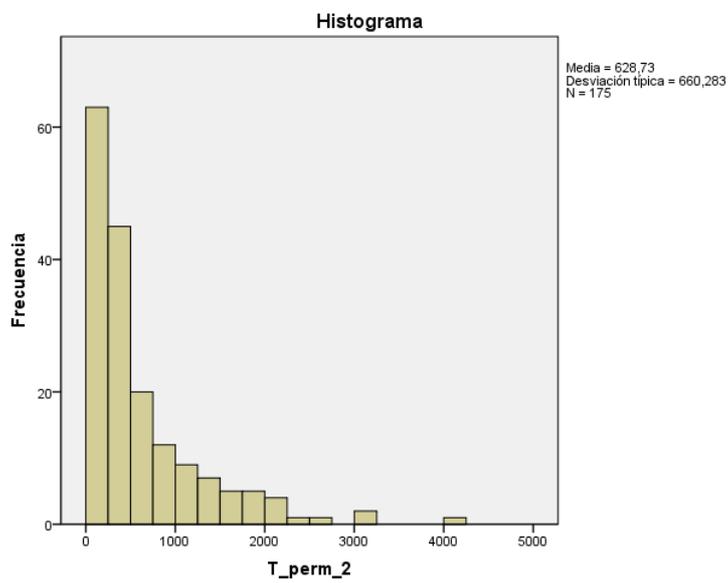


Figura 2.4. Histograma correspondiente a la variable *Tiempo de permanencia en 2*, a partir de 175 casos.

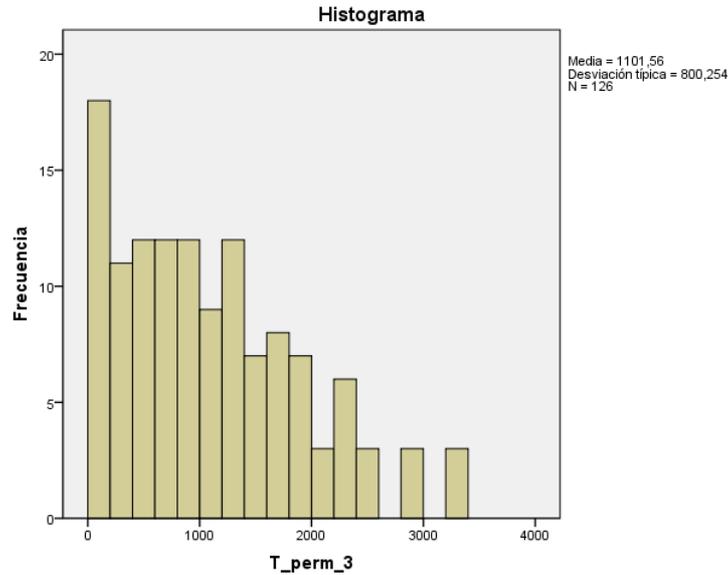


Figura 2.5. Histograma correspondiente a la variable *Tiempo de permanencia en 3*, a partir de 126 casos.

## 2.2. Ajuste para la distribución de los tiempos de permanencia en cada estado

### Análisis de la correlación entre los tiempos de permanencia en los estados transitorios

Se trata de analizar primero la correlación entre dos variables *Tiempo de permanencia en 0* y *Tiempo de permanencia en 1*, considerando pares de valores para cada paciente. Se trabajará luego con las dos variables *Tiempo de permanencia en 1* y *Tiempo de permanencia en 2*, como con *Tiempo de permanencia en 2* y *Tiempo de permanencia en 3*, sin embargo la cantidad de casos varía en cada análisis.

Para el primer estudio consideramos los casos de aquellos pacientes que transitan de 0 a 1, se analizan entonces 322 casos que corresponden a los pacientes que transitan de 0 a 1 y

posteriormente de 1 a 2, a *P* o a *NP*. Ver las Figuras 2.6 y 2.7.

Para el segundo análisis consideraremos como casos el de aquellos pacientes que transitan de 1 a 2, se analizan entonces 175 casos que corresponden a los pacientes que transitan de 1 a 2 y posteriormente de 2 a 3, a *P* o a *NP*. Ver las Figuras 2.8 y 2.9.

Para el tercer análisis, consideraremos como casos el de aquellos pacientes que transitan de 2 a 3, se analizan entonces 126 casos que corresponden a los pacientes que transitan de 2 a 3 y posteriormente de 3 a *P* o a *NP*. Ver las Figuras 2.10 y 2.11.

		T_perm_0	T_perm_1
T_perm_0	Correlación de Pearson	1	,163**
	Sig. (bilateral)		,003
	N	322	322
T_perm_1	Correlación de Pearson	,163**	1
	Sig. (bilateral)	,003	
	N	322	322

\*\* . La correlación es significativa al nivel 0,01 (bilateral).

Figura 2.6. Análisis paramétrico de correlación entre las variables *Tiempo de permanencia en 0* y *Tiempo de permanencia en 1*, a partir de 322 casos.

**Correlaciones**

			T_perm_0	T_perm_1
Tau_b de Kendall	T_perm_0	Ceficiente de correlación	1,000	,175**
		Sg. (bilateral)	.	,000
		N	322	322
	T_perm_1	Ceficiente de correlación	,175**	1,000
		Sg. (bilateral)	,000	.
		N	322	322
Rho de Spearman	T_perm_0	Ceficiente de correlación	1,000	,258**
		Sg. (bilateral)	.	,000
		N	322	322
	T_perm_1	Ceficiente de correlación	,258**	1,000
		Sg. (bilateral)	,000	.
		N	322	322

Figura 2.7. Análisis no paramétrico de correlación entre las variables *Tiempo de permanencia en 0* y *Tiempo de permanencia en 1*, a partir de 322 casos.

**Correlaciones**

		T_perm_1	T_perm_2
T_perm_1	Correlación de Pearson	1	,133
	Sig. (bilateral)		,080
	N	175	175
T_perm_2	Correlación de Pearson	,133	1
	Sig. (bilateral)	,080	
	N	175	175

Figura 2.8. Análisis paramétrico de correlación entre las variables *Tiempo de permanencia en 1* y *Tiempo de permanencia en 2*, a partir de 175 casos.

**Correlaciones**

			T_perm_1	T_perm_2
Tau_b de Kendall	T_perm_1	Coefficiente de correlación	1,000	,142**
		Sig. (bilateral)	.	,005
		N	175	175
	T_perm_2	Coefficiente de correlación	,142**	1,000
		Sig. (bilateral)	,005	.
		N	175	175
Rho de Spearman	T_perm_1	Coefficiente de correlación	1,000	,212**
		Sig. (bilateral)	.	,005
		N	175	175
	T_perm_2	Coefficiente de correlación	,212**	1,000
		Sig. (bilateral)	,005	.
		N	175	175

\*\* . La correlación es significativa al nivel 0,01 (bilateral).

Figura 2.9. Análisis no paramétrico de correlación entre las variables *Tiempo de permanencia en 1* y *Tiempo de permanencia en 2*, a partir de 175 casos.

**Correlaciones**

		T_perm_2	T_perm_3
T_perm_2	Correlación de Pearson	1	-,183*
	Sig. (bilateral)		,041
	N	126	126
T_perm_3	Correlación de Pearson	-,183*	1
	Sig. (bilateral)	,041	
	N	126	126

\*. La correlación es significativa al nivel 0,05 (bilateral).

Figura 2.10. Análisis paramétrico de correlación entre las variables *Tiempo de permanencia en 2* y *Tiempo de permanencia en 3*, a partir de 126 casos.

Correlaciones			T_perm_2	T_perm_3
Tau_b de Kendall	T_perm_2	Coefficiente de correlación	1,000	-,068
		Sig. (bilateral)	.	,263
		N	126	126
	T_perm_3	Coefficiente de correlación	-,068	1,000
		Sig. (bilateral)	,263	.
		N	126	126
Rho de Spearman	T_perm_2	Coefficiente de correlación	1,000	-,108
		Sig. (bilateral)	.	,231
		N	126	126
	T_perm_3	Coefficiente de correlación	-,108	1,000
		Sig. (bilateral)	,231	.
		N	126	126

Figura 2.11. Análisis no paramétrico de correlación entre las variables *Tiempo de permanencia en 2* y *Tiempo de permanencia en 3*, a partir de 126 casos.

Para aquellos casos en que el contraste es significativo, para un umbral de 0.05, los distintos coeficientes de correlación dejan evidencia de correlación débil entre cada par de variables consideradas. Podremos presuponer entonces que los tiempos de permanencia en los distintos estados, corresponden a variables que pueden considerarse independientes.

### **Análisis de los tiempos de permanencia en cada estado transitorio**

Consideraremos todos los tiempos de permanencia en cada estado transitorio 0, 1, 2 y 3, sin discriminar por estado al que luego transita cada paciente.

Se ha realizado a través del paquete *riskDistributions* de *R*, los contrastes correspondientes para analizar la bondad de los distintos ajustes con las distribuciones más conocidas en supervivencia, para los tiempos de permanencia en cada estado transitorio. Véase las Figuras 2.12, 2.13, 2.14 y 2.15.

```

> datos<-read.table("TIEMPOS_EN_0_POSITIVOS.txt",header=T)
> fit_0<-fit.cont(datos$tiempos)

Begin fitting distributions -----
* fitting normal distribution ... OK
* fitting Cauchy distribution ... OK
* fitting logistic distribution ... OK
* fitting beta distribution ... failed
* fitting exponential distribution ... failed
* fitting chi-square distribution ... OK
* fitting uniform distribution ... OK
* fitting gamma distribution ... failed
* fitting lognormal distribution ... OK
* fitting Weibull distribution ... OK
* fitting F-distribution ... OK
* fitting Student's t-distribution ... OK
* fitting Gompertz distribution ... failed
* fitting triangular distribution ... failed
End fitting distributions -----

      logL      AIC      BIC Chisq(value) Chisq(p) AD(value)
Normal    -6944.57  13893.14  13902.62      770.37      0      44.75
Cauchy    -6941.59  13887.17  13896.66      583.83      0      60.07
Logistic  -6921.51  13847.02  13856.5      661.10      0      34.98
Chi-square -145375.23 290752.46 290757.2      Inf      0      Inf
Uniform      NULL      NULL      NULL      Inf      0      Inf
Lognormal  -6636.16  13276.32  13285.81      74.53      0      3.81
Weibull   -6659.95  13323.9  13333.38      105.15      0      7.16
F         -8077.14  16158.27  16167.76     4393.75      0     345.90
Student   -8677.76  17357.52  17362.26     9786.69      0     734.07

      H(AD) KS(value) H(KS)
Normal    rejected      0.16 rejected
Cauchy    rejected      0.21 rejected
Logistic  rejected      0.17 rejected
Chi-square  NULL      0.45 rejected
Uniform    NULL      0.21 rejected
Lognormal  rejected      0.05 rejected
Weibull    rejected      0.06 rejected
F          NULL      0.57 rejected
Student    NULL      0.78 rejected

```

Figura 2.12. Pruebas de bondad de ajuste para la distribución de la variable *Tiempo de permanencia en el estado 0*.

```

> fit_1<-fit.cont(datos$tiempos)

Begin fitting distributions -----
* fitting normal distribution ... OK
* fitting Cauchy distribution ... OK
* fitting logistic distribution ... OK
* fitting beta distribution ... failed
* fitting exponential distribution ... failed
* fitting chi-square distribution ... OK
* fitting uniform distribution ... OK
* fitting gamma distribution ... failed
* fitting lognormal distribution ... OK
* fitting Weibull distribution ... OK
* fitting F-distribution ... OK
* fitting Student's t-distribution ... OK
* fitting Gompertz distribution ... failed
* fitting triangular distribution ... failed
End fitting distributions -----

      logL      AIC      BIC Chisq(value) Chisq(p) AD(value) H(AD) KS(value) H(KS)
Normal    -2552.45  5109.9  5116.45      435.96      0      22.71 rejected      0.19 rejected
Cauchy    -2510.07  5024.15  5031.7      262.50      0      32.72 rejected      0.23 rejected
Logistic  -2532.62  5069.23  5076.78      374.30      0      18.04 rejected      0.20 rejected
Chi-square -39359.83 78721.66 78725.43      Inf      0      Inf NULL      0.47 rejected
Uniform    NULL      NULL      NULL      Inf      0      Inf NULL      0.24 rejected
Lognormal  -2392.54  4789.08  4796.63      54.75      0      3.36 rejected      0.09 rejected
Weibull   -2418.85  4841.7  4849.25      94.76      0      6.56 rejected      0.10 rejected
F         -2937.38  5878.76  5886.31     1662.87      0     132.46 NULL      0.59 rejected
Student   -3166.9  6335.8  6339.57     3725.78      0     280.05 NULL      0.79 rejected

Chosen continuous distribution is: Normal (norm)
Fitted parameters are:
      mean      sd
679.5528 670.4350

```

Figura 2.13. Pruebas de bondad de ajuste para la distribución de la variable *Tiempo de permanencia en el estado 1*.

```

> datos<-read.table("TIEMPOS_EN_2_POSITIVOS.txt",header=T)
> fit_2<-fit.cont(datos$tiempos)

Begin fitting distributions -----
* fitting normal distribution ... OK
* fitting Cauchy distribution ... OK
* fitting logistic distribution ... OK
* fitting beta distribution ... failed
* fitting exponential distribution ... failed
* fitting chi-square distribution ... OK
* fitting uniform distribution ... OK
* fitting gamma distribution ... OK
* fitting lognormal distribution ... OK
* fitting Weibull distribution ... OK
* fitting F-distribution ... OK
* fitting Student's t-distribution ... OK
* fitting Gompertz distribution ... failed
* fitting triangular distribution ... failed
End fitting distributions -----

      logL      AIC      BIC Chisq(value) Chisq(p) AD(value)  H(AD) KS(value)  H(KS)
Normal -1384.03 2772.06 2778.39      282.36      0      14.43 rejected 0.20 rejected
Cauchy -1342.23 2688.45 2694.78      115.66      0      17.74 rejected 0.23 rejected
Logistic -1365.95 2735.91 2742.24      223.30      0      10.85 rejected 0.21 rejected
Chi-square -19819.71 39639.43 39642.59      Inf      0      Inf NULL 0.47 rejected
Uniform NULL NULL NULL      334.07      0      Inf NULL 0.26 rejected
Gamma -1298.87 2601.74 2608.06      67.48      0      4.85 rejected 0.12 rejected
Lognormal -1282.42 2568.83 2575.16      35.43      0      2.24 rejected 0.09 not rejected
Weibull -1301.4 2606.81 2613.14      69.85      0      4.51 rejected 0.12 rejected
F -1580.87 3165.73 3172.06      915.38      0      72.65 NULL 0.59 rejected
Student -1705.76 3413.51 3416.68      2053.41      0      152.63 NULL 0.79 rejected

Chosen continuous distribution is: Log-normal (lnorm)
Fitted parameters are:
  meanlog      sdlog
6.0180591 0.8968114

```

Figura 2.14. Pruebas de bondad de ajuste para la distribución de la variable *Tiempo de permanencia en el estado 2*.

```

> datos<-read.table("TIEMPOS_EN_3_POSITIVOS.txt",header=T)
> fit_3<-fit.cont(datos$tiempos)

Begin fitting distributions -----
* fitting normal distribution ... OK
* fitting Cauchy distribution ... OK
* fitting logistic distribution ... OK
* fitting beta distribution ... failed
* fitting exponential distribution ... failed
* fitting chi-square distribution ... OK
* fitting uniform distribution ... OK
* fitting gamma distribution ... failed
* fitting lognormal distribution ... OK
* fitting Weibull distribution ... OK
* fitting F-distribution ... OK
* fitting Student's t-distribution ... OK
* fitting Gompertz distribution ... failed
* fitting triangular distribution ... failed
End fitting distributions -----

      logL      AIC      BIC Chisq(value) Chisq(p) AD(value)  H(AD) KS(value)  H(KS)
Normal -1020.59 2045.17 2050.84      24.77      0.01      2.09 rejected 0.11 not rejected
Cauchy -1044.64 2093.28 2098.95      56.63      0.00      3.97 rejected 0.18 rejected
Logistic -1022.25 2048.51 2054.18      26.23      0.00      1.82 rejected 0.11 not rejected
Chi-square -21489.1 42978.21 42981.04      Inf      0.00      Inf NULL 0.52 rejected
Uniform NULL NULL NULL      Inf      0.00      Inf NULL 0.13 rejected
Lognormal -1011.19 2026.38 2032.05      33.18      0.00      2.76 rejected 0.10 not rejected
Weibull -1001.11 2006.23 2011.9      10.79      0.37      0.71 not rejected 0.06 not rejected
F -1228.78 2461.56 2467.24      850.37      0.00      51.94 NULL 0.56 not rejected
Student -1317.91 2837.82 2840.66      1846.24      0.00      109.56 NULL 0.77 rejected

Chosen continuous distribution is: Normal (norm)
Fitted parameters are:
  mean      sd
1101.5556 797.0717

```

Figura 2.15. Pruebas de bondad de ajuste para la distribución de la variable *Tiempo de permanencia en el estado 3*.

Nótese que en los dos primeros estados transitorios no se han logrado ajustes adecuados y sin embargo en 2 y 3 se ha admitido una distribución *lognormal*. No obstante lo anterior y buscando construir un modelo versátil desde el punto de vista computacional, veremos que será posible aproximar la distribución en todos los estados transitorios y en el sistema global a partir de las propiedades de las distribuciones tipo fase, apoyándonos en la aplicación

de los métodos analíticos matriciales de Neuts [30] para la construcción del generador del sistema. Sobre esto trabajaremos a continuación, buscando primero un ajuste para la distribución del *Tiempo de permanencia en el sistema*, construyendo posteriormente un modelo que permita calcular las medidas de interés para la supervivencia de la enfermedad, teniendo en cuenta los tiempos de permanencia en cada estado transitorio y la estimación de sus distribuciones.

### **Ajuste de la distribución para los tiempos de permanencia en cada estado con una distribución *tipo fase***

En virtud de lo observado hasta el momento, nos proponemos construir una distribución *tipo fase* con un ajuste satisfactorio para los tiempos de permanencia en los estados transitorios del modelo que representa la evolución de la enfermedad y a partir de los datos. Así obtendremos un modelo más apropiado que los que ya se han aplicado anteriormente. Una vez presentado el modelo, se validará a través de contrastes de hipótesis de Kolmogorov-Smirnov.

Para cada estado transitorio 0, 1, 2 y 3, según se ha explicado en los objetivos de este trabajo, se construirá la distribución PH que se ajuste satisfactoriamente a la de cada variable aleatoria *Tiempo de permanencia en el estado*.

Se empleará el algoritmo EM de ajuste con distribuciones de este tipo, creado por el grupo de investigación de Chalmers University of Technology and Göteborg University en 1998 (véase [32]). Dicho algoritmo permite ingresar desde un fichero con formato de texto, las observaciones de la variable cuya distribución se quiere aproximar, codificadas según sean o no censuradas, elegir el orden para la distribución tipo fase ( $p$ ) y el número de iteraciones ( $N$ ) para el ajuste.

El procedimiento de estimar los parámetros  $(\alpha, A)$  de la distribución tipo fase, se realiza con estimaciones sucesivas que optimizan el valor de la función de verosimilitud para la muestra específica  $y = (y_1, \dots, y_n)$ :

$$L^k(\alpha^k, A^k; y) \leq L^{k+1}(\alpha^{k+1}, A^{k+1}; y), k \in \{1, \dots, N-1\}$$

Las estimaciones  $(\hat{\alpha}, \hat{A})$  convergen al par  $(\alpha, A)$ , cuya aproximación obtendremos luego de  $N$  iteraciones.

A partir de tales estimaciones  $(\hat{\alpha}, \hat{A})$  y de la lectura de las observaciones desde un archivo, se ha ejecutado en *Matlab* el algoritmo contenido en el archivo *PHplot.m* creado para ser implementado en dicho software por los mismos autores. De esa forma ha sido sencillo obtener los gráficos de las funciones de interés.

Por otro lado, se ha modificado el código contenido (ver el Apéndice III) para obtener los valores en la distribución de ajuste, para las observaciones de tiempo de la muestra. Así se ha podido calcular  $D$  del estadístico de Kolmogorov-Smirnov. Luego de comparar dicho valor con el límite de cada región crítica,  $D_\alpha$ , se ha podido decidir si rechazar o admitir la hipótesis nula de ajuste satisfactorio con la correspondiente distribución.

Se ha logrado un buen ajuste con distribuciones *tipo fase*, en algunos casos con  $p = 2$  y en otros casos, con  $p = 4$ . En cualquier caso, se han realizado  $N = 1000$  iteraciones.

#### ■ Ajuste para los tiempos de permanencia en el estado 0

A partir del algoritmo EM se ha realizado un ajuste con distribución PH para las 847 observaciones correspondientes a los tiempos de permanencia en el estado inicial 0,

antes de la primera intervención quirúrgica a partir de él con o sin progresión de la enfermedad ó antes de la salida del sistema. De los 847 pacientes registrados en la base de datos y a partir de la primera intervención quirúrgica para extirpar el tumor no invasivo de los músculos de la vejiga (en el instante 0):

- 499 salen del sistema sin progresión (estado *NP*), dado que en la primera consulta son diagnosticados sin recidiva, luego de lo cual no se han realizado nuevos registros de controles médicos.
- 322 realizan una primera consulta de control en la que se diagnostica recidiva sin progresión y
- 26 presentan progresión, según el diagnóstico de la primera consulta.

Se obtuvo las siguientes estimaciones del vector  $\alpha_0$ , la matriz  $A_0$  y la matriz columna de absorción  $A_0^0$ :

$$\widehat{\alpha}_0 = \begin{bmatrix} 1.0000 \\ 0.0000 \end{bmatrix}, \widehat{A}_0 = \begin{bmatrix} -0.0114 & 0.0114 \\ 0.0003 & -0.0015 \end{bmatrix} \text{ y } \widehat{A}_0^0 = \begin{bmatrix} 0.0000 \\ 0.0012 \end{bmatrix}$$

En las Figuras 2.16 y 2.17, se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 0* y la función de supervivencia, ambas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

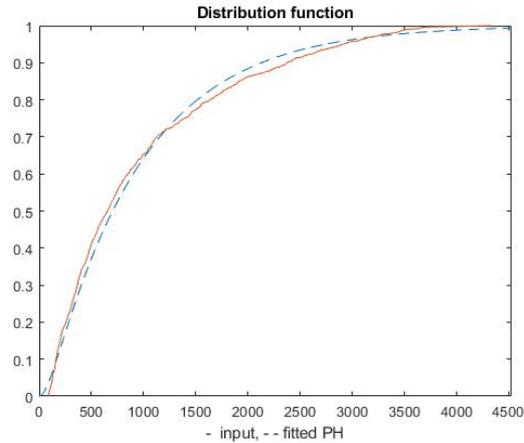


Figura 2.16. Gráfico de la función de distribución tipo fase estimada, para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 847$  casos

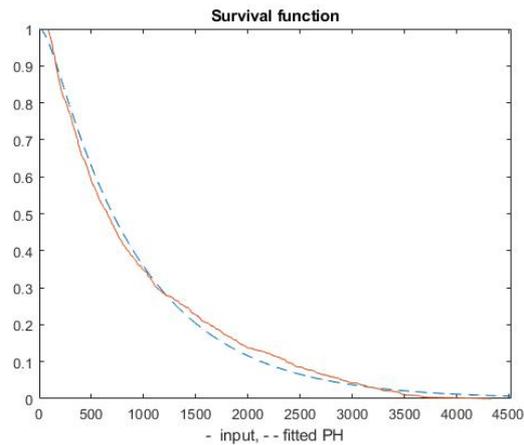


Figura 2.17. Gráfico de la función de supervivencia estimada, para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 847$  casos.

$D_{\alpha=0.05} = 0.0465$  y  $D_{\alpha=0.01} = 0.0557$ . El valor del estadístico de Kolmogorov-Smirnov es  $D = 0.0417$  que resulta ser menor que  $D_{\alpha=0.05} = 0.0465$ . Admitimos que el ajuste con esta distribución  $PH(\hat{\alpha}, \hat{T})$  es satisfactorio, para un nivel de significación de 0.05 y claro que también para 0.01.

En la Figura 2.18 se presenta el gráfico de la función *densidad*, que se ha obtenido a partir del ajuste con una distribución tipo fase, para la variable *Tiempo de permanencia en 0*.

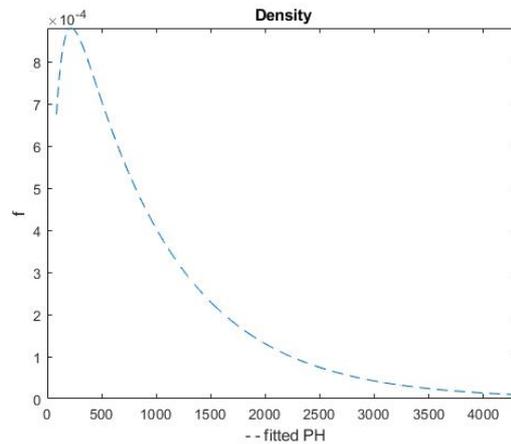


Figura 2.18. Gráfico de la función densidad estimada, para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 847$  casos.

En la Figura 2.19 se presenta el gráfico de la función *razón de fallo*, que se ha obtenido a partir del ajuste con una distribución tipo fase, para la variable *Tiempo de permanencia en 0*.

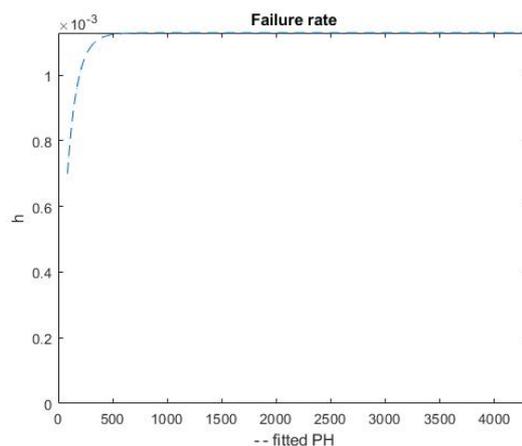


Figura 2.19. Gráfico de la función razón de fallo estimada, para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 847$  casos.

■ **Ajuste para los tiempos de permanencia en el estado 1**

A partir del algoritmo EM se ha realizado un ajuste con distribución PH y a partir de las 322 observaciones correspondientes a los tiempos de permanencia en el estado 1, antes de la segunda intervención quirúrgica posterior a la inicial (estado 2), pasaje a progresión de la enfermedad (estado  $P$ ) o salida del sistema sin progresión (estado  $NP$ ). De los 322 pacientes:

- 175 transitan a 2 ya que en la siguiente consulta presentan recidiva sin progresión, siendo intervenidos nuevamente quirúrgicamente.
- 128 realizan una consulta posterior en la que son diagnosticados sin recidiva, luego de lo cual no presentan nuevos registros de consultas. Se trata de pacientes que salen del sistema luego de un tiempo de estancia positivo en el estado 1 siendo absorbidos por  $NP$ .
- 19 realizan una segunda consulta en la que se diagnostica la progresión de la enfermedad, transitando a  $P$ .

Para lograr un buen ajuste, fue necesario darle al parámetro  $p$  valor  $p = 4$ , obteniéndose las siguientes estimaciones del vector  $\alpha_1$ , de la matriz  $A_1$  y de la matriz columna de absorción  $A_1^0$ :

$$\widehat{\alpha}_1 = \begin{bmatrix} 0.0000 \\ 0.0000 \\ 1.0000 \\ 0.0000 \end{bmatrix}, \widehat{A}_1 = \begin{bmatrix} -0.0028 & 0.0003 & 0.0024 & 0.0001 \\ 0.0035 & -0.0109 & 0.0000 & 0.0074 \\ 0.0001 & 0.0108 & -0.0109 & 0.0000 \\ 0.0020 & 0.0000 & 0.0000 & -0.0120 \end{bmatrix} \text{ y}$$

$$\widehat{A}_1^0 = \begin{bmatrix} 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0100 \end{bmatrix}$$

En las Figuras 2.20 y 2.21, se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 1* y la función de supervivencia, ambas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

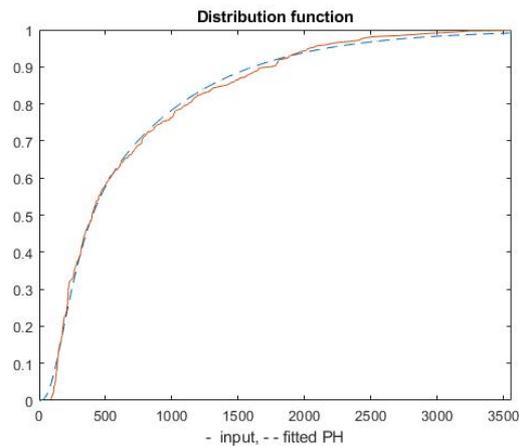


Figura 2.20. Gráfico de la función de distribución tipo fase estimada, para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 322$  casos

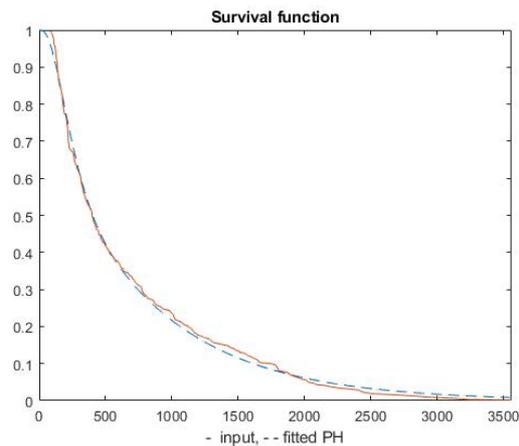


Figura 2.21. Gráfico de la función de supervivencia estimada, para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 322$  casos.

El valor del estadístico de Kolmogorov-Smirnov es aproximadamente  $D = 0.0510$ , que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.0752$  y  $D_{\alpha=0.01} = 0.0901$ . Al no pertenecer el valor de  $D$  a la región crítica para ambos niveles de significación, puede considerarse que el ajuste es satisfactorio.

De un total de 52 pacientes que salen del sistema con progresión ( $P$ ) desde alguno de los estados transitorios, la mayor proporción corresponde al estado 1: 19 casos de un total de 322 pacientes que ingresaron a dicho estado con primera recidiva. Esto representa el 5.90% de los casos. En los demás estados la proporción es mucho más baja. Así por ejemplo en 0 los 26 pacientes que salen a  $P$  representan un 3.10%. En 2 y 3, salen del sistema a progresión apenas 4 y 3 pacientes respectivamente, correspondiendo a un 2.29% y un 2.38% en cada caso.

En la Figura 2.22 se presenta el gráfico de la función *densidad* que se ha obtenido a partir del ajuste con una distribución tipo fase, para la variable *Tiempo de permanencia en 1*.

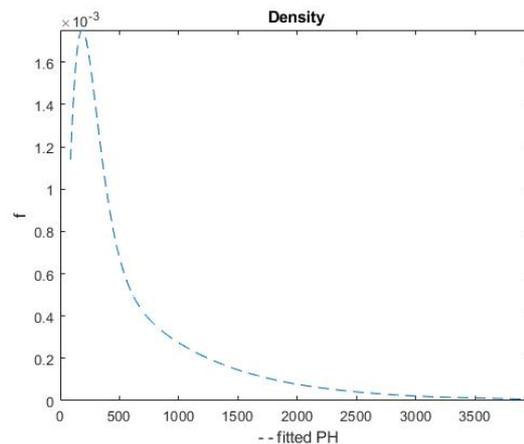


Figura 2.22. Gráfico de la función densidad estimada, para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 322$  casos.

En la Figura 2.23 se presenta el gráfico de la función *razón de fallo*, que se ha obtenido a partir del ajuste con una distribución tipo fase, para la variable *Tiempo de permanencia en 1*.

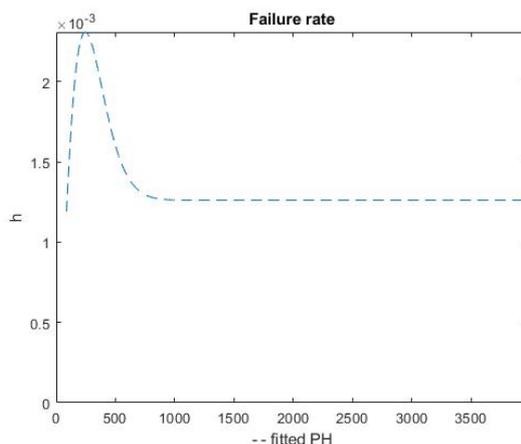


Figura 2.23. Gráfico de la función razón de fallo estimada, para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 322$  casos.

#### ■ Ajuste para los tiempos de permanencia en el estado 2

A partir del algoritmo EM se ha realizado un ajuste con distribución PH para las 175 observaciones correspondientes a los tiempos de permanencia en el estado 2, antes de la tercera intervención quirúrgica posterior a la inicial (estado 3), pasaje a progresión (estado  $P$ ) o salida del sistema sin progresión (estado  $NP$ ). De los 175 pacientes:

- 126 transitan a 3 ya que en la siguiente consulta presentan recidiva sin progresión, siendo intervenidos nuevamente quirúrgicamente.
- 45 realizan una consulta posterior en la que son diagnosticados sin recidiva, luego de lo cual no presentan nuevos registros de consultas. Se trata de pacientes que salen del sistema luego de un tiempo de estancia positivo en el estado 2, siendo absorbidos por  $NP$ .

- 4 realizan una tercera consulta en la que se diagnostica progresión de la enfermedad, transitando a  $P$ .

Se obtuvo las siguientes estimaciones del vector  $\alpha_2$ , de la matriz  $A_2$  y de la matriz columna  $A_2^0$ :

$$\widehat{\alpha}_2 = \begin{bmatrix} 0.0000 \\ 0.0000 \\ 1.0000 \\ 0.0000 \end{bmatrix}, \widehat{A}_2 = \begin{bmatrix} -0.0024 & 0.0002 & 0.0020 & 0.0001 \\ 0.0033 & -0.0109 & 0.0000 & 0.0076 \\ 0.0001 & 0.0109 & -0.0110 & 0.0000 \\ 0.0013 & 0.0000 & 0.0000 & -0.0117 \end{bmatrix} \text{ y}$$

$$\widehat{A}_2^0 = \begin{bmatrix} 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0103 \end{bmatrix}$$

En las Figuras 2.24 y 2.25 se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 2* y la función de supervivencia, ambas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

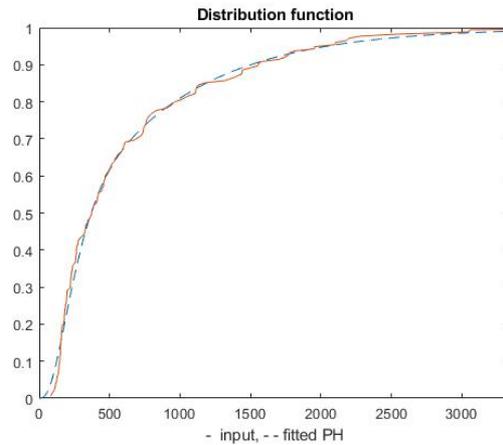


Figura 2.24. Gráfico de la función de distribución tipo fase estimada, para el *Tiempo de permanencia en 2*, a partir de la muestra de  $n = 175$  casos

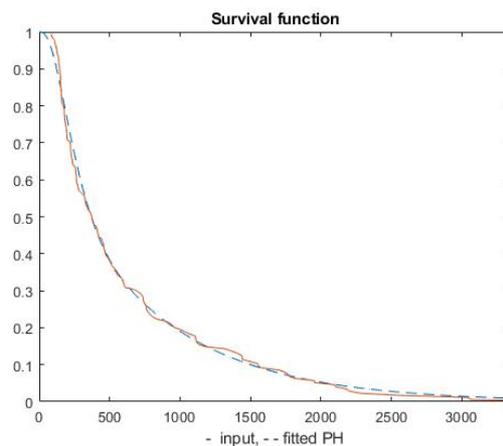


Figura 2.25. Gráfico de la función de supervivencia estimada, para el *Tiempo de permanencia en 2*, a partir de la muestra de  $n = 175$  casos.

El valor del estadístico de Kolmogorov-Smirnov es aproximadamente  $D = 0.0608$ , que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.1017$  y  $D_{\alpha=0.01} = 0.1219$ , con lo cual se evidencia la bondad del ajuste para ambos niveles de significación.

En la Figura 2.26 se presenta el gráfico de la función *densidad* que se ha obtenido a partir del ajuste con una distribución tipo fase, para la variable *Tiempo de permanencia en 1*.

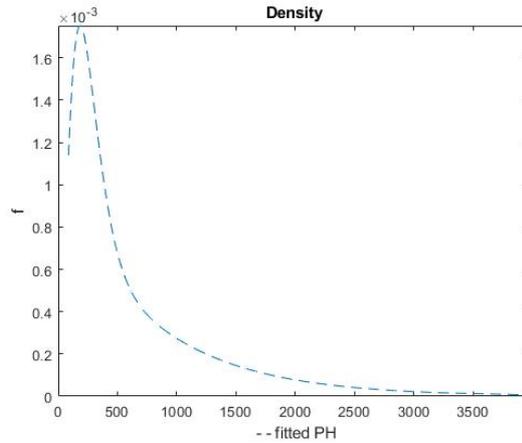


Figura 2.26. Gráfico de la función densidad estimada, para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 175$  casos.

En la Figura 2.27 se presenta el gráfico de la función *razón de fallo*, que se ha obtenido a partir del ajuste con la distribución tipo fase antes expuesta, para la variable *Tiempo de permanencia en 2*.

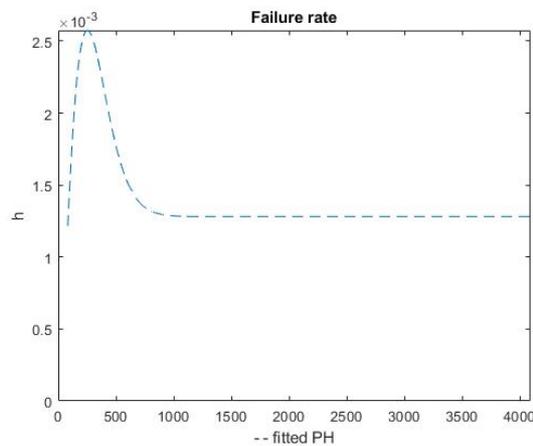


Figura 2.27. Gráfico de la función razón de fallo estimada, para el *Tiempo de permanencia en 2*, a partir de la muestra de  $n = 175$  casos.

### ■ Ajuste para los tiempos de permanencia en el estado 3

A partir del algoritmo EM se ha realizado un ajuste con distribución PH para las 126 observaciones correspondientes a los tiempos de permanencia en el estado 3, antes de

que el paciente salga del sistema con progresión (estado  $P$ ) o sin progresión (estado  $NP$ ) luego de 3 o más recidivas. De los 126 pacientes:

- 123 realizan una cuarta consulta o alguna posterior en la que son diagnosticados sin recidiva, luego de lo cual no presentan nuevos registros de consultas en la base de datos. Se trata de pacientes que salen del sistema luego de un tiempo de estancia positivo en el estado 3 siendo el proceso absorbido por  $NP$ .
- 3 realizan una posterior consulta en la que se diagnostica progresión de la enfermedad, transitando a  $P$ .

Fue posible obtener un muy buen ajuste con una distribución *tipo fase* y  $p = 2$ , con las siguientes estimaciones del vector  $\alpha_3$ , de la matriz  $A_3$  y de la matriz columna  $A_3^0$ :

$$\widehat{\alpha}_3 = \begin{bmatrix} 0.7776 \\ 0.2224 \end{bmatrix}, \widehat{A}_3 = \begin{bmatrix} -0.0016 & 0.0016 \\ 0.0000 & -0.0016 \end{bmatrix} \text{ y } \widehat{A}_3^0 = \begin{bmatrix} 0.0000 \\ 0.0016 \end{bmatrix}$$

En las Figuras 2.28 y 2.29 se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 3* y la función de supervivencia, ambas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

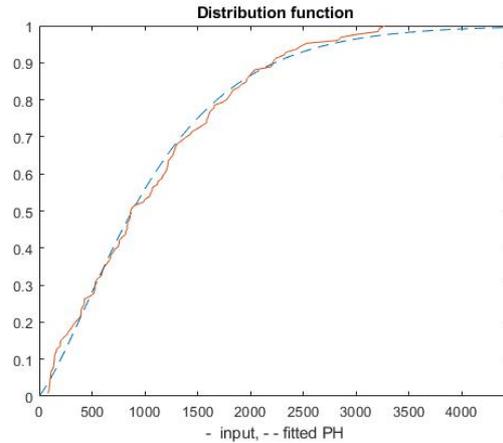


Figura 2.28. Gráfico de la función de distribución tipo fase estimada, para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 126$  casos

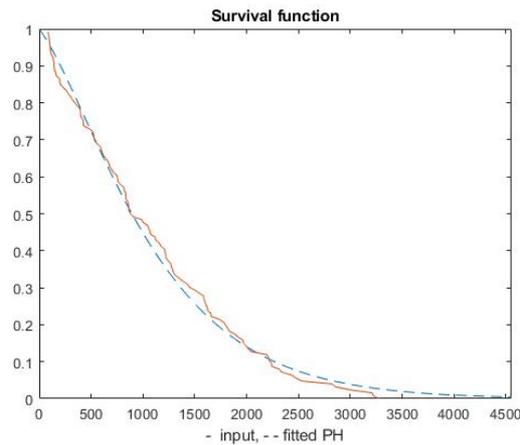


Figura 2.29. Gráfico de la función de supervivencia estimada, para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 126$  casos.

El valor del estadístico de Kolmogorov-Smirnov es aproximadamente  $D = 0.0505$ , menor que los dos valores  $D_{\alpha=0.05} = 0.1196$  y  $D_{\alpha=0.01} = 0.1434$ , con lo cual el ajuste logrado es más que satisfactorio para ambos niveles de significación.

Nótese que esta función de supervivencia, es la que decrece menos rápidamente, en comparación con las de los otros estados transitorios. Este resultado se vincula a que los pacientes que ingresan a 3 sobreviven casi en su totalidad.

En la Figura 2.30 se presenta el gráfico de la función *densidad* que se ha obtenido a partir del ajuste con una distribución tipo fase, para la variable *Tiempo de permanencia en 3*.

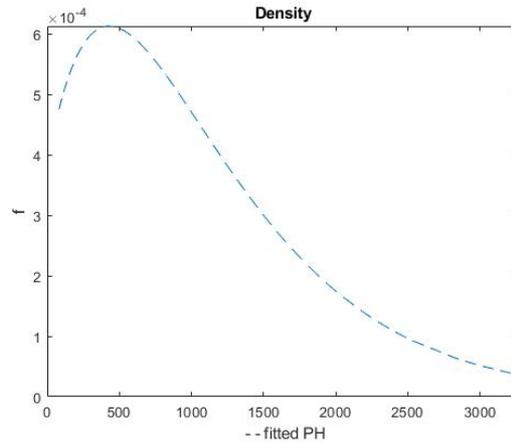


Figura 2.30. Gráfico de la función densidad estimada, para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 126$  casos.

En la Figura 2.31 se presenta el gráfico de la función *razón de fallo*, que se ha obtenido a partir del ajuste con una distribución tipo fase, para la variable *Tiempo de permanencia en 3*.

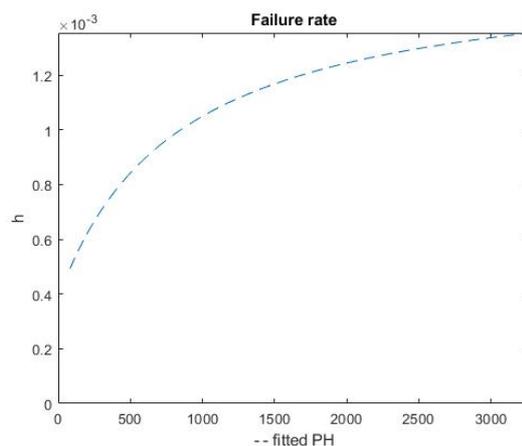


Figura 2.31. Gráfico de la función razón de fallo estimada, para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 126$  casos.

En la Tabla 2.3 se presentan algunos valores de la función  $S(t)$  hasta los instantes especificados en la primera columna, a partir de las funciones de supervivencia estimadas en cada estado transitorio. En la última columna, se registran los valores de la función  $S(t)$  para los tiempos de permanencia en el sistema. Téngase en cuenta que la evaluación de la función de supervivencia en un instante  $t$ , representa la probabilidad de que el paciente permanezca sin progresión hasta  $t$ .

En la Tabla 2.4 se presentan algunos valores de la función razón de fallo  $h(t)$ , hasta los instantes especificados en la primera columna, a partir de las funciones estimadas en cada estado transitorio.

t (en días)	$S_0(t)$	$S_1(t)$	$S_2(t)$	$S_3(t)$	$S_{sys}(t)$
200	0.8724	0.7780	0.7607	0.9030	0.9959
400	0.7032	0.5030	0.4663	0.7842	0.9896
600	0.5618	0.3637	0.3256	0.6628	0.9831
800	0.4484	0.2789	0.2463	0.5492	0.9772
1000	0.3578	0.2164	0.1900	0.4481	0.9719
1200	0.2855	0.1681	0.1470	0.3612	0.9673
1400	0.2278	0.1306	0.1138	0.2883	0.9632
1600	0.1818	0.1015	0.0880	0.2283	0.9597
1800	0.1451	0.0788	0.0681	0.1795	0.9567
2000	0.1158	0.0613	0.0527	0.1404	0.9541
2200	0.0924	0.0476	0.0408	0.1092	0.9519
2400	0.0737	0.0370	0.0316	0.0846	0.9499
2600	0.0588	0.0287	0.0244	0.0652	0.9483
2800	0.0469	0.0223	0.0189	0.0502	0.9469
3000	0.0375	0.0173	0.0146	0.0384	0.9456
3200	0.0299	0.0135	0.0113	0.0294	0.9446
3400	0.0238	0.0105	0.0088	0.0224	0.9437
3600	0.0190	0.0081	0.0068	0.0170	0.9430
3800	0.0152	0.0063	0.0053	0.0129	0.9423
4000	0.0121	0.0049	0.0041	0.0098	0.9418
4200	0.0097	0.0038	0.0031	0.0074	0.9413
4400	0.0077	0.0030	0.0024	0.0056	0.9409
4600	0.0062	0.0023	0.0019	0.0042	0.9406
4800	0.0049	0.0018	0.0015	0.0032	0.9403
5000	0.0039	0.0014	0.0011	0.0024	0.9401

Tabla 2.3. Comparativa de valores de  $S(t)$ , para los distintos estados transitorios y en el sistema.

t (en días)	$h_0(t)$	$h_1(t)$	$h_2(t)$	$h_3(t)$
200	0.0010	0.0022	0.0025	0.0006
400	0.0011	0.0019	0.0022	0.0008
600	0.0011	0.0014	0.0015	0.0009
800	0.0011	0.0013	0.0013	0.0010
1000	0.0011	0.0013	0.0013	0.0010
1200	0.0011	0.0013	0.0013	0.0011
1400	0.0011	0.0013	0.0013	0.0011
1600	0.0011	0.0013	0.0013	0.0012
1800	0.0011	0.0013	0.0013	0.0012
2000	0.0011	0.0013	0.0013	0.0012
2200	0.0011	0.0013	0.0013	0.0013
2400	0.0011	0.0013	0.0013	0.0013
2600	0.0011	0.0013	0.0013	0.0013
2800	0.0011	0.0013	0.0013	0.0013
3000	0.0011	0.0013	0.0013	0.0013
3200	0.0011	0.0013	0.0013	0.0013
3400	0.0011	0.0013	0.0013	0.0014
3600	0.0011	0.0013	0.0013	0.0014
3800	0.0011	0.0013	0.0013	0.0014
4000	0.0011	0.0013	0.0013	0.0014
4200	0.0011	0.0013	0.0013	0.0014
4400	0.0011	0.0013	0.0013	0.0014
4600	0.0011	0.0013	0.0013	0.0014
4800	0.0011	0.0013	0.0013	0.0014
5000	0.0011	0.0013	0.0013	0.0014

Tabla 2.4. Comparativa de valores de  $h(t)$ , para los distintos estados transitorios.

Se resume en la Tabla 2.5 las características de los ajustes logrados con distribuciones *tipo fase*, para los tiempos de permanencia en cada estado transitorio:  $n$  el tamaño muestral, los valores  $D$  de los estadísticos de Kolmogorov Smirnov y los límites de las regiones críticas  $D_\alpha$ , evidenciando en cada caso la bondad de los mismos.

Estados	p	n	D	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
0	2	847	0,0417	0,0419	0,0465	0,0557
1	4	322	0,0510	0,0677	0,0752	0,0901
2	4	175	0,0608	0,0916	0,1017	0,1219
3	2	126	0,0543	0,1078	0,1196	0,1434
Tiempos en el sistema		52	0.0223	0,1666	0,1849	0,2216

Tabla 2.5. Análisis de la bondad de los ajustes logrados en cada distribución, a partir de la distribución PH especificada.

### 2.3. Generador del Sistema y las principales medidas de la supervivencia a la enfermedad

A partir de las estimaciones realizadas y cuyos resultados se han expuesto, disponemos de  $(\alpha_i, A_i)$  para  $(i \in \{0, 1, 2, 3\})$ . Estos pares determinan las estimaciones de las distribuciones *tipo fase* para los tiempos de permanencia en cada estado transitorio 0, 1, 2 y 3. A partir de ellos, será posible determinar el generador  $Q$  que caracterizará, junto al vector de probabilidades iniciales  $\alpha$ , la distribución *tipo fase* del sistema y permitirá calcular todas las medidas de interés.

Nótese que el vector  $\widehat{\alpha}_0$  de las probabilidades iniciales estimadas para las fases del estado 0, es el que brindará las probabilidades iniciales para el sistema global, corrigiendo su dimensión a través de la asignación de valores 0 a las demás entradas.

Cada matriz  $A_i$  da a conocer las razones de transición entre las fases, virtuales, dentro del subsistema con el que se ha caracterizado al estado transitorio  $i$ . A su vez, la salida desde un subsistema  $i$  y entrada a otro  $i + 1$ , se dará según las razones de absorción de

la matriz columna  $A_i^0$ , afectadas por la probabilidad de transición entre esos estados y el vector de probabilidades de entrada a las fases virtuales del estado  $i + 1$ . Así por ejemplo, del subsistema (virtual) vinculado al estado 0 puede transitarse al del estado 1, a  $P$  o a  $NP$ , con probabilidades de transición estimadas a partir de las proporciones muestrales de transiciones desde 0 a 1,  $P$  y  $NP$  respectivamente. Es natural entonces que las razones de transición desde 0 al estado transitorio 1, queden definidas no solamente a partir de  $A_0^0$  que es el vector en el que figuran las razones de absorción por parte de la fase absorbente en 0, dado que quedarán afectadas por la probabilidad de transición de 0 a 1 y también por las probabilidades de entrada a las fases, virtuales, del sistema asociado a 1. La probabilidad de transición entre los estados transitorios 0 y 1, se ha estimado puntualmente a partir de la proporción muestral de transiciones entre estos estados  $\widehat{p}_{01}$  y las probabilidades de entrada a las fases del sistema de 1 se han estimado y se registran en  $\widehat{\alpha}_1$ .

En búsqueda de una distribución tipo fase, con buen ajuste para la variable *Tiempo de permanencia en el sistema*, ha de tenerse en cuenta que el vector de probabilidades iniciales será  $\alpha = (1.0000, 0.0000, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ , según el modelo ajustado para el estado transitorio 0.

La matriz generadora  $Q$ , cuadrada y de orden 14, tendrá dos bloques principales  $A$  y  $B$ . Los bloques de  $A$  que definen las 12 primeras filas y 12 primeras columnas, tienen diferente dimensión dado que los ajustes se han logrado con diferentes valores de orden  $p$  y caracterizan las transiciones entre los cuatro estados transitorios. Por otro lado las dos últimas columnas del bloque  $B$ , contienen las razones de absorción por parte de los dos estados absorbentes  $NP$  y  $P$ . La matriz  $Q$  se completa con bloques de matrices nulas de dimensiones correspondientes a cada caso. Cada una de las matrices que componen los bloques de  $Q$ , surgen de una estimación realizada a partir de la muestra de observaciones

y las notaciones correspondientes se han simplificado. Cabe destacar la generalidad del procedimiento y la conformabilidad de los productos matriciales involucrados en los distintos bloques del generador. Nótese que el valor de  $p$  de cada distribución PH puede cambiar (como es el caso de la del estado 1), con el objetivo de mejorar el ajuste en los distintos estados transitorios del sistema.

$$Q = \begin{bmatrix} A_{12 \times 12} & B_{12 \times 2} \\ 0_{2 \times 12} & 0_{2 \times 2} \end{bmatrix}$$

$$A = \begin{bmatrix} A_0 & A_0^0 p_{01} \alpha_1 & 0 & 0 \\ 0 & A_1 & A_1^0 p_{12} \alpha_2 & 0 \\ 0 & 0 & A_2 & A_2^0 p_{23} \alpha_3 \\ 0 & 0 & 0 & A_3 \end{bmatrix}$$

$$B = \begin{bmatrix} p_{0NP} A_0^0 & p_{0P} A_0^0 \\ p_{1NP} A_1^0 & p_{1P} A_1^0 \\ p_{2NP} A_2^0 & p_{2P} A_2^0 \\ p_{3NP} A_3^0 & p_{3P} A_3^0 \end{bmatrix}$$

Teniendo en cuenta que las probabilidades de transición deben sumar 1, se tiene:

$$p_{i(i+1)} + p_{iNP} + p_{iP} = 1, \forall i \in \{0, 1, 2\}$$

Estas probabilidades se han estimado a partir de las proporciones muestrales:

$$\begin{aligned} p_{01} &= 322/847 \approx 0.3802 & p_{0NP} &= 499/847 \approx 0.5891 & p_{0P} &= 26/847 \approx 0.0307 \\ p_{12} &= 175/322 \approx 0.5435 & p_{1NP} &= 128/322 \approx 0.3975 & p_{1P} &= 19/322 \approx 0.0590 \\ p_{23} &= 126/175 = 0.7200 & p_{2NP} &= 45/175 \approx 0.2571 & p_{2P} &= 4/175 \approx 0.0229 \\ & & p_{3NP} &= 123/126 \approx 0.9762 & p_{3P} &= 3/126 \approx 0.0238 \end{aligned}$$

Las matrices  $A$  y  $B$  que componen los dos bloques principales de  $Q$  son respectivamente:

$$A = \begin{bmatrix} -0.0114 & 0.0114 & | & 0.0000 & 0.0000 & 0.0000 & 0.0000 & | & 0 & 0 & 0 & 0 & | & 0 & 0 \\ 0.0003 & -0.0015 & | & 0.0000 & 0.0000 & 0.0004 & 0.0000 & | & 0 & 0 & 0 & 0 & | & 0 & 0 \\ --- & --- & | & --- & --- & --- & --- & | & --- & --- & --- & --- & | & --- & --- \\ 0 & 0 & | & -0.0028 & 0.0003 & 0.0024 & 0.0001 & | & 0.0000 & 0.0000 & 0.0000 & 0.0000 & | & 0 & 0 \\ 0 & 0 & | & 0.0035 & -0.0109 & 0.0000 & 0.0074 & | & 0.0000 & 0.0000 & 0.0000 & 0.0000 & | & 0 & 0 \\ 0 & 0 & | & 0.0001 & 0.0108 & -0.0109 & 0.0000 & | & 0.0000 & 0.0000 & 0.0000 & 0.0000 & | & 0 & 0 \\ 0 & 0 & | & 0.0002 & 0.0000 & 0.0000 & -0.0012 & | & 0.0000 & 0.0000 & 0.0054 & 0.0000 & | & 0 & 0 \\ --- & --- & | & --- & --- & --- & --- & | & --- & --- & --- & --- & | & --- & --- \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & -0.0024 & 0.0002 & 0.0002 & 0.0001 & | & 0.0000 & 0.0000 \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0.0033 & -0.0109 & 0.0000 & 0.0076 & | & 0.0000 & 0.0000 \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0.0001 & 0.0109 & -0.0011 & 0.0000 & | & 0.0000 & 0.0000 \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0.0013 & 0.0000 & 0.0000 & -0.0117 & | & 0.0058 & 0.0016 \\ --- & --- & | & --- & --- & --- & --- & | & --- & --- & --- & --- & | & --- & --- \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0 & 0 & 0 & 0 & | & -0.0016 & 0.0016 \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0.0000 & -0.0016 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0000 & 0.0000 \\ 0.0007 & 0.0000 \\ --- & --- \\ 0.0000 & 0.0000 \\ 0.0000 & 0.0000 \\ 0.0000 & 0.0000 \\ 0.0040 & 0.0006 \\ --- & --- \\ 0.0000 & 0.0000 \\ 0.0000 & 0.0000 \\ 0.0000 & 0.0000 \\ 0.0026 & 0.0002 \\ --- & --- \\ 0.0000 & 0.0000 \\ 0.0016 & 0.0000 \end{bmatrix}$$

Luego, a partir de cálculo simbólico, es posible obtener los dos bloques principales de la matriz  $P(t)$ ,  $A(t)$  y  $B(t)$ , que se han propuesto en el marco teórico. La matriz  $A(t)$  en su elemento  $a_{ij}(t)$  en cada fila  $i$  contiene las probabilidades de ocupación de la fase y estado indicado por la columna en el instante  $t$ , habiendo ingresado al sistema en la fase del estado que indica la fila. La matriz  $B(t)$  en su elemento  $b_{ik}(t)$ , contiene la probabilidad de ocupación del estado absorbente  $k$  en el instante  $t$ , habiendo ingresado al sistema (cuando  $t = 0$ ) en la fase del estado que la fila indica.

Por lo antes expuesto, la supervivencia en el instante  $t$ , que corresponde a la probabilidad de que no haya ocurrido progresión de la enfermedad hasta el instante  $t$ , corresponde a la probabilidad de que hasta ese instante no se haya ocupado el estado  $P$ . Puede calcularse entonces a partir de la segunda columna de la matriz  $B(t)$  de la siguiente forma:

$$S(t) = 1 - \alpha B(t) \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Luego se tiene:

$$\begin{aligned} F(t) &= \alpha B(t) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \alpha (A^{-1}(\exp(At) - I)B) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

Derivando la expresión matricial anterior en  $t$ , se tiene:

$$\begin{aligned} f(t) = F'(t) &= \alpha \exp(At)B \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \alpha A(t)B \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

Así, la función razón de fallo podrá definirse como sigue:

$$h(t) = \frac{f(t)}{S(t)} = \alpha A(t) B \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{1}{S(t)}$$

Teniendo en cuenta que el modelo de sistema multiestado representa las etapas de la enfermedad, cualquiera de los pacientes tiene dos alternativas excluyentes: salir del sistema con progresión de la enfermedad (suceso al que denominaremos P) ó sin progresión desde alguno de los cuatro estados transitorios (suceso complementario de P que denominaremos NP). Con esta interpretación y teniendo en cuenta que la salida del sistema sin progresión se ha representado con uno de los estados absorbentes del modelo consideraremos, en coherencia con el modelo que hemos desarrollado, que hay supervivencia en el instante  $t$  si el paciente es de los que recorren el sistema hasta salir de él a través de algún estado transitorio y sin progresión ó, si es uno de los sujetos que culmina su trayectoria en progresión de la enfermedad, en ese instante todavía no ha ingresado a dicho estado absorbente.

Luego, si para un paciente que termina en progresión, se define la variable aleatoria  $T$  *Tiempo que tarda en ser absorbido por el estado P*, no ser absorbido por P en el instante  $t$  puede expresarse como  $T > t$ , cuyo suceso complementario es  $T \leq t$ . Es así que la función empírica de supervivencia comparable con la función teórica  $S(t)$  puede expresarse de la siguiente forma:

$$S_{emp}(t) = 1 \cdot p(NP) + p(T > t | \overline{NP}) \cdot p(\overline{NP}) = p(NP) + S_{P,emp}(t) \cdot p(P)$$

En la expresión anterior  $S_{P,emp}(t)$  representa la función de supervivencia empírica dentro del subgrupo de pacientes que salen del sistema con progresión de la enfermedad. Por otro lado,  $p(NP)$  y  $p(P)$  designan las probabilidades de salir del sistema, en algún instante, sin progresión y con progresión respectivamente. Estas probabilidades serán estimadas como es habitual con las proporciones muestrales:

$$\widehat{p(NP)} = \frac{795}{847} \quad \widehat{p(P)} = \frac{52}{847}$$

La representación gráfica de ambas funciones de supervivencia teórica  $S(t)$  y  $S_{emp}(t)$  es la de la Figura 2.32.

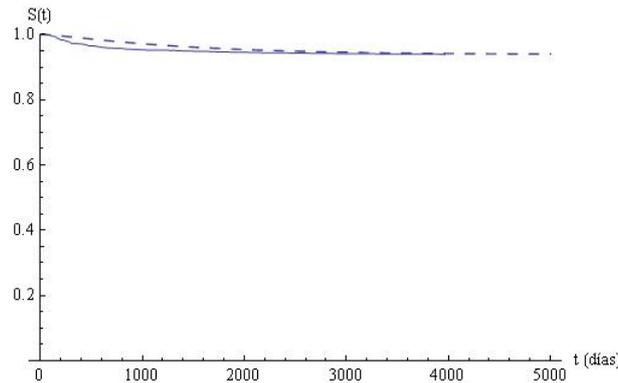


Figura 2.32. Gráfico de la función de supervivencia  $S(t)$  estimada(- -), en contraste con la supervivencia empírica.

El valor del estadístico de Kolmogorov-Smirnov es aproximadamente  $D = 0.0223$ , que resulta ser notablemente menor que los dos valores  $D_{\alpha=0.05} = 0.1849$  y  $D_{\alpha=0.01} = 0.2216$ . Al no pertenecer el valor de  $D$  a las regiones críticas con extremos inferiores 0.1849 y 0.2216, admitimos que el ajuste logrado a partir de la distribución estimada  $F(t) = 1 - S(t)$  con la empírica, es satisfactorio para ambos niveles de significación.

Es posible observar que la función de supervivencia decrece con mayor velocidad hasta los 1000 días, con marcada tendencia a largo plazo hacia un valor apenas inferior a 0.95. Esto tiene consecuencias en la función razón de fallos, que tenderá a 0 rápidamente, dada su relación con la derivada de  $S(t)$ :

$$h(t) = -\frac{dS}{dt} \cdot S^{-1}(t)$$

Nótese que 0.95 es un valor considerablemente alto como límite de la función de supervivencia, marcando gran diferencia en relación a otras enfermedades que ocasionan mayor desgaste de la salud.

Los gráficos de la función *densidad*  $f(t)$  y de la función *razón de fallo*, se presentan en las Figuras 2.33 y 2.34. Véase los valores de dicha función y su decrecimiento con rapidez hacia 0, tal como se esperaba.

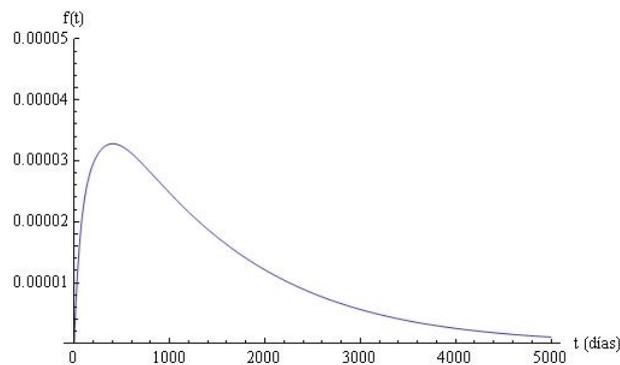


Figura 2.33. Gráfico de la función densidad  $f(t)$ , para el sistema multiestado.

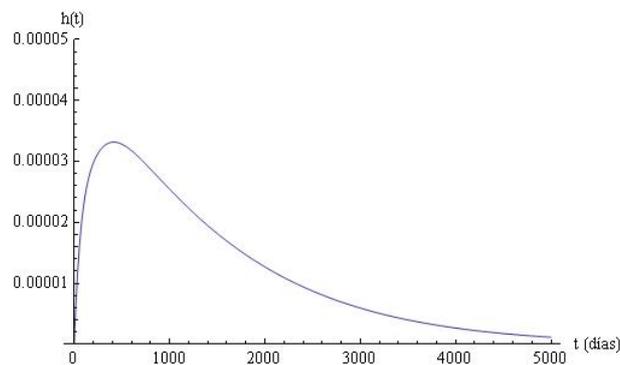


Figura 2.34. Gráfico de la función razón de fallo  $h(t)$ , para el sistema multiestado.

En la Tabla 2.6 se evalúa la función  $S(t)$  en ciertos valores de tiempo. Los resultados corresponden a las respectivas probabilidades de que hasta ese instante, el paciente

permanezca sin progresión del cáncer. También se presenta los respectivos valores de la función razón de fallo  $h(t)$ , que como se explicó tiende a 0 rápidamente.

t (en días)	S(t)	h(t)
200	0.9959	$3.0 \times 10^{-5}$
400	0.9896	$3.3 \times 10^{-5}$
600	0.9831	$3.2 \times 10^{-5}$
800	0.9772	$2.9 \times 10^{-5}$
1000	0.9719	$2.5 \times 10^{-5}$
1200	0.9673	$2.2 \times 10^{-5}$
1400	0.9632	$1.9 \times 10^{-5}$
1600	0.9597	$1.7 \times 10^{-5}$
1800	0.9567	$1.5 \times 10^{-5}$
2000	0.9541	$1.3 \times 10^{-5}$
2200	0.9519	$1.1 \times 10^{-5}$
2400	0.9499	$9.4 \times 10^{-6}$
2600	0.9483	$8.1 \times 10^{-6}$
2800	0.9469	$6.9 \times 10^{-6}$
3000	0.9456	$5.9 \times 10^{-6}$
3200	0.9446	$5.1 \times 10^{-6}$
3400	0.9437	$4.3 \times 10^{-6}$
3600	0.9430	$3.7 \times 10^{-6}$
3800	0.9423	$3.1 \times 10^{-6}$
4000	0.9418	$2.7 \times 10^{-6}$
4200	0.9413	$2.2 \times 10^{-6}$
4400	0.9409	$1.9 \times 10^{-6}$
4600	0.9406	$1.6 \times 10^{-6}$
4800	0.9403	$1.4 \times 10^{-6}$
5000	0.9401	$1.1 \times 10^{-6}$

Tabla 2.6. Estimación de la supervivencia y razón de fallo, en distintos valores de tiempo.

En la Tabla 2.7 se evalúa la función  $S(t)$  en los valores de tiempo muestrales. Los resultados corresponden a las respectivas probabilidades de que hasta ese instante, el paciente permanezca sin progresión. También se presenta los respectivos valores empíricos de supervivencia.

t (en días)	S(t)	Semp(t)
92	0.9989	0.9988
95	0.9987	0.9976
96	0.9987	0.9965
128	0.9979	0.9953
135	0.9978	0.9941
155	0.9972	0.9929
156	0.9972	0.9917
158	0.9971	0.9906
159	0.9971	0.9894
178	0.9966	0.9882
187	0.9963	0.9870
187	0.9963	0.9858
194	0.9961	0.9847
221	0.9953	0.9835
237	0.9948	0.9823
242	0.9947	0.9811
254	0.9943	0.9799
274	0.9937	0.9787
281	0.9934	0.97757
290	0.9932	0.9764
298	0.9929	0.9752
303	0.9927	0.9740
304	0.9927	0.9728
403	0.9895	0.9717
414	0.9891	0.9705
440	0.9883	0.9693
459	0.9876	0.9681
464	0.9875	0.9669
469	0.9873	0.9658
540	0.9850	0.9646
559	0.9844	0.9634
586	0.9836	0.9622
607	0.9829	0.9610
636	0.9820	0.9599
669	0.9810	0.9587
806	0.9770	0.9575
809	0.9769	0.9563
869	0.9753	0.9551
962	0.9729	0.9540
1036	0.9710	0.9528
1313	0.9649	0.9516
1338	0.9644	0.9504
1462	0.9621	0.9492
1687	0.9584	0.9481
1746	0.9575	0.9469
2010	0.9540	0.9457
2159	0.9523	0.9445
2506	0.9490	0.9433
2724	0.9474	0.9421
3127	0.9450	0.9410
3907	0.9420	0.9398
3970	0.9419	0.9386

Tabla 2.7. Comparación de valores de la supervivencia empírica y estimada, en distintos instantes.

La siguiente matriz determina en cada uno de sus elementos los valores esperados para los tiempos de permanencia en  $j$  habiendo partido del estado  $i$  en el instante  $t = 0$ , en el intervalo  $(0, t)$ :

$$\int_0^t P(\tau) d\tau = \begin{bmatrix} N(t) & M(t) \\ 0 & tI \end{bmatrix}$$

A continuación, veremos cómo calcular a partir de las componentes de la matriz anterior el valor esperado del tiempo de permanencia en un estado transitorio, suponiendo que el paciente ingresa al sistema en el estado  $i = 0$ , en el instante  $t = 0$ . Para dichos cálculos, centraremos nuestra atención en las dos primeras filas de la matriz  $N(t)$ , dado que corresponden a las transiciones desde  $i = 0$  a los demás estados transitorios y que se había logrado un buen ajuste a la distribución del Tiempo de permanencia en el estado 0 con  $p = 2$  fases en la distribución PH.

Para interpretar los elementos de la matriz anterior veamos cómo calcular el tiempo medio de permanencia en el estado 1 hasta un instante  $t$ . Recordando además que para el estado transitorio 1, se había logrado un buen ajuste para  $p = 4$  fases en la distribución PH, definimos:

- $X_{11}(t)$  la variable aleatoria que cuantifica el tiempo de permanencia en la fase 1 (virtual) del estado transitorio  $j = 1$  del sistema, en el intervalo  $(0, t)$ , habiendo iniciado el proceso en alguna de las fases del estado 0,
- $X_{12}(t)$  la variable aleatoria que cuantifica el tiempo de permanencia en la fase 2 (virtual) del estado  $j = 1$  del sistema, en el intervalo  $(0, t)$ , habiendo iniciado el proceso en alguna de las fases del estado 0,
- $X_{13}(t)$  la variable aleatoria que cuantifica el tiempo de permanencia en la fase 3 (virtual) del estado transitorio  $j = 1$  del sistema, en el intervalo  $(0, t)$ , habiendo iniciado el proceso en alguna de las fases del estado 0,
- $X_{14}(t)$  la variable aleatoria que cuantifica el tiempo de permanencia en la fase 4 (virtual) del estado  $j = 1$  del sistema, en el intervalo  $(0, t)$ , habiendo iniciado el proceso en alguna de las fases del estado 0,

- $X_1(t) = X_{11}(t) + X_{12}(t) + X_{13}(t) + X_{14}(t)$  la variable Tiempo de permanencia en el estado  $j = 1$  del sistema, en el intervalo  $(0, t)$ , habiendo iniciado el proceso en alguna de las fases del estado  $i = 0$ ,
- $Y$  la variable aleatoria discreta que representa la fase en la que se inicia el proceso en el estado  $i = 0$ . Aunque sabemos que son fases virtuales, tienen asignadas probabilidades en el modelo de ajuste para el estado 0, que ya han sido estimadas y constituyen las entradas del vector  $\alpha_0$ . En este caso el recorrido de dicha variable es  $Rec(Y) = \{1, 2\}$ , teniendo en cuenta que se logró un excelente ajuste para la variable *Tiempo de permanencia* en 0, con una distribución PH de orden 2 y que por lo tanto en el estado  $i = 0$  hay solamente dos fases virtuales.

En el modelo de este sistema, la matriz  $N(t)$  es cuadrada de orden 12 y dado que en cualquier caso el proceso se inicia en el estado 0, estaremos interesados en las dos primeras filas, que corresponden a sus dos fases virtuales. Así por ejemplo para el estado transitorio  $j = 1$  (aunque este puede variar  $j = 0, 1, 2, 3$  y por tanto las columnas de  $N(t)$  que tengamos en cuenta):

$$\begin{aligned}
 n_{13}(t) &= E[X_{11}(t)|Y = 1] & n_{14}(t) &= E[X_{12}(t)|Y = 1] \\
 n_{15}(t) &= E[X_{13}(t)|Y = 1] & n_{16}(t) &= E[X_{14}(t)|Y = 1] \\
 n_{23}(t) &= E[X_{11}(t)|Y = 2] & n_{24}(t) &= E[X_{12}(t)|Y = 2] \\
 n_{25}(t) &= E[X_{13}(t)|Y = 2] & n_{26}(t) &= E[X_{14}(t)|Y = 2]
 \end{aligned}$$

Aplicando la definición y propiedades de esperanza condicional, se tiene:

$$\begin{aligned}
 E[X_1(t)|Y = 1] &= E[X_{11}(t) + X_{12}(t) + X_{13}(t) + X_{14}(t) | Y = 1] \\
 &= E[X_{11}(t)|Y = 1] + E[X_{12}(t)|Y = 1] + E[X_{13}(t)|Y = 1] + E[X_{14}(t)|Y = 1] \\
 &= n_{13}(t) + n_{14}(t) + n_{15}(t) + n_{16}(t)
 \end{aligned}$$

$$\begin{aligned}
 E[X_1(t)|Y = 2] &= E[X_{11}(t) + X_{12}(t) + X_{13}(t) + X_{14}(t)|Y = 2] \\
 &= E[X_{11}(t)|Y = 2] + E[X_{12}(t)|Y = 2] + E[X_{13}(t)|Y = 2] + E[X_{14}(t)|Y = 2] \\
 &= n_{23}(t) + n_{24}(t) + n_{25}(t) + n_{26}(t)
 \end{aligned}$$

$$\begin{aligned}
 E[X_1(t)] &= \int_0^{+\infty} E[X_1(t)|Y = y] dP_Y(y) \\
 &= \sum_{y=1}^{y=2} E[X_1(t)|Y = y] P(Y = y)
 \end{aligned}$$

Luego se tiene:

$$\begin{aligned}
 m_1(t) = E[X_1(t)] &= E[X_1(t)|Y = 1]P(Y = 1) + E[X_1(t)|Y = 2]P(Y = 2) \\
 &= [n_{13}(t) + n_{14}(t) + n_{15}(t) + n_{16}(t)]P(Y = 1) + \\
 &+ [n_{23}(t) + n_{24}(t) + n_{25}(t) + n_{26}(t)]P(Y = 2)
 \end{aligned}$$

**A nivel muestral:** para cada estado transitorio  $j = 0$ ,  $j = 1$ ,  $j = 2$  o  $j = 3$  es posible recuperar los tiempos de permanencia en el estado antes de que el paciente transite a otro estado o salga del sistema (hacia  $P$  o  $NP$ ), ordenarlos en forma ascendente y una vez establecida una partición conveniente del rango de variación  $(t_1, \dots, t_k, \dots, t_n)$ , calcular el *tiempo medio de permanencia en  $j$* , hasta el instante  $t = t_k$ . Tendremos en cuenta que

cualquier paciente  $u$  ingresa al sistema en el instante  $t = 0$  a través del estado  $i = 0$  y permanece en el estado  $j$  durante el intervalo de tiempo definido por  $[a_{uj}, b_{uj}]$ , razón por la cual omitiremos el indicador  $i$  de estado de partida.

Siendo  $\{t_1 = 400, \dots, t_k, \dots, t_{10} = 4000\}$  una partición razonable establecida durante el período de observación, se realizará un seguimiento en esos valores. Si denotamos  $\tau_{ujk}$  el Tiempo de permanencia en el estado  $j$ , para el paciente  $u$ , en el intervalo  $[0, t_k]$ , con  $k = 1, 2, \dots, 10$ , tenemos:

$$\tau_{ujk} = \begin{cases} 0, & t_k < a_{uj} \\ t_k - a_{uj}, & a_{uj} \leq t_k \leq b_{uj} \\ b_{uj} - a_{uj}, & t_k > b_{uj} \end{cases}$$

Idéntico procedimiento, puede hacerse para los tiempos medios de permanencia en el sistema, en cuyo caso cada paciente  $u$  ingresa al sistema en  $t = 0$  en el estado 0, razón por la cual  $a_u = 0$ .

Posteriormente, se han calculado los tiempos medios empíricos para cada una de estas variables  $\overline{\tau}_{jk} = m_{jk}$ , en cada estado transitorio  $j$ :  $m_{jk} = m_j(t_k)$  y análogamente los tiempos medios de permanencia en el sistema, se denotarán como  $m_k = m(t_k)$ .

La representación gráfica de las funciones  $E[X_j(t)]$  en contraste con los valores medios muestrales  $m_j(t_k)$ , siendo  $j \in \{0, 1, 2, 3\}$ , son las que se presentan en las Figuras [2.35](#), [2.36](#), [2.37](#), [2.38](#):

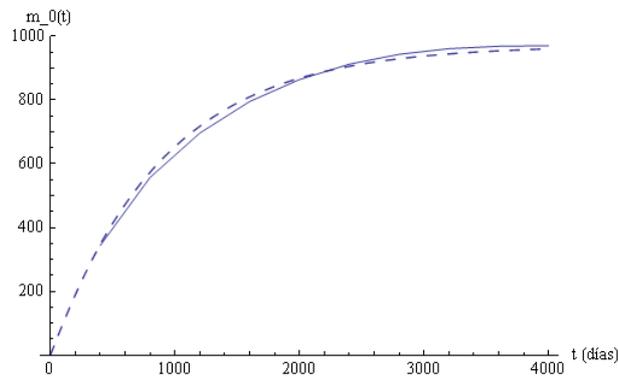


Figura 2.35. Representación gráfica de los tiempos medios muestrales de permanencia en el estado transitorio 0, para cada intervalo  $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante  $t = 0$ , en comparación con los tiempos medios estimados (- -).

Nótese la bondad del ajuste entre la función teórica, que modela el comportamiento de los tiempos medios de permanencia en 0 en forma teórica, en relación a los valores muestrales. No será así en los siguientes casos, donde el modelo creado predice un crecimiento moderado en los tiempos medios de permanencia en 1, 2 y 3, que en la práctica no ocurre. Los valores de los tiempos de permanencia en cada uno de esos estados transitorios incrementan cuando el paciente transita desde dicho estado al de absorción  $NP$ , salida del sistema sin progresión. Es esta la razón por la que se estudiarán separados dos sistemas, cada uno con un estado absorbente  $P$  ó  $NP$ , estableciendo una participación en la población de pacientes y obteniendo mayor homogeneidad en los valores de las variables en consideración. Esta necesidad de indagar en mayor profundidad, ha emergido del propio análisis desarrollado hasta el momento.

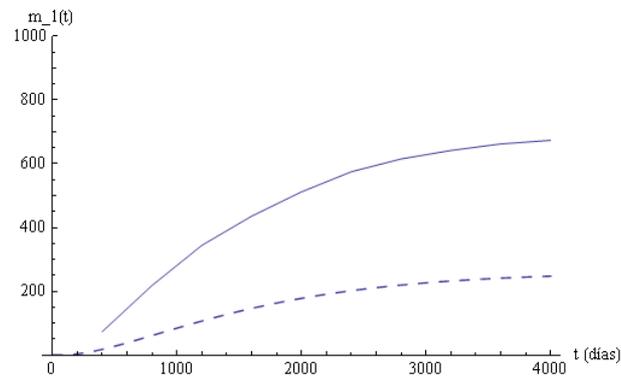


Figura 2.36. Representación gráfica de los tiempos medios muestrales de permanencia en el estado transitorio 1, para cada intervalo  $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante  $t = 0$ , en comparación con los tiempos medios estimados (- -).

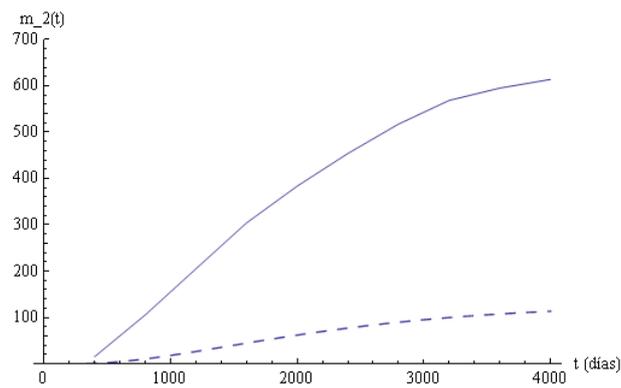


Figura 2.37. Gráfico de los tiempos medios muestrales de permanencia en el estado transitorio 2, para cada intervalo  $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante  $t = 0$ , en comparación con los tiempos medios estimados (- -).

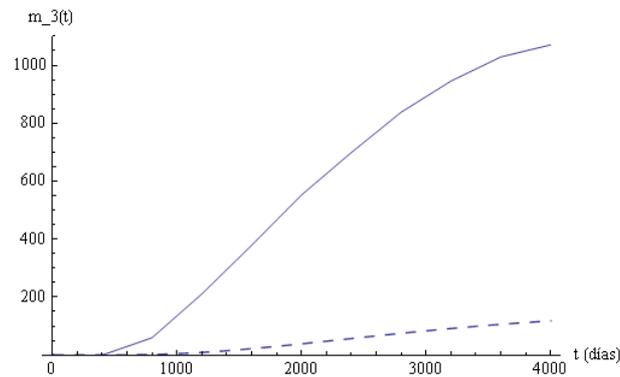


Figura 2.38. Gráfico de los tiempos medios muestrales de permanencia en el estado transitorio 3, para cada intervalo  $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante  $t = 0$ , en comparación con los tiempos medios estimados (- -).

La representación gráfica de la función  $E[X(t)]$ , tiempo medio o valor esperado de la variable *tiempo de permanencia en el sistema* (- -), en contraste con los valores medios muestrales  $m(t_k)$ , es la que se presenta en las Figura 2.39. Nótese la bondad del ajuste, a pesar de que en el sistema se incluyen todos los pacientes, aquellos que salen del sistema por el estado absorbente  $P$  o por  $NP$ .

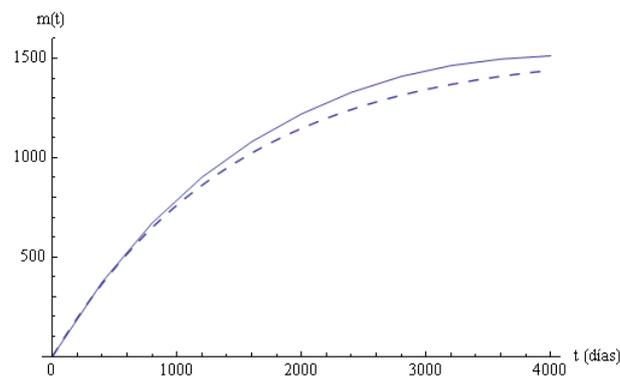


Figura 2.39. Gráfico de los tiempos medios muestrales de permanencia en el sistema, para cada intervalo  $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante  $t = 0$ , en comparación con los tiempos medios estimados (- -).

En la Tabla 2.8, para la misma partición  $\{t_1 = 400, t_2 = 800, \dots, t_{10} = 4000\}$  se explicitan los tiempos medios muestrales de permanencia en cada estado transitorio hasta cada instante  $t_k$ . En la última columna se registran los tiempos medios de permanencia en el sistema  $m(t_k)$ , estimados por el modelo, hasta dicho instante. Los valores de la última fila corresponden a los tiempos medios de permanencia en el estado transitorio o en el sistema, según corresponda, hasta el instante  $t = 4000$  (días).

$t$ (en días)	$m_0(t)$	$m_1(t)$	$m_2(t)$	$m_3(t)$	$m_{Sist.emp}(t)$	$m_{Sist.est}(t)$
400	343.74	73.43	15.38	0.00	374.83	366.18
800	557.33	218.38	105.54	59.94	671.07	647.48
1200	696.55	344.17	204.95	210.82	901.30	860.67
1600	795.45	435.91	304.14	379.21	1080.42	1023.11
2000	864.24	512.02	383.55	552.66	1220.35	1147.25
2400	912.58	575.06	454.09	698.99	1329.00	1242.00
2800	943.44	615.08	517.39	837.98	1408.83	1314.01
3200	961.20	641.85	568.52	945.43	1463.32	1368.39
3600	968.66	662.31	595.12	1028.79	1496.45	1409.14
4000	970.44	673.63	613.78	1070.39	1512.58	1439.42

Tabla 2.8. Tiempos medios de permanencia en cada estado y en el sistema, hasta cierto instante  $t$ .

Si se analizan las dos últimas columnas de la tabla anterior es claro que tiempos medios muestrales y estimados a partir del modelo que hemos creado, se ajustan muy bien. Para visualizarlo más claramente hemos considerado las variables  $X = m_{Sist.emp}(t)$  e  $Y = m_{Sist.est}(t)$ , realizando un análisis de la correlación entre ambas y planteado un modelo lineal que ha presentado excelente ajuste:  $Y = aX + b + \varepsilon$ , con un valor para  $R_a^2 \cong 0.9999$ . El mejor ajuste lineal, se ha obtenido para  $\hat{a} \cong 0.94$  y  $\hat{b} \cong 0$ , no habiendo rechazado la hipótesis nula en el contraste:  $H_0 : b = 0$ ,  $H_1 : b \neq 0$ . En el contraste  $H_0 : a = 0$ ,  $H_1 : a \neq 0$  se ha rechazado la hipótesis nula, a partir de un  $p$ -valor igual a  $2 \times 10^{-16}$ , claramente significativo.

La diferencia que se ha mencionado entre los procesos que siguen los pacientes que salen del sistema sin progresión y los que lo hacen con progresión genera desajuste entre los tiempos medios muestrales de permanencia en cada estado transitorio y los que se obtienen en forma teórica, a partir del modelo probabilístico.

El sistema con el que se ha trabajado hasta ahora, con dos estados absorbentes, ha permitido calcular las medidas de interés para establecer inferencias sobre la población de todos los pacientes con esta enfermedad, que derivan o no en progresión. Sin embargo, los valores esperados de los tiempos de permanencia en cada estado transitorio para distintos instantes  $t$ , se han visto distorsionados por la heterogeneidad del grupo en relación a la variable estudiada. A pesar de lo expuesto, el modelo global sí ha presentado una bondad de ajuste satisfactoria para la distribución del Tiempo de permanencia en el sistema, así como en cada estado.

Las medidas de tendencia central y de dispersión, resultan ser notablemente distintas si se analizan por bloques, para los dos tipos de pacientes. Tratándose de pacientes que derivan en  $P$ , los tiempos de permanencia en cada estado transitorio y en general en el sistema disminuyen hasta en un 50% en relación a los de los pacientes que salen del sistema sin progresión. Esto pone en evidencia la necesidad de separar la población en bloques, dado que se requiere analizar todas las medidas de interés en uno y otro caso. Será posible entonces analizar los tiempos medios de permanencia en los distintos estados y las implicancias que esto tiene en la supervivencia de los pacientes y también en los costes por sus tratamientos.

A continuación y en virtud de lo observado, la población será subdividida en dos estratos más homogéneos, definiendo dos modelos con ajuste satisfactorio que describen las etapas de la enfermedad, en cada caso.

## 2.4. Sistema con las transiciones de los pacientes que derivan en progresión

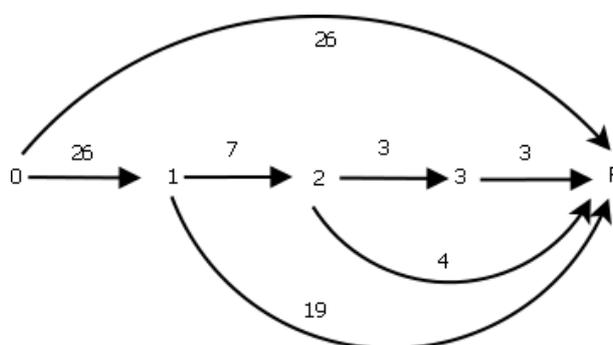


Figura 2.40. Diagrama del sistema multiestado, con un estado absorbente P, para los  $n=52$  pacientes que derivan en progresión

El resumen descriptivo de los tiempos de permanencia en el sistema y en cada estado transitorio, se presenta en la Tabla 2.9. Nótese cómo en cada estado transitorio, al menos en los que el tamaño muestral así lo permite, las medidas de tendencia central y de dispersión son notablemente inferiores a las correspondientes en el sistema global, como cabe esperar.

	n	Rango	Mínimo	Máximo	Mediana	Media	Desviación Típica
Sistema	52	3878	92	3970	449.50	832.71	961.33
0	52	3878	92	3970	250.00	453.77	640.33
1	26	2261	87	2348	284.50	525.08	559.21
2	7	928	140	1068	227.00	416.29	362.87
3	3	2010	206	2216	717.00	1046.33	1044.69

Tabla 2.9. Descripción estadística de los tiempos en cada estado transitorio.

La distribución tipo fase, con buen ajuste para la variable *Tiempo de permanencia en el sistema*, tiene como vector de probabilidades iniciales  $\alpha = (0.0826, 0.9174)$ .

La matriz generadora  $Q$ , cuadrada y de orden 3, tendrá dos bloques principales  $A$  y  $B$ . Cada una de las matrices que componen los bloques de  $Q$ , surgen de una estimación realizada a partir de la muestra de observaciones y las notaciones correspondientes se han simplificado:

$$Q = \begin{bmatrix} A_{2 \times 2} & B_{2 \times 1} \\ 0_{1 \times 2} & 0_{1 \times 1} \end{bmatrix}$$

Donde los dos bloques principales vienen dados por:

$$A = \begin{bmatrix} -0.0011 & 0.0008 \\ 0.0004 & -0.0022 \end{bmatrix} \quad B = \begin{bmatrix} 0.0003 \\ 0.0017 \end{bmatrix}$$

Las funciones de interés, se presentan en las Figuras 2.41, 2.42, 2.43 y 2.44.

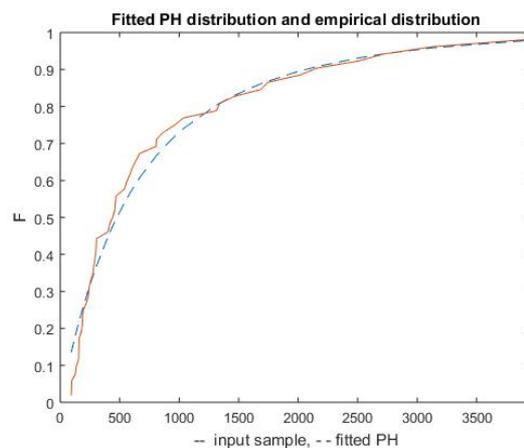


Figura 2.41. Gráfico de la función de distribución para el sistema multiestado, con un estado absorbente P y  $n=52$  pacientes que derivan en progresión

El valor del estadístico de Kolmogorov-Smirnov es aproximadamente  $D = 0.1352$ , que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.1849$  y  $D_{\alpha=0.01} = 0.2216$ . Al no pertenecer el valor de  $D$  a las regiones críticas con extremos inferiores 0.1849 y 0.2216, admitimos que el ajuste logrado a partir de la distribución estimada con su correspondiente

función de distribución empírica antes representada, es satisfactorio, para ambos niveles de significancia.

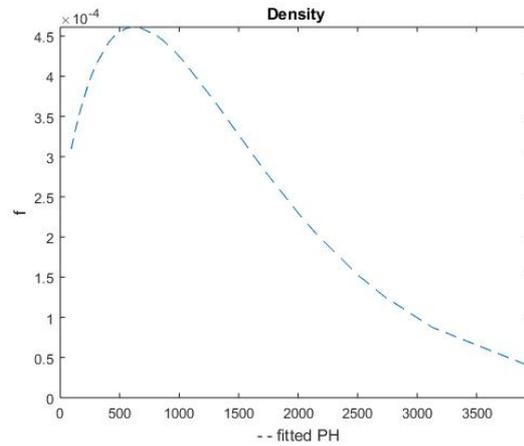


Figura 2.42. Gráfico de la función de densidad para el sistema multiestado, con un estado absorbente P y  $n=52$  pacientes que derivan en progresión

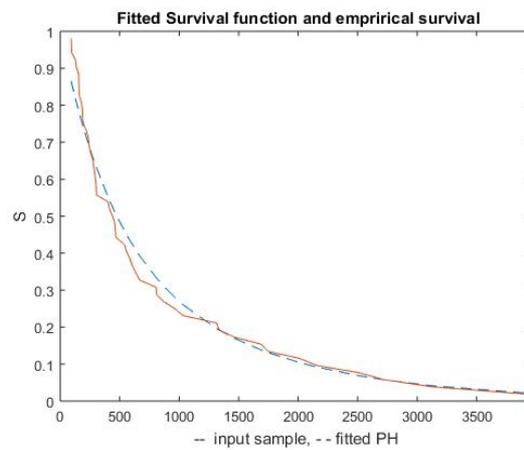


Figura 2.43. Gráfico de la función de supervivencia para el sistema multiestado, con un estado absorbente P y  $n=52$  pacientes que derivan en progresión

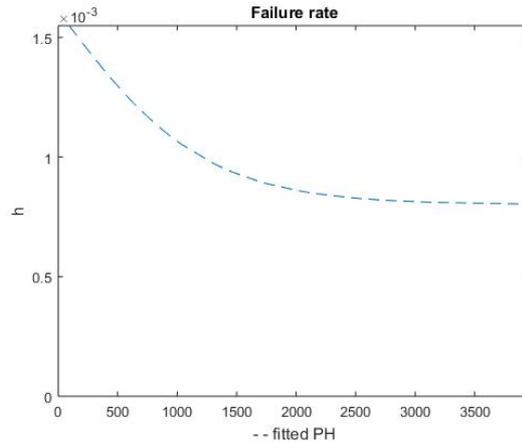


Figura 2.44. Gráfico de la función razón de fallo para el sistema multiestado, con un estado absorbente P y  $n=52$  pacientes que derivan en progresión

En la Tabla 2.10 se evalúa la función supervivencia  $S(t)$  en algunos valores de tiempo que se han especificado. Los resultados corresponden a las respectivas probabilidades de que hasta ese instante, el paciente permanezca sin progresión. En la última columna de la misma tabla, se presenta los respectivos valores de la función razón de fallo  $h(t)$  que es también decreciente.

t (en días)	S(t)	h(t)
92	0.8648	0.0015
95	0.8608	0.0015
96	0.8594	0.0015
128	0.8182	0.0015
135	0.8096	0.0015
155	0.7854	0.0015
156	0.7842	0.0015
158	0.7819	0.0015
159	0.7807	0.0015
178	0.7588	0.0015
187	0.7487	0.0015
187	0.7487	0.0015
194	0.7409	0.0015
221	0.7120	0.0015
237	0.6956	0.0015
242	0.6905	0.0015
254	0.6786	0.0014
274	0.6594	0.0014
281	0.6528	0.0014
290	0.6445	0.0014
298	0.6373	0.0014
303	0.6328	0.0014
304	0.6319	0.0014
403	0.5510	0.0014
414	0.5429	0.0013
440	0.5244	0.0013
459	0.5113	0.0013
464	0.5080	0.0013
469	0.5046	0.0013
540	0.4603	0.0013
559	0.4493	0.0013
586	0.4344	0.0012
607	0.4232	0.0012
636	0.4083	0.0012
669	0.3923	0.0012
806	0.3340	0.0011
809	0.3328	0.0011
869	0.3110	0.0011
962	0.2808	0.0011
1036	0.2595	0.0011
1313	0.1962	0.0010
1338	0.1915	0.0010
1462	0.1702	0.0009
1687	0.1385	0.0009
1746	0.1314	0.0009
2010	0.1043	0.0009
2159	0.0919	0.0008
2506	0.0687	0.0008
2724	0.0574	0.0008
3127	0.0414	0.0008
3907	0.0221	0.0008
3970	0.0210	0.0008

Tabla 2.10. Estimación de la supervivencia y razón de fallo, en distintos valores de tiempo.

Es posible calcular los tiempos medios de permanencia en este sistema con estado absorbente  $P$  y compararlos con los tiempos medios muestrales. La función  $m(t)$  de *tiempos medios de permanencia en el sistema*, en el intervalo  $[0, t]$ , se obtiene a partir de las componentes de la matriz  $N(t)$  aplicando esperanzas condicionadas y a partir del vector de probabilidades iniciales estimado en el modelo en forma similar al caso que ya se explicó sobre el sistema global anterior:

$$m(t) = 0.0826 \cdot (n_{11}(t) + n_{12}(t)) + 0.9174 \cdot (n_{21}(t) + n_{22}(t))$$

En la Figura 2.45 se presenta el gráfico de la función *tiempos medios de permanencia en el sistema*  $m(t)$  estimada (- -), en comparación con los valores de los tiempos medios muestrales calculados a partir de los  $n = 52$  casos de pacientes que derivan en progresión. Puede apreciarse el buen ajuste logrado por el modelo.

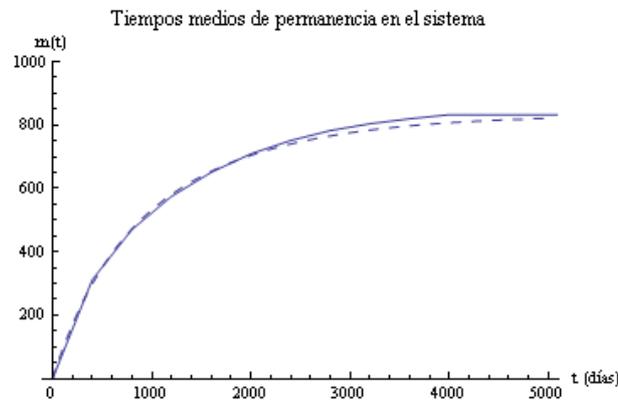


Figura 2.45. Tiempos medios estimados de permanencia en el sistema multiestado y estado absorbente P (- -), en comparación con los tiempos medios en la muestra de  $n=52$  pacientes que derivan en progresión

### Análisis de los tiempos en cada estado transitorio

En este caso, al haber homogeneizado los valores de la variable considerada a partir de la subdivisión de la población en dos estratos, podrá observarse que se obtienen ajustes satisfactorios no solamente en el modelo global para  $n = 52$  pacientes, sino que además el ajuste es muy bueno en cada estado. Esto llama la atención, ya que el número de pacientes que accede a cada estado transitorio es cada vez menor y muy limitado.

En virtud de lo observado hasta el momento, nos proponemos construir una distribución *tipo fase* con ajuste satisfactorio para los tiempos de permanencia en los estados transitorios del modelo que representa la evolución de la enfermedad para los pacientes que derivan en progresión, a partir de los datos.

Para cada estado transitorio 0, 1, 2 y 3, según se ha explicado en los objetivos de este trabajo, se construirá la distribución *tipo fase* (PH) que mejor se ajuste a la de cada variable aleatoria *Tiempo de permanencia en el estado*.

Igual que en el sistema global, se empleará el algoritmo EM de ajuste con distribuciones de este tipo (véase [32]).

Los ajustes han resultado ser adecuados para  $p = 2$  y  $N = 1000$  en los cuatro casos, para un nivel de significación de 0.05 y también para 0.01.

#### ■ Ajuste para los tiempos de permanencia en el estado 0

A partir del algoritmo EM se ha realizado un ajuste con distribución *PH* para las 52 observaciones correspondientes a los tiempos de permanencia en el estado inicial 0, antes de la primera intervención quirúrgica a partir de él ó antes de la salida del sistema. De los 52 pacientes registrados en la base de datos y a partir de la primera intervención quirúrgica para extirpar el tumor no invasivo de los músculos de la vejiga (en el instante 0):

- 26 salen del sistema con progresión (estado *P*).
- 26 realizan una primera consulta de control en la que se diagnostica recidiva sin progresión.

Se obtuvo las siguientes estimaciones del vector  $\alpha_0$ , la matriz  $A_0$  y la matriz columna de absorción  $A_0^0$ :

$$\widehat{\alpha}_0 = \begin{bmatrix} 1.0000 \\ 0.0000 \end{bmatrix}, \widehat{A}_0 = \begin{bmatrix} -0.0159 & 0.0159 \\ 0.0010 & -0.0037 \end{bmatrix} \text{ y } \widehat{A}_0^0 = \begin{bmatrix} 0.0000 \\ 0.0027 \end{bmatrix}$$

En las Figuras 2.46 y 2.47, se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la

función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 0* y la función de supervivencia, ambas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

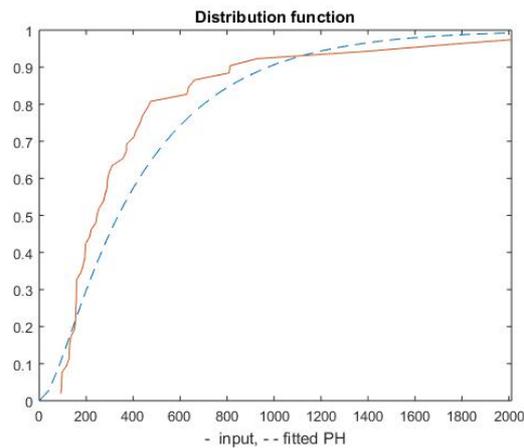


Figura 2.46. Gráfico de la función de distribución tipo fase estimada para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 52$  casos

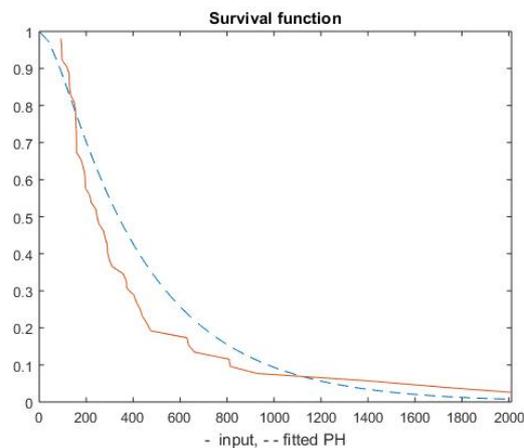


Figura 2.47. Gráfico de la función de supervivencia estimada para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 52$  casos.

El valor del estadístico de Kolmogorov-Smirnov es  $D = 0.1682$  que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.1849$  y  $D_{\alpha=0.01} = 0.2216$ . Al no pertenecer el valor

de  $D$  a la región crítica admitimos que el ajuste con esta distribución  $PH(\hat{\alpha}, \hat{T})$  es satisfactorio, para ambos niveles de significación.

■ **Ajuste para los tiempos de permanencia en el estado 1**

A partir del algoritmo EM se ha realizado un ajuste con distribución tipo  $PH$  para las 26 observaciones correspondientes a los tiempos de permanencia en el estado 1, antes de la segunda intervención quirúrgica posterior a la inicial (estado 2) ó pasaje a progresión de la enfermedad (estado  $P$ ). De las 26 observaciones:

- 7 transitan a 2 sin salir del sistema.
- 19 salen del sistema con progresión de la enfermedad.

Se obtuvo las siguientes estimaciones del vector  $\alpha_1$ , de la matriz  $A_1$  y de la matriz columna de absorción  $A_1^0$ :

$$\hat{\alpha}_1 = \begin{bmatrix} 1.0000 \\ 0.0000 \end{bmatrix}, \hat{A}_1 = \begin{bmatrix} -0.0166 & 0.0166 \\ 0.0011 & -0.0034 \end{bmatrix} \text{ y } \hat{A}_1^0 = \begin{bmatrix} 0.0000 \\ 0.0023 \end{bmatrix}$$

En las Figuras 2.48 y 2.49 se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 1* y la función de supervivencia, ambas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

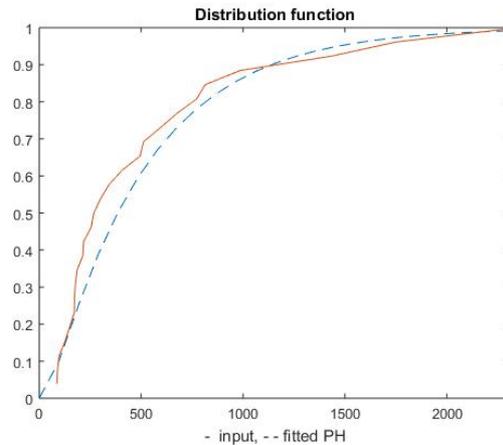


Figura 2.48. Gráfico de la función de distribución tipo fase estimada para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 26$  casos

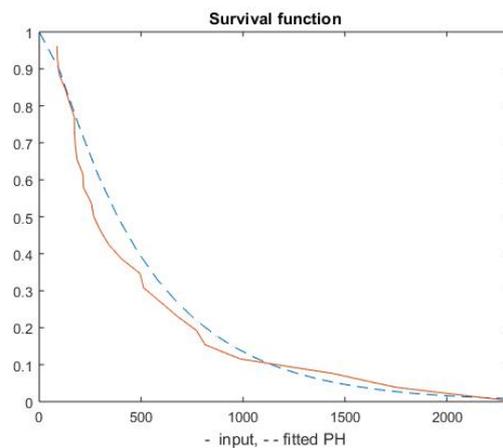


Figura 2.49. Gráfico de la función de supervivencia estimada para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 26$  casos.

El valor del estadístico de Kolmogorov-Smirnov es  $D = 0.1404$  que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.2591$  y  $D_{\alpha=0.01} = 0.3107$ . Al no pertenecer el valor de  $D$  a la región crítica admitimos que el ajuste con esta distribución  $PH(\hat{\alpha}, \hat{T})$  es satisfactorio para ambos niveles de significación.

- **Ajuste para los tiempos de permanencia en el estado 2**

A partir del algoritmo EM se ha realizado un ajuste con distribución tipo *PH* para las 7 observaciones correspondientes a los tiempos de permanencia en el estado 2, antes de la tercera intervención quirúrgica por recidiva recurrente ó antes de la salida del sistema con progresión. De los 7 casos:

- 4 salen del sistema con progresión (estado *P*).
- 3 transitan al estado 3 con recidiva sin progresión.

Se obtuvo las siguientes estimaciones del vector  $\alpha_2$ , la matriz  $A_2$  y la matriz columna de absorción  $A_2^0$ :

$$\widehat{\alpha}_2 = \begin{bmatrix} 1.0000 \\ 0.0000 \end{bmatrix}, \widehat{A}_2 = \begin{bmatrix} -0.0120 & 0.0120 \\ 0.0009 & -0.0041 \end{bmatrix} \text{ y } \widehat{A}_2^0 = \begin{bmatrix} 0.0000 \\ 0.0032 \end{bmatrix}$$

En las Figuras 2.50 y 2.51 se presentan las salidas que ofrece Matlab, a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 2* y la función de supervivencia, ambas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste a pesar del tamaño de la muestra.

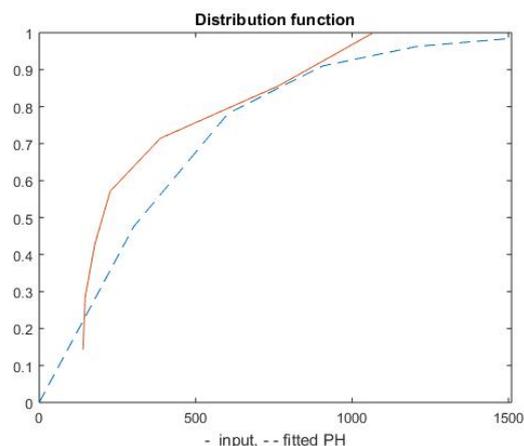


Figura 2.50. Gráfico de la función de distribución tipo fase estimada para el *Tiempo de permanencia en 2*, a partir de la muestra de  $n = 7$  casos

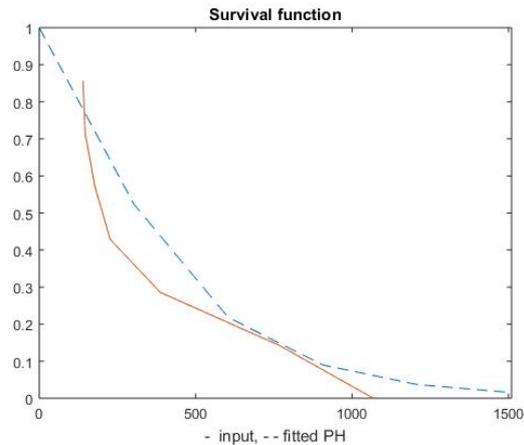


Figura 2.51. Gráfico de la función de supervivencia estimada para el *Tiempo de permanencia en 2*, a partir de la muestra de  $n = 7$  casos.

El valor del estadístico de Kolmogorov-Smirnov es  $D = 0.2174$  que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.4837$  y  $D_{\alpha=0.01} = 0.5799$ . Al no pertenecer el valor de  $D$  a la región crítica admitimos que el ajuste con esta distribución  $PH(\hat{\alpha}, \hat{T})$  es satisfactorio, para ambos niveles de significación.

#### ■ Ajuste para los tiempos de permanencia en el estado 3

A partir del algoritmo EM se ha realizado un ajuste con distribución tipo  $PH$  para las 3 observaciones correspondientes a los tiempos de permanencia en el estado 3, antes del pasaje a progresión de la enfermedad (estado  $P$ ).

Se obtuvo las siguientes estimaciones del vector  $\alpha_3$ , de la matriz  $A_3$  y de la matriz columna de absorción  $A_3^0$ :

$$\hat{\alpha}_3 = \begin{bmatrix} 1.0000 \\ 0.0000 \end{bmatrix}, \hat{A}_3 = \begin{bmatrix} -0.0120 & 0.0120 \\ 0.0004 & -0.0015 \end{bmatrix} \text{ y } \hat{A}_3^0 = \begin{bmatrix} 0.0000 \\ 0.0011 \end{bmatrix}$$

En las Figuras 2.52 y 2.53 se presentan las salidas que ofrece Matlab, a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 3* y la función de supervivencia, ambas en contraste con

los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste a pesar de tan solo  $n = 3$  casos.

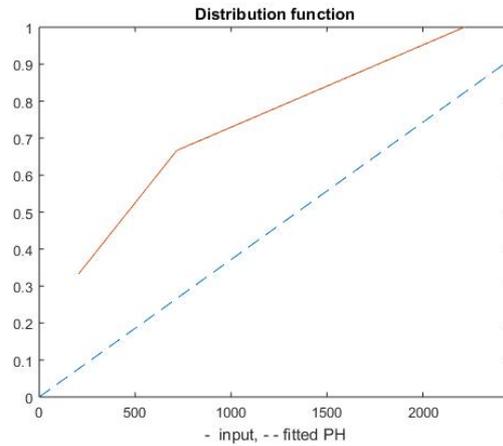


Figura 2.52. Gráfico de la función de distribución tipo fase estimada para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 3$  casos

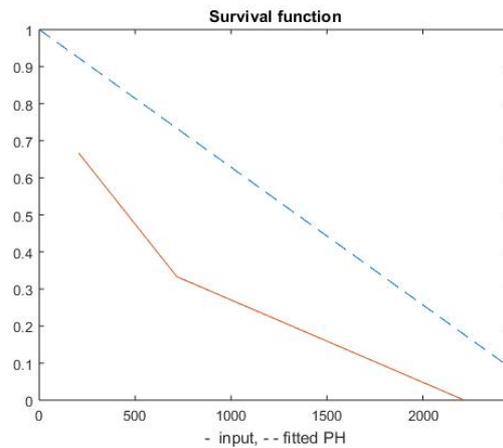


Figura 2.53. Gráfico de la función de supervivencia estimada para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 3$  casos.

El valor del estadístico de Kolmogorov-Smirnov es  $D = 0.2233$  que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.7089$  y  $D_{\alpha=0.01} = 0.8499$ . Al no pertenecer el valor de  $D$  a la región crítica admitimos que el ajuste con esta distribución  $PH(\hat{\alpha}, \hat{T})$  es satisfactorio, para ambos niveles de significación.

En la Tabla 2.11 se evalúan las funciones de supervivencia en los estados transitorios en distintos valores de tiempo  $t$ , que se especifican en la primera columna. En la última columna se registran los valores de la función supervivencia estimada en el sistema:

t(en días)	$S_0(t)$	$S_1(t)$	$S_2(t)$	$S_3(t)$	$S_{Sys}(t)$
200	0.7021	0.7378	0.6944	0.8791	0.7344
400	0.4265	0.4840	0.3959	0.7202	0.5533
600	0.2572	0.3160	0.2207	0.5860	0.4268
800	0.1551	0.2063	0.1227	0.4764	0.3363
1000	0.0935	0.1346	0.0682	0.3874	0.2696
1200	0.0564	0.0879	0.0379	0.3149	0.2193
1400	0.0340	0.0574	0.0211	0.2560	0.1805
1600	0.0205	0.0374	0.0117	0.2082	0.1498
1800	0.0123	0.0244	0.0065	0.1692	0.1252
2000	0.0074	0.0160	0.0036	0.1376	0.1052
2200	0.0045	0.0104	0.0020	0.1119	0.0887
2400	0.0027	0.0068	0.0011	0.0910	0.0750
2600	0.0016	0.0044	0.0006	0.0739	0.0636
2800	0.0010	0.0029	0.0003	0.0601	0.0540
3000	0.0006	0.0019	0.0002	0.0489	0.0459
3200	0.0004	0.0012	0.0001	0.0397	0.0390
3400	0.0002	0.0008	0.0001	0.0323	0.0332
3600	0.0001	0.0005	0.0000	0.0263	0.0282
3800	0.0001	0.0003	0.0000	0.0214	0.0240
4000	0.0000	0.0002	0.0000	0.0174	0.0205
4200	0.0000	0.0001	0.0000	0.0141	0.0174
4400	0.0000	0.0001	0.0000	0.0115	0.0148
4600	0.0000	0.0001	0.0000	0.0093	0.0126
4800	0.0000	0.0000	0.0000	0.0076	0.0108
5000	0.0000	0.0000	0.0000	0.0062	0.0092

Tabla 2.11. Estimación de valores de la función supervivencia en cada estado transitorio y en el sistema con estado absorbente P.

### Generador del Sistema y cálculo de las medidas de interés

La matriz generadora  $Q$ , cuadrada y de orden 9, tendrá dos bloques principales  $A$  y  $B$ . Los bloques de  $A$  que definen las ocho primeras filas y ocho primeras columnas, tienen

dimensión  $2 \times 2$  y caracterizan las transiciones entre los cuatro estados transitorios. Por otro lado la última columna que conforma  $B$  contiene las razones de absorción por parte del estado absorbente  $P$ . La matriz  $Q$  se completa con bloques de matrices nulas de dimensiones correspondientes a cada caso. Cada una de las matrices que componen los bloques de  $Q$  surgen de una estimación realizada a partir de la muestra de observaciones y las notaciones correspondientes se han simplificado:

$$Q = \begin{bmatrix} A_{8 \times 8} & B_{8 \times 1} \\ 0_{1 \times 8} & 0_{1 \times 1} \end{bmatrix}$$

$$A = \begin{bmatrix} A_0 & A_0^0 p_{01} \cdot \alpha_1 & 0 & 0 \\ 0 & A_1 & A_1^0 p_{12} \alpha_2 & 0 \\ 0 & 0 & A_2 & A_2^0 p_{23} \alpha_3 \\ 0 & 0 & 0 & A_3 \end{bmatrix}$$

$$B = \begin{bmatrix} p_{0P} A_0^0 \\ p_{1P} A_1^0 \\ p_{2P} A_2^0 \\ p_{3P} A_3^0 \end{bmatrix}$$

Teniendo en cuenta que las probabilidades de transición deben sumar 1, se tiene:

$$p_{i(i+1)} + p_{iP} = 1, \forall i \in \{0, 1, 2\}$$

Estas probabilidades se han estimado a partir de las proporciones muestrales:

$$p_{01} = 26/52 \approx 0.5000 \quad p_{0P} = 26/52 \approx 0.5000$$

$$p_{12} = 7/26 \approx 0.2692 \quad p_{1P} = 19/26 \approx 0.7308$$

$$p_{23} = 3/7 \approx 0.4286 \quad p_{2P} = 4/7 \approx 0.5714$$

$$p_{3P} = 3/3 \approx 1.0000$$

Las matrices  $A$  y  $B$  que componen los dos bloques principales de  $Q$ , son respectivamente:

$$A = \left[ \begin{array}{cc|cc|cc|cc} -0.0159 & 0.0159 & 0.0000 & 0.0000 & 0 & 0 & 0 & 0 \\ 0.0010 & -0.0037 & 0.0014 & 0.0000 & 0 & 0 & 0 & 0 \\ \text{---} & \text{---} \\ 0 & 0 & -0.0166 & 0.0166 & 0.0000 & 0.0000 & 0 & 0 \\ 0 & 0 & 0.0011 & -0.0034 & 0.0006 & 0.0000 & 0 & 0 \\ \text{---} & \text{---} \\ 0 & 0 & 0 & 0 & -0.0120 & 0.0120 & 0.0000 & 0.0000 \\ 0 & 0 & 0 & 0 & 0.0009 & -0.0041 & 0.0014 & 0.0000 \\ \text{---} & \text{---} \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.0120 & 0.0120 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.0004 & -0.0015 \end{array} \right]$$

$$B = \begin{bmatrix} 0.0000 \\ 0.0014 \\ 0.0000 \\ 0.0017 \\ 0.0000 \\ 0.0018 \\ 0.0000 \\ 0.0011 \end{bmatrix}$$

Luego, a partir de cálculo simbólico es posible obtener los dos bloques principales de la matriz  $P(t)$ ,  $A(t)$  y  $B(t)$ , que se han propuesto en el marco teórico. La matriz  $A(t)$  en cada fila contiene las probabilidades de ocupación de la fase correspondiente a cierto estado transitorio en el instante  $t$ , habiendo ingresado al sistema en la fase del estado que indica la fila. La matriz  $B(t)$  en su elemento  $b_{i1}(t)$ , contiene la probabilidad de ocupación del estado absorbente  $P$  en el instante  $t$ , habiendo ingresado al sistema (cuando  $t = 0$ ) en la fase correspondiente al estado transitorio que la fila  $i$  indica.

Por lo antes expuesto, la supervivencia en el instante  $t$ , que corresponde a la probabilidad de que no haya ocurrido progresión de la enfermedad hasta el instante  $t$ , corresponde a la probabilidad de que hasta ese instante no se haya ocupado el estado  $P$ . Puede calcularse entonces a partir de la matriz  $B(t)$  de la siguiente forma:

$$S(t) = 1 - \alpha.B(t)$$

En la Figura 2.54 se ha representado la función de supervivencia estimada (- -) en el sistema con estado absorbente  $P$ , teniendo en cuenta los tiempos en los estados transitorios y en contraste con la función empírica de supervivencia.

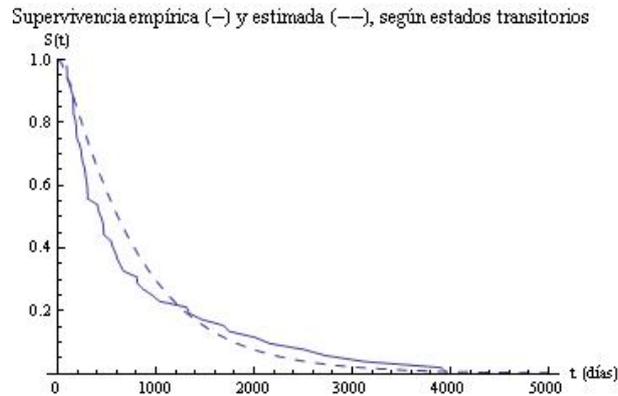


Figura 2.54. Gráfico de la función de supervivencia estimada en el sistema con estado absorbente P, teniendo en cuenta los tiempos en los estados transitorios (- -) en contraste con la función empírica de supervivencia.

En la Figura 2.55 puede apreciarse la similitud de las dos funciones de supervivencia que se han estimado en este sistema. Una de ellas es la que se obtuvo a partir de los tiempos de absorción por parte del estado P, sin tener en cuenta los tiempos en los estados transitorios. La segunda función de supervivencia, se ha obtenido teniendo en cuenta dichos estados y a partir de la construcción del generador  $Q$  de orden  $9 \times 9$ . Cabe esperar dicha semejanza, pero esta última construcción permite obtener mayor información de las distintas etapas de la enfermedad.

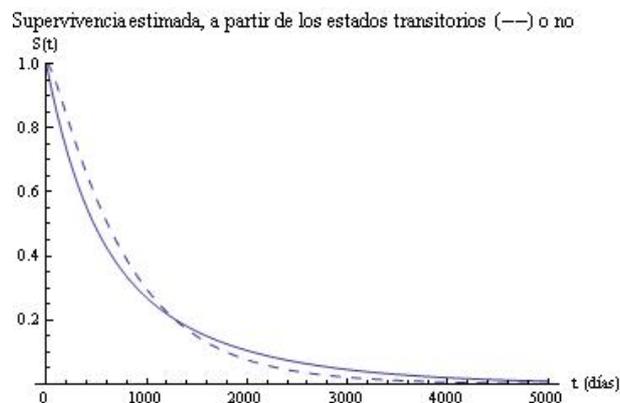


Figura 2.55. Gráficos de las funciones de supervivencia estimadas: sin tener en cuenta los tiempos en los estados transitorios (-) y teniéndolos en cuenta (- -).

Se resume en la Tabla 2.12, las características de los ajustes logrados con distribuciones *tipo fase*, para los tiempos de permanencia en cada estado transitorio:  $n$  el tamaño muestral, los valores  $D$  de los estadísticos de Kolmogorov Smirnov y los límites de las regiones críticas  $D_\alpha$ , evidenciando en cada caso la bondad de los mismos.

Estados	p	n	D	$D_{0.05}$	$D_{0.01}$
0	2	52	0.1682	0.1849	0.2216
1	2	26	0.1404	0.2591	0.03107
2	2	7	0.2174	0.4837	0.5799
3	2	3	0.2233	0.7089	0.8499
Tiempos en el sistema		52	0.1352	0,1849	0,2216

Tabla 2.12. Bondad de los ajustes en el sistema con estado absorbente P.

La siguiente matriz determina en cada uno de sus elementos, los valores esperados para los tiempos de permanencia en  $j$  habiendo partido del estado  $i$  en el instante  $t = 0$ , en el intervalo  $(0, t)$ :

$$\int_0^t P(\tau) d\tau = \begin{bmatrix} N(t) & M(t) \\ 0 & tI \end{bmatrix}$$

Como ya se ha explicado con anterioridad es posible calcular los tiempos medios de permanencia en cada estado transitorio de este sistema (con estado absorbente  $P$ ) y compararlos con los tiempos medios muestrales. La función  $m_j(t)$  de los *tiempos medios de permanencia en cada estado transitorio  $j$*  habiendo partido de 0 en el instante  $t = 0$ , en el intervalo  $[0, t]$ , se obtiene a partir de las componentes de la matriz  $N(t)$ , aplicando esperanzas condicionadas y a partir del vector de probabilidades iniciales estimado en el modelo:

$$m_0(t) = 1 \cdot (n_{11}(t) + n_{12}(t)) + 0 \cdot (n_{21}(t) + n_{22}(t))$$

$$m_1(t) = 1 \cdot (n_{13}(t) + n_{14}(t)) + 0 \cdot (n_{23}(t) + n_{24}(t))$$

$$m_2(t) = 1 \cdot (n_{15}(t) + n_{16}(t)) + 0 \cdot (n_{25}(t) + n_{26}(t))$$

$$m_3(t) = 1 \cdot (n_{17}(t) + n_{18}(t)) + 0 \cdot (n_{27}(t) + n_{28}(t))$$

El ajuste entre la función  $m_j$  y los valores de tiempos medios muestrales, ha sido muy bueno cuando  $j = 0$ , lo que puede apreciarse en la Figura 2.56. En los demás estados transitorios se cuenta con muy pocas observaciones como para obtener buen ajuste, lo que permite deducir grandes discrepancias.

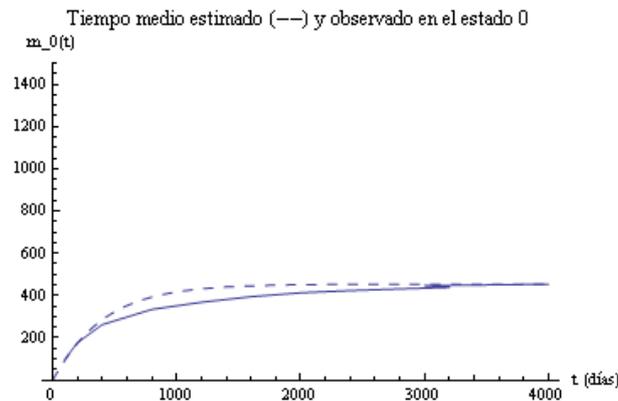


Figura 2.56. Tiempos esperados (- -) y tiempos medios muestrales de permanencia en el estado 0 en el intervalo  $[0, t]$ , habiendo partido en  $t = 0$  del estado 0. Muestra de  $n=52$  pacientes que salen del sistema con progresión

En la Tabla 2.13 se presentan los tiempos medios muestrales de permanencia en los distintos estados transitorios, en distintos intervalos  $[0, t]$ . A su vez, se registran en la sexta columna los tiempos medios empíricos de permanencia en el sistema, hasta el correspondiente instante ( $t$ ) y en la última columna, los tiempos medios estimados en el sistema, a partir del ajuste global que se desarrolló al comienzo, dada la similaridad de las funciones implicadas al tenerlos en cuenta.

$t$ (en días)	$m_0(t)$	$m_1(t)$	$m_2(t)$	$m_3(t)$	$m_{Sist,emp}(t)$	$m_{Sist,est}(t)$
400	262.37	94.69	0	0	312.00	299.34
800	334.42	235.73	74.57	143	470.58	472.46
1200	368.06	317.73	174.86	376	575.31	581.39
1600	394.73	372.08	248.14	642.67	651.25	654.13
2000	412.92	418.23	305.29	805	709.58	704.54
2400	423.58	463.04	312.43	938.33	751.29	740.23
2800	431.27	482.5	369.57	1046.33	782.63	765.79
3200	438.96	497.88	416.29	1046.33	804.31	784.21
3600	446.65	513.27	416.29	1046.33	819.69	797.54
4000	453.77	525.08	416.29	1046.33	832.71	807.20

Tabla 2.13. Tiempos medios de permanencia en cada estado y en el sistema con estado absorbente P, hasta cierto instante  $t$ .

Si se analizan las dos últimas columnas de la tabla anterior, es claro que tiempos medios muestrales y estimados a partir del modelo que hemos creado, se ajustan muy bien. Para visualizarlo más claramente hemos considerado las variables  $X = m_{Sist,emp}(t)$  e  $Y = m_{Sist,est}(t)$ , realizando un análisis de la correlación entre ambas y planteado un modelo lineal que ha presentado excelente ajuste:  $Y = aX + b + \varepsilon$ , con un valor para  $R_a^2 \cong 0.9998$ . El mejor ajuste lineal se ha obtenido para  $\hat{a} \cong 0.98$  y  $\hat{b} \cong 0$ , no habiendo rechazado la hipótesis nula en el contraste:  $H_0 : b = 0$ ,  $H_1 : b \neq 0$ . En el contraste  $H_0 : a = 0$ ,  $H_1 : a \neq 0$  se ha rechazado la hipótesis nula a partir de un  $p$ -valor igual a  $2 \times 10^{-16}$ , claramente significativo.

## 2.5. Sistema con las transiciones de los pacientes que salen del sistema sin progresión

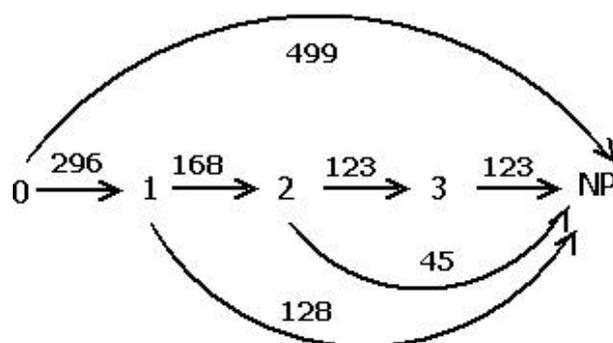


Figura 2.57. Diagrama del sistema multiestado con un estado absorbente NP, para los  $n=795$  pacientes que salen del sistema sin progresión

El resumen descriptivo de los tiempos de permanencia en el sistema y en cada estado transitorio, se presenta en la Tabla 2.14.

	n	Rango	Mínimo	Máximo	Mediana	Media	Desviación Típica
Sistema	795	4729	89	4818	1275.00	1568.29	1129.48
0	795	4248	85	4333	695.00	1004.85	883.94
1	296	3887	95	3982	404.50	693.12	679.58
2	168	4009	83	4092	375.50	637.58	668.99
3	123	3192	82	3274	904.00	1102.90	798.87

Tabla 2.14. Descripción estadística de los tiempos en cada estado transitorio del sistema con estado absorbente NP.

En la tabla anterior puede apreciarse en cada caso que la mediana sigue estando notablemente por debajo del valor medio, el rango es similar en cada caso a los rangos obtenidos en el sistema global, así como también los valores de la mediana, media y desviación típica. Esta última con valores extremos en cada caso, que se justifica por la distorsión de los datos censurados.

La distribución tipo fase con ajuste satisfactorio para la variable *Tiempo de permanencia en el sistema*, ha requerido considerar para el parámetro  $p$  el valor  $p = 4$  y tiene como vector de probabilidades iniciales  $\alpha = (0.0000, 0.0000, 1.0000, 0.0000)$ .

La matriz generadora  $Q$ , cuadrada y de orden 5, tendrá dos bloques principales  $A$  y  $B$ . Cada una de las matrices que componen los bloques de  $Q$ , surgen de una estimación realizada a partir de la muestra de observaciones y las notaciones correspondientes se han simplificado:

$$Q = \begin{bmatrix} A_{4 \times 4} & B_{4 \times 1} \\ 0_{1 \times 4} & 0_{1 \times 1} \end{bmatrix}$$

Donde los dos bloques principales vienen dados por:

$$A = \begin{bmatrix} -0.0015 & 0.0002 & 0.0000 & 0.0013 \\ 0.0024 & -0.0070 & 0.0000 & 0.0006 \\ 0.0004 & 0.0010 & -0.0014 & 0.0000 \\ 0.0000 & 0.0001 & 0.0000 & -0.0015 \end{bmatrix} \quad B = \begin{bmatrix} 0.0001 \\ 0.0040 \\ 0.0000 \\ 0.0014 \end{bmatrix}$$

Las funciones de interés se presentan en las Figuras [2.58](#), [2.59](#), [2.60](#) y [2.61](#).

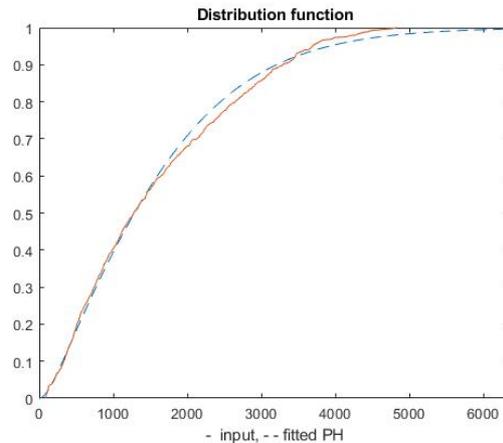


Figura 2.58. Gráfico de la función de distribución para el sistema multiestado con un estado absorbente NP y  $n=795$  pacientes que salen del sistema sin progresión

El valor del estadístico de Kolmogorov-Smirnov es aproximadamente  $D = 0.0415$  que resulta ser menor que  $D_{\alpha=0.05} = 0.0480$  y que  $D_{\alpha=0.01} = 0.0575$ . Al no pertenecer el valor de  $D$  a dichas regiones críticas admitimos que el ajuste logrado a partir de la distribución *tipo fase* con  $p = 4$ , es satisfactorio para ambos niveles de significación. Cabe aclarar que a pesar de tratarse de un conjunto de pacientes que constituyen un estrato de tamaño considerablemente mayor que el de los pacientes que derivan en progresión, los valores de la variable para este caso siguen siendo heterogéneos, teniendo en cuenta que se trata de los tiempos antes de la salida del sistema sin progresión que presentan gran dispersión según pudo observarse en el análisis estadístico descriptivo inicial.

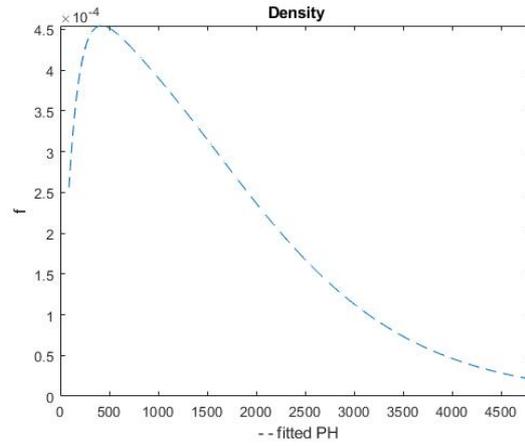


Figura 2.59. Gráfico de la función de densidad para el sistema multiestado, con un estado absorbente NP y  $n=795$  pacientes que salen del sistema sin progresión

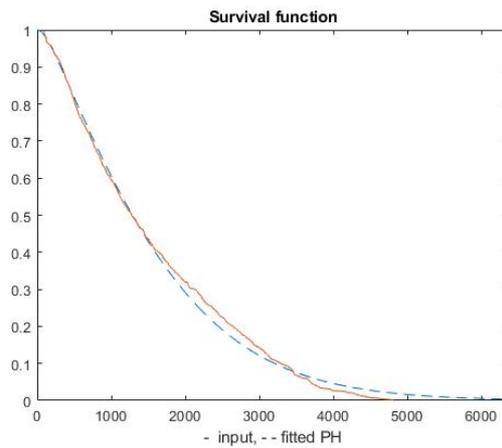


Figura 2.60. Gráfico de la función de supervivencia para el sistema multiestado, con un estado absorbente NP y  $n=795$  pacientes que salen del sistema sin progresión

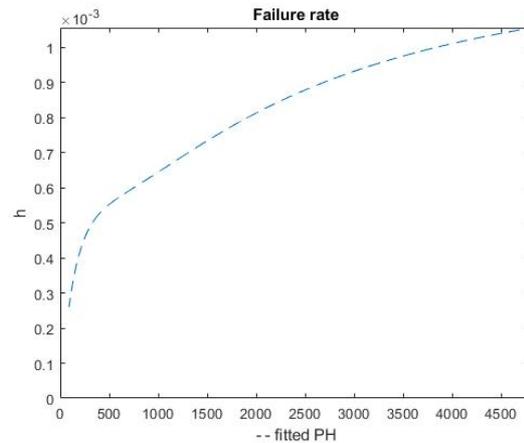


Figura 2.61. Gráfico de la función razón de fallo para el sistema multiestado, con un estado absorbente NP y  $n=795$  pacientes que salen del sistema sin progresión

En la Tabla 2.15 se evalúa la función  $S(t)$  en algunos valores de tiempo que se han especificado. Los resultados corresponden a las respectivas probabilidades de que hasta ese instante el paciente permanezca sin salir del sistema. Notemos que esta función definida para los pacientes que salen del sistema sin progresión, tiene una interpretación distinta que la propuesta en el sistema anterior con estado absorbente  $P$ . En este caso es la probabilidad de que hasta ese instante (y esto es porque el modelo es Coxiano), un paciente se mantenga dentro del sistema de control de su enfermedad. En la última columna de la misma tabla se presenta los respectivos valores de la función razón de fallo  $h(t)$ , que en este caso es creciente. Es esperable que la razón de salida del sistema, para estos pacientes que no sufren progresión, vaya en aumento.

t (en días)	S(t)	h(t)
200	0.9498	0.0004
400	0.8625	0.0005
600	0.7724	0.0006
800	0.6863	0.0006
1000	0.6054	0.0006
1200	0.5303	0.0007
1400	0.4611	0.0007
1600	0.3981	0.0008
1800	0.3415	0.0008
2000	0.2911	0.0008
2200	0.2467	0.0008
2400	0.2080	0.0009
2600	0.1744	0.0009
2800	0.1457	0.0009
3000	0.1211	0.0009
3200	0.1003	0.0010
3400	0.0828	0.0010
3600	0.0681	0.0010
3800	0.0559	0.0010
4000	0.0457	0.0010
4200	0.0373	0.0010
4400	0.0304	0.0010
4600	0.0247	0.0010
4800	0.0200	0.0011
5000	0.0162	0.0011

Tabla 2.15. Estimación de la supervivencia y razón de fallo en distintos valores de tiempo, para el sistema con estado absorbente NP .

Es posible calcular los tiempos medios de permanencia en este sistema con estado absorbente  $NP$  y compararlos con los tiempos medios muestrales. La función  $m(t)$  de *tiempos medios de permanencia en el sistema*, en el intervalo  $[0, t]$ , se obtiene a partir de las componentes de la matriz  $N(t)$ , aplicando esperanzas condicionadas y a partir del vector de probabilidades iniciales estimado en el modelo. Dado que el vector de probabilidades de ingreso al sistema tiene como única componente no nula la tercera, la función  $\tau(t)$  se estima

de la siguiente forma:

$$m(t) = 0.0000 \cdot (n_{11}(t) + n_{12}(t) + n_{13}(t) + n_{14}(t)) + 0.0000 \cdot (n_{21}(t) + n_{22}(t) + n_{23}(t) + n_{24}(t)) + 1.0000 \cdot (n_{31}(t) + n_{32}(t) + n_{33}(t) + n_{34}(t)) + 0.0000 \cdot (n_{41}(t) + n_{42}(t) + n_{43}(t) + n_{44}(t))$$

En la Figura 2.62 se presenta el gráfico de la función  $m(t)$  en comparación con los valores de los tiempos medios muestrales calculados a partir de los  $n = 795$  casos de pacientes que salen del sistema sin progresión. Puede apreciarse el buen ajuste logrado por el modelo.

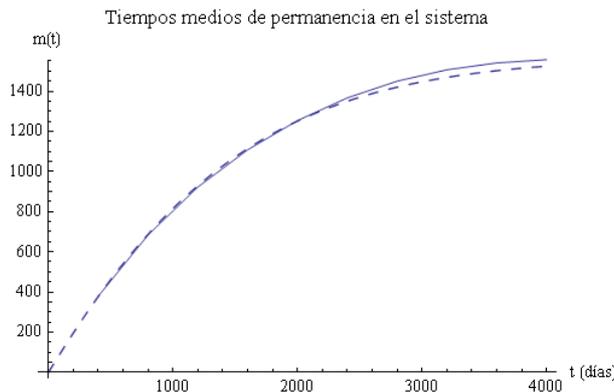


Figura 2.62. Tiempos medios de permanencia estimados (- -) en el sistema con estado absorbente NP, en comparación con los tiempos medios en la muestra de  $n=795$  pacientes que salen del sistema sin progresión

### Análisis de los tiempos en cada estado transitorio

En este caso no todos los ajustes en cada estado transitorio para un orden  $p = 2$  de la distribución tipo fase serán satisfactorios, lo que está vinculado a lo que se explicó anteriormente acerca de la heterogeneidad de valores del tiempo de permanencia en el sistema. En cada caso se buscará un ajuste satisfactorio incrementando  $p$  hasta el valor que sea necesario, que como se verá, no supera en ningún caso  $p = 4$ .

En virtud de lo observado hasta el momento, nos proponemos construir una distribución *tipo fase* con ajuste satisfactorio para los tiempos de permanencia en los estados transitorios

del modelo que representa la evolución de la enfermedad para los pacientes que salen del sistema sin progresión a partir de los datos.

Para cada estado transitorio 0, 1, 2 y 3, según se ha explicado en los objetivos de este trabajo, se construirá la distribución PH con un ajuste satisfactorio a la de cada variable aleatoria *Tiempo de permanencia en el estado*.

Igual que en el sistema global, se empleará el algoritmo EM de ajuste con distribuciones de este tipo (véase [32]).

#### ■ Ajuste para los tiempos de permanencia en el estado 0

A partir del algoritmo EM se ha realizado un ajuste con distribución tipo PH con  $p = 2$ , para las 795 observaciones correspondientes a los tiempos de permanencia en el estado inicial 0, antes de la primera intervención quirúrgica posterior ó antes de la salida del sistema. De los 795 pacientes registrados en la base de datos y a partir de la primera intervención quirúrgica para extirpar el tumor no invasivo de los músculos de la vejiga (en el instante 0):

- 499 salen del sistema sin progresión (estado NP).
- 296 realizan una primera consulta de control en la que se diagnostica recidiva sin progresión.

Se obtuvo las siguientes estimaciones del vector  $\alpha_0$ , la matriz  $A_0$  y la matriz columna de absorción  $A_0^0$ :

$$\widehat{\alpha}_0 = \begin{bmatrix} 1.0000 \\ 0.0000 \end{bmatrix}, \widehat{A}_0 = \begin{bmatrix} -0.0100 & 0.0100 \\ 0.0003 & -0.0014 \end{bmatrix} \text{ y } \widehat{A}_0^0 = \begin{bmatrix} 0.0000 \\ 0.0011 \end{bmatrix}$$

En las Figuras 2.63, 2.64, 2.65 y 2.66 se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los

gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 0* y la función de supervivencia, así como los de las funciones densidad y razón de fallo, todas ellas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

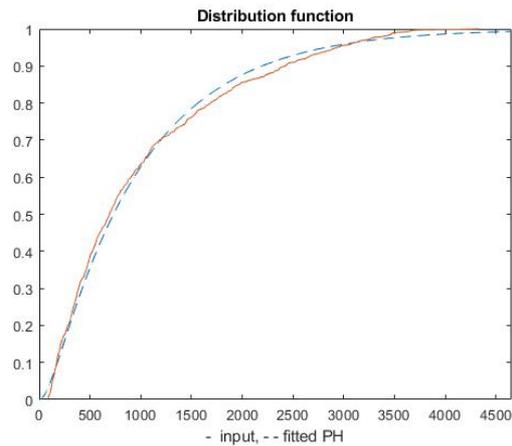


Figura 2.63. Gráfico de la función de distribución tipo fase estimada para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 795$  casos

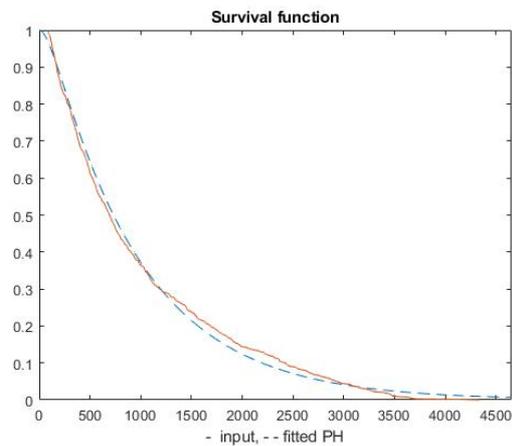


Figura 2.64. Gráfico de la función de supervivencia estimada para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 795$  casos.

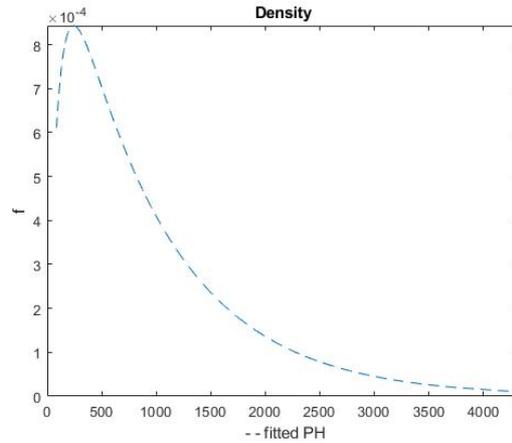


Figura 2.65. Gráfico de la función densidad tipo fase estimada para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 795$  casos

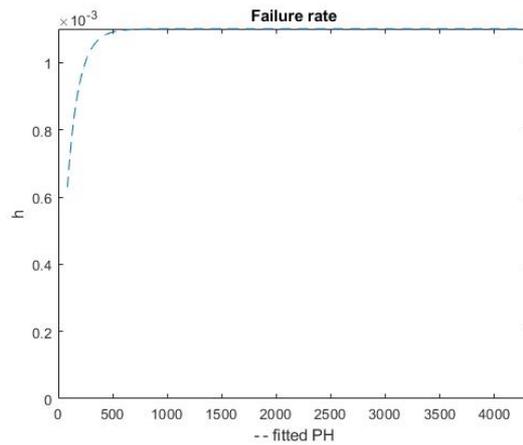


Figura 2.66. Gráfico de la función razón de fallo estimada para el *Tiempo de permanencia en 0*, a partir de la muestra de  $n = 795$  casos.

El valor del estadístico de Kolmogorov-Smirnov es  $D = 0.0330$  que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.0480$  y  $D_{\alpha=0.01} = 0.0575$ . Al no pertenecer el valor de  $D$  a la región crítica admitimos que el ajuste con esta distribución  $PH(\hat{\alpha}, \hat{T})$  es satisfactorio, para ambos niveles de significación.

- **Ajuste para los tiempos de permanencia en el estado 1**

A partir del algoritmo EM se ha realizado un ajuste con distribución tipo *PH* con  $p = 4$ , para las 296 observaciones correspondientes a los tiempos de permanencia en el estado 1 antes de la segunda intervención quirúrgica posterior a la inicial (estado 2) ó salida del sistema sin progresión de la enfermedad (hacia el estado *NP*). De las 296 observaciones:

- 168 transitan a 2 sin salir del sistema.
- 128 salen del sistema sin progresión de la enfermedad.

Se obtuvo las siguientes estimaciones del vector  $\alpha_1$ , de la matriz  $A_1$  y de la matriz columna de absorción  $A_1^0$ :

$$\widehat{\alpha}_1 = \begin{bmatrix} 0.0000 \\ 0.0000 \\ 1.0000 \\ 0.0000 \end{bmatrix}, \widehat{A}_1 = \begin{bmatrix} -0.0028 & 0.0002 & 0.0024 & 0.0001 \\ 0.0033 & -0.0106 & 0.0000 & 0.0073 \\ 0.0001 & 0.0106 & -0.0107 & 0.0000 \\ 0.0021 & 0.0000 & 0.0000 & -0.0118 \end{bmatrix} \text{ y}$$

$$\widehat{A}_1^0 = \begin{bmatrix} 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0097 \end{bmatrix}$$

En las Figuras 2.67, 2.68, 2.69 y 2.70 se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 1* y la función de supervivencia, así como los de las funciones densidad y razón de fallo, todas ellas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

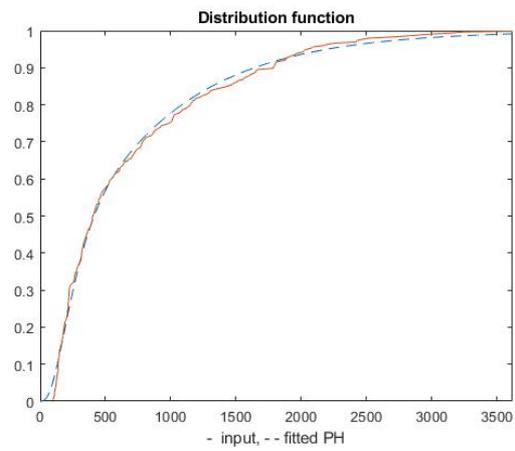


Figura 2.67. Gráfico de la función de distribución tipo fase estimada para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 296$  casos

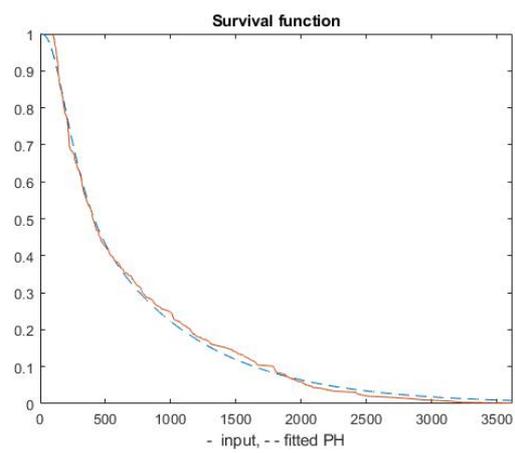


Figura 2.68. Gráfico de la función de supervivencia estimada para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 296$  casos.

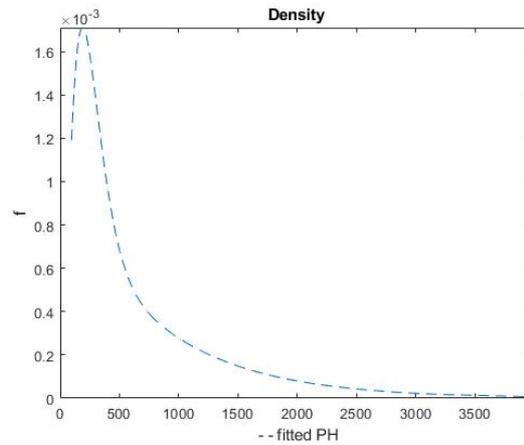


Figura 2.69. Gráfico de la función densidad tipo fase estimada para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 296$  casos

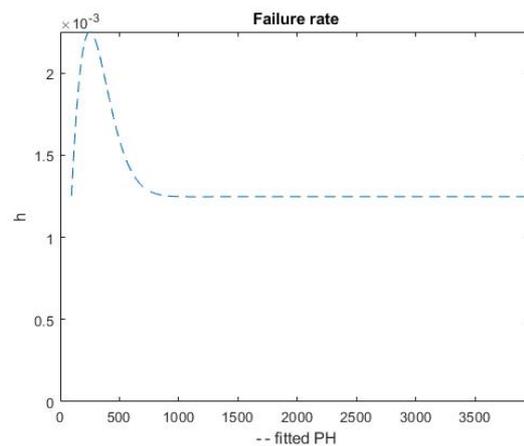


Figura 2.70. Gráfico de la función razón de fallo estimada para el *Tiempo de permanencia en 1*, a partir de la muestra de  $n = 296$  casos.

El valor del estadístico de Kolmogorov-Smirnov es  $D = 0.0562$  que resulta menor a  $D_{\alpha=0.05} = 0.0784$  y  $D_{\alpha=0.01} = 0.0939$ . Al no pertenecer el valor de  $D$  a la región crítica en cada caso es posible afirmar que el ajuste para  $p = 4$  ha resultado ser satisfactorio.

■ **Ajuste para los tiempos de permanencia en el estado 2**

A partir del algoritmo EM se ha realizado un ajuste con distribución tipo *PH* con  $p = 3$ , para las 168 observaciones correspondientes a los tiempos de permanencia en el estado 2 antes de la tercera intervención quirúrgica por recidiva posterior a la inicial ó antes de la salida del sistema sin progresión de la enfermedad. De los 168 casos:

- 45 salen del sistema sin progresión (estado *NP*) desde el estado 2.
- 123 transitan al estado 3 por presentar nueva recidiva sin progresión.

Se obtuvo las siguientes estimaciones del vector  $\alpha_2$ , la matriz  $A_2$  y la matriz columna de absorción  $A_2^0$ :

$$\widehat{\alpha}_2 = \begin{bmatrix} 0.0000 \\ 0.0118 \\ 0.9882 \end{bmatrix}, \widehat{A}_2 = \begin{bmatrix} -0.0061 & 0.0016 & 0.0000 \\ 0.0001 & -0.0021 & 0.0020 \\ 0.0060 & 0.0001 & -0.0061 \end{bmatrix} \text{ y } \widehat{A}_2^0 = \begin{bmatrix} 0.0044 \\ 0.0000 \\ 0.0000 \end{bmatrix}$$

En las Figuras 2.71, 2.72, 2.73 y 2.74 se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 2* y la función de supervivencia, así como los de las funciones densidad y razón de fallo, todas ellas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

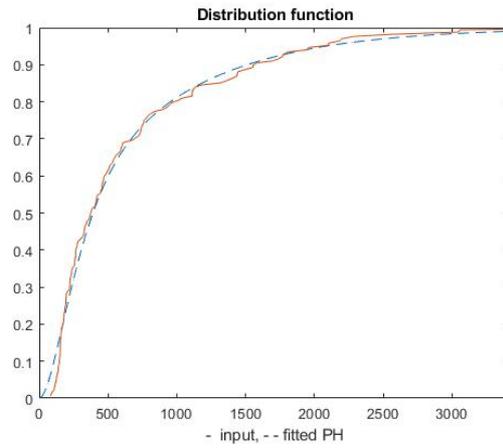


Figura 2.71. Gráfico de la función de distribución tipo fase estimada para el *Tiempo de permanencia en 2*, a partir de la muestra de  $n = 168$  casos

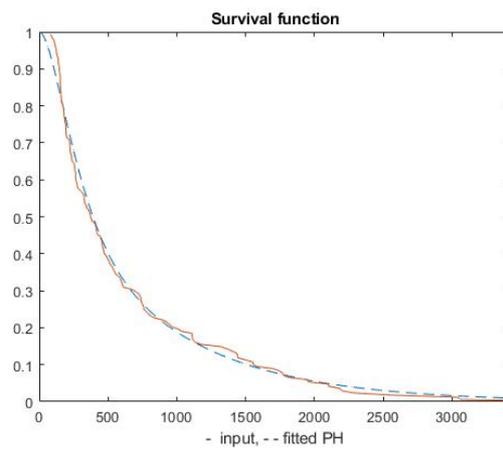


Figura 2.72. Gráfico de la función de supervivencia estimada para el *Tiempo de permanencia en 2*, a partir de la muestra de  $n = 168$  casos.

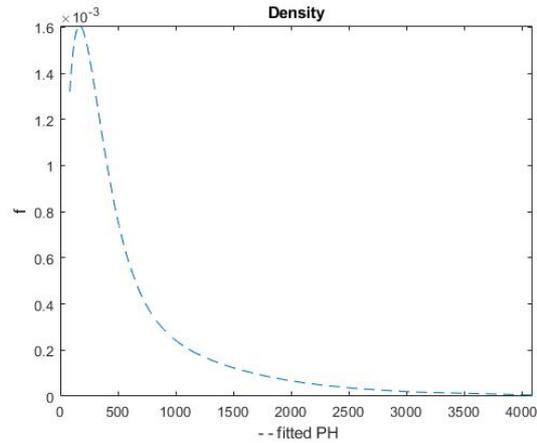


Figura 2.73. Gráfico de la función densidad tipo fase estimada para el *Tiempo de permanencia en 2*, a partir de la muestra de  $n = 168$  casos

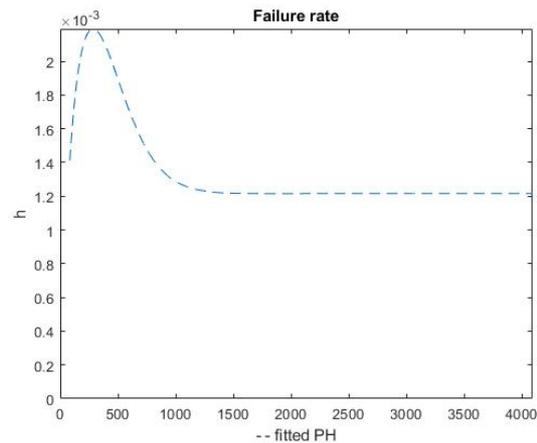


Figura 2.74. Gráfico de la función razón de fallo estimada para el *Tiempo de permanencia en 2*, a partir de la muestra de  $n = 168$  casos.

El valor del estadístico de Kolmogorov-Smirnov es  $D = 0.0854$  que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.1037$  y  $D_{\alpha=0.01} = 0.1244$ . De lo anterior se deduce que en este caso el ajuste puntual para el estado 2 y con  $p = 3$  es satisfactorio.

- **Ajuste para los tiempos de permanencia en el estado 3**

A partir del algoritmo EM se ha realizado un ajuste con distribución tipo  $PH$  con  $p = 2$  para las 123 observaciones correspondientes a los tiempos de permanencia en el estado 3 antes de la salida del sistema sin progresión de la enfermedad (estado  $NP$ ). Se obtuvo las siguientes estimaciones del vector  $\alpha_3$ , de la matriz  $A_3$  y de la matriz columna de absorción  $A_3^0$ :

$$\widehat{\alpha}_3 = \begin{bmatrix} 0.7813 \\ 0.2187 \end{bmatrix}, \widehat{A}_3 = \begin{bmatrix} -0.0016 & 0.0016 \\ 0.0000 & -0.0016 \end{bmatrix} \text{ y } \widehat{A}_3^0 = \begin{bmatrix} 0.0000 \\ 0.0016 \end{bmatrix}$$

En las Figuras 2.75, 2.76, 2.77 y 2.78 se presentan las salidas que ofrece Matlab a partir de la ejecución del algoritmo modificado, conteniendo respectivamente los gráficos de la función de distribución tipo fase que se ha obtenido a partir del ajuste para la variable *Tiempo de permanencia en 3* y la función de supervivencia, así como los de las funciones densidad y razón de fallo, todas ellas en contraste con los valores obtenidos a partir del diseño muestral. Nótese la bondad de dicho ajuste.

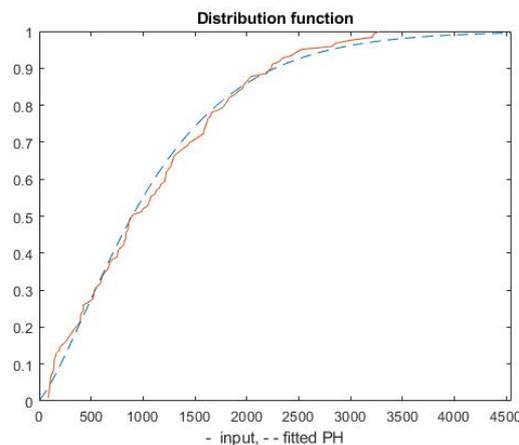


Figura 2.75. Gráfico de la función de distribución tipo fase estimada para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 123$  casos

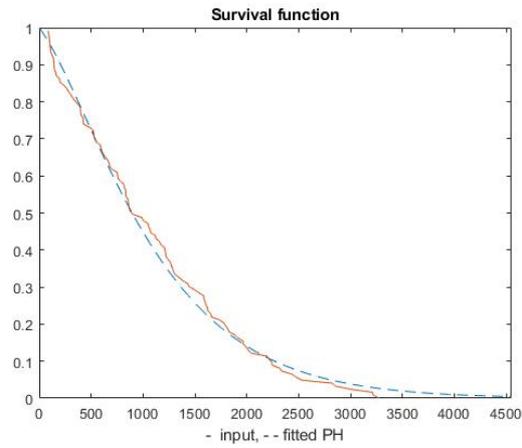


Figura 2.76. Gráfico de la función de supervivencia estimada para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 123$  casos.

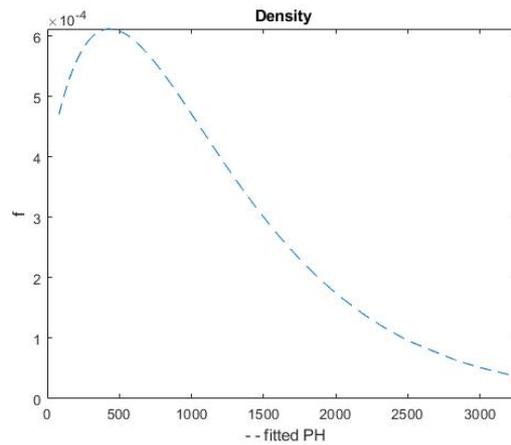


Figura 2.77. Gráfico de la función densidad tipo fase estimada para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 123$  casos

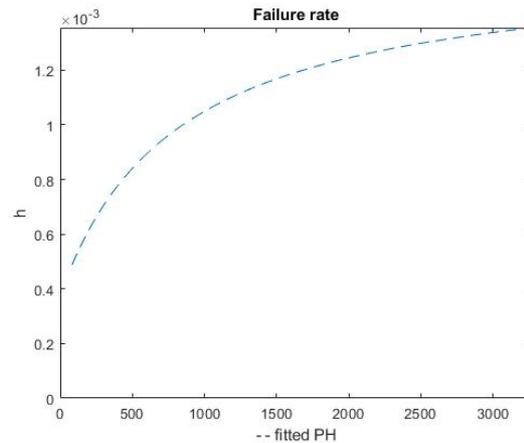


Figura 2.78. Gráfico de la función razón de fallo estimada para el *Tiempo de permanencia en 3*, a partir de la muestra de  $n = 123$  casos.

El valor del estadístico de Kolmogorov-Smirnov es  $D = 0.0555$  que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.1210$  y  $D_{\alpha=0.01} = 0.1451$ . De lo antes dicho, el ajuste para este estado es satisfactorio, habiendo elegido por ser suficiente el orden  $p = 2$  para la distribución tipo fase.

En la Tabla 2.16 se evalúan las funciones de supervivencia en los estados transitorios, en distintos valores de tiempo  $t$ . Nótese que la interpretación de esta función que recibe este nombre en la teoría de fiabilidad, tiene una interpretación distinta a la de las funciones obtenidas en el sistema parcial anterior, con estado absorbente  $P$ . En este caso, el valor de la función supervivencia en un instante  $t$  hace referencia a la probabilidad de que en ese instante de tiempo transcurrido mientras el paciente está en ese estado transitorio, no haya salido del sistema (sin progresión).

t(en días)	$S_0(t)$	$S_1(t)$	$S_2(t)$	$S_3(t)$
200	0.8828	0.7859	0.7546	0.9038
600	0.5778	0.3714	0.3356	0.6640
1000	0.3721	0.2222	0.1881	0.4489
1400	0.2396	0.1349	0.1145	0.2888
1800	0.1542	0.0819	0.0704	0.1798
2200	0.0993	0.0498	0.0433	0.1093
2600	0.0639	0.0302	0.0266	0.0653
3000	0.0411	0.0183	0.0164	0.0384
3400	0.0265	0.0111	0.0101	0.0224
3800	0.0171	0.0068	0.0062	0.0129
4200	0.0110	0.0041	0.0038	0.0074
4600	0.0071	0.0025	0.0023	0.0042
5000	0.0045	0.0015	0.0014	0.0024

Tabla 2.16. Evaluación de las funciones de *supervivencia* en los estados transitorios del sistema con estado absorbente NP

En la Tabla 2.17 se evalúan las funciones razones de fallo en los estados transitorios, en distintos valores de tiempo  $t$ .

t(en días)	$h_0(t)$	$h_1(t)$	$h_2(t)$	$h_3(t)$
200	0.0009	0.0022	0.0021	0.0006
600	0.0011	0.0014	0.0017	0.0009
1000	0.0011	0.0012	0.0013	0.0010
1400	0.0011	0.0012	0.0012	0.0011
1800	0.0011	0.0012	0.0012	0.0012
2200	0.0011	0.0012	0.0012	0.0013
2600	0.0011	0.0012	0.0012	0.0013
3000	0.0011	0.0012	0.0012	0.0013
3400	0.0011	0.0012	0.0012	0.0014
3800	0.0011	0.0012	0.0012	0.0014
4200	0.0011	0.0012	0.0012	0.0014
4600	0.0011	0.0012	0.0012	0.0014
5000	0.0011	0.0012	0.0012	0.0014

Tabla 2.17. Evaluación de las funciones *razón de fallo* en los estados transitorios del sistema con estado absorbente NP

### Generador del Sistema y cálculo de las medidas de interés

La matriz generadora  $Q$ , cuadrada y de orden 12, tendrá dos bloques principales  $A$  y  $B$ . Los bloques de  $A$  que definen las once primeras filas y columnas, caracterizan las transiciones entre los cuatro estados transitorios. Por otro lado la última columna que conforma  $B$ , contiene las razones de absorción por parte del estado absorbente  $NP$ . La matriz  $Q$  se completa con bloques de matrices nulas de dimensiones correspondientes a cada caso. Cada una de las matrices que componen los bloques de  $Q$  surgen de una estimación realizada a partir de la muestra de observaciones y las notaciones correspondientes se han simplificado:

$$Q = \begin{bmatrix} A_{11 \times 11} & B_{11 \times 1} \\ 0_{1 \times 11} & 0_{1 \times 1} \end{bmatrix}$$

$$A = \begin{bmatrix} A_0 & A_0^0 p_{01} \cdot \alpha_1 & 0 & 0 \\ 0 & A_1 & A_1^0 p_{12} \alpha_2 & 0 \\ 0 & 0 & A_2 & A_2^0 p_{23} \alpha_3 \\ 0 & 0 & 0 & A_3 \end{bmatrix}$$

$$B = \begin{bmatrix} p_{0NP} A_0^0 \\ p_{1NP} A_1^0 \\ p_{2NP} A_2^0 \\ p_{3NP} A_3^0 \end{bmatrix}$$

Teniendo en cuenta que las probabilidades de transición deben sumar 1, se tiene:

$$p_{i(i+1)} + p_{iNP} = 1, \forall i \in \{0, 1, 2\}$$

Estas probabilidades se han estimado a partir de las proporciones muestrales:

$$p_{01} = 296/795 \approx 0.3723 \quad p_{0NP} = 499/795 \approx 0.6277$$

$$p_{12} = 168/296 \approx 0.5676 \quad p_{1NP} = 128/296 \approx 0.4324$$

$$p_{23} = 123/168 \approx 0.7321 \quad p_{2NP} = 45/168 \approx 0.2679$$

$$p_{3NP} = 123/123 = 1.0000$$

El vector  $\alpha$  para el sistema resulta de la extensión con componentes nulas del vector de la estimación realizada en el estado de ingreso al sistema  $\alpha_0$ :  $\alpha = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$

Las matrices  $A$  y  $B$  que componen los dos bloques principales de  $Q$ , son respectivamente:

$$A = \begin{bmatrix} -0.0100 & 0.0100 & | & 0 & 0 & 0 & 0 & | & 0 & 0 & 0 & | & 0 & 0 \\ 0.0003 & -0.0014 & | & 0 & 0 & 0.0004 & 0 & | & 0 & 0 & 0 & | & 0 & 0 \\ --- & --- & | & --- & --- & --- & --- & | & --- & --- & --- & | & --- & --- \\ 0 & 0 & | & -0.0028 & 0.0002 & 0.0024 & 0.0001 & | & 0 & 0.0000 & 0.0000 & | & 0 & 0 \\ 0 & 0 & | & 0.0033 & -0.0106 & 0.0000 & 0.0073 & | & 0 & 0 & 0 & | & 0 & 0 \\ 0 & 0 & | & 0.0001 & 0.0106 & -0.0107 & 0.0000 & | & 0 & 0 & 0 & | & 0 & 0 \\ 0 & 0 & | & 0.0021 & 0.0000 & 0.0000 & -0.0118 & | & 0 & 0.0001 & 0.0054 & | & 0 & 0 \\ --- & --- & | & --- & --- & --- & --- & | & --- & --- & --- & | & --- & --- \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & -0.0061 & 0.0016 & 0.0000 & | & 0.0025 & 0.0007 \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0.0001 & -0.0021 & 0.0020 & | & 0 & 0 \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0.0060 & 0.0001 & -0.0061 & | & 0 & 0 \\ --- & --- & | & --- & --- & --- & --- & | & --- & --- & --- & | & --- & --- \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0 & 0 & 0 & | & -0.0016 & 0.0016 \\ 0 & 0 & | & 0 & 0 & 0 & 0 & | & 0 & 0 & 0 & | & 0.0000 & -0.0016 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0000 \\ 0.0007 \\ 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0042 \\ 0.0012 \\ 0.0000 \\ 0.0000 \\ 0.0000 \\ 0.0016 \end{bmatrix}$$

Luego, a partir de cálculo simbólico, es posible obtener los dos bloques principales de la matriz  $P(t)$ ,  $A(t)$  y  $B(t)$ , que se han propuesto en el marco teórico. La matriz  $A(t)$  en cada fila contiene las probabilidades de ocupación de la fase virtual en el estado que indica la columna correspondiente en el instante  $t$ , habiendo ingresado al sistema en la fase del estado que indica la fila. La matriz  $B(t)$  en su elemento  $b_{i1}(t)$ , contiene la probabilidad de

ocupación del estado absorbente  $NP$  en el instante  $t$ , habiendo ingresado al sistema (cuando  $t = 0$ ) en la fase del estado que indica  $i$ . Luego, la probabilidad de estar en el sistema en el instante  $t$ , viene dada por la expresión siguiente que coincide con la definición de la función *supervivencia* en el caso de un estado absorbente:

$$S(t) = 1 - \alpha \cdot B(t)$$

En la Figura 2.79 puede apreciarse la similaridad de las dos funciones de supervivencia que se han estimado en este modelo. Una de ellas es la que se obtuvo a partir de los tiempos de absorción por parte del estado  $NP$  y a partir de un generador de dimensión  $5 \times 5$ , sin tener en cuenta los tiempos en los estados transitorios. La segunda función de supervivencia se ha obtenido teniendo en cuenta dichos estados y a partir de la construcción del generador  $Q$  de orden  $12 \times 12$  (- -). Cabe esperar dicha semejanza, pero esta última construcción permite obtener mayor información de las distintas etapas de la enfermedad.

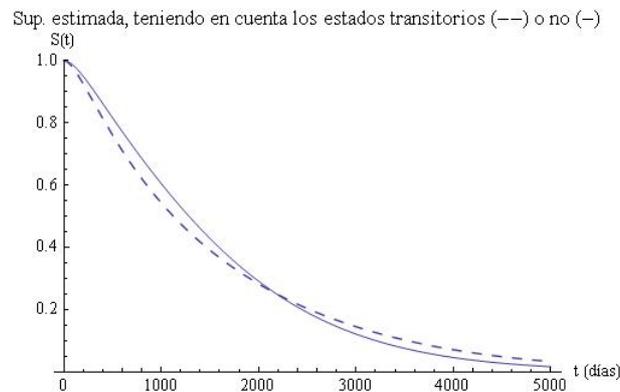


Figura 2.79. Gráficos de las funciones de supervivencia estimadas: sin tener en cuenta los tiempos transitorios y a partir de considerarlos (- -).

En la Figura 2.80 se ha representado la función de supervivencia estimada en el sistema con estado absorbente  $NP$ , en contraste con la función empírica de supervivencia.

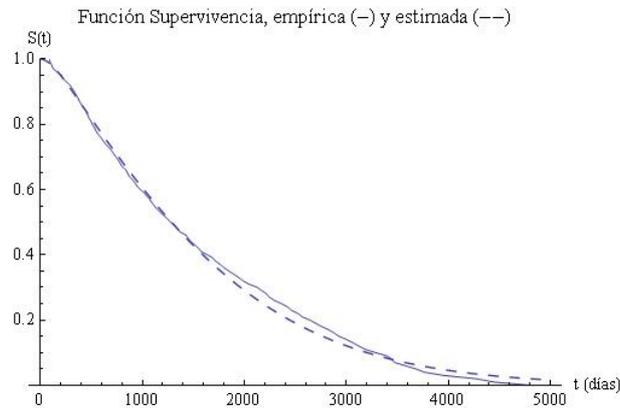


Figura 2.80. Gráfico de la función de supervivencia estimada (---) en el sistema con estado absorbente NP, en contraste con la función empírica de supervivencia.

Se resume en la Tabla 2.18 las características de los ajustes logrados con distribuciones *tipo fase*, para los tiempos de permanencia en cada estado transitorio:  $n$  el tamaño muestral, los valores  $D$  de los estadísticos de Kolmogorov Smirnov y los límites de las regiones críticas  $D_\alpha$ , evidenciando en cada caso la bondad de los mismos.

Estados	p	n	D	$D_{0.10}$	$D_{0.05}$	$D_{0.01}$
0	2	795	0,0330	0,0432	0,0480	0,0575
1	4	296	0,0562	0,0706	0,0784	0,0939
2	3	168	0,0854	0,0935	0,1037	0,1244
3	2	123	0,0555	0,1091	0,1210	0,1451
Tiempos en el sistema		795	0.0415	0,0432	0,0480	0,575

Tabla 2.18. Análisis de la bondad de los ajustes, en el sistema con estado absorbente NP.

La siguiente matriz determina en cada uno de sus elementos los valores esperados para los tiempos de permanencia en  $j$  habiendo partido del estado  $i$  en el instante  $t = 0$ , en el intervalo  $(0, t)$ :

$$\int_0^t P(\tau) d\tau = \begin{bmatrix} N(t) & M(t) \\ 0 & tI \end{bmatrix}$$

Como ya se ha explicado con anterioridad, es posible calcular los tiempos medios de permanencia en cada estado transitorio de este sistema (con estado absorbente  $NP$ ) y compararlos con los tiempos medios muestrales. La función  $\tau_{0j}(t)$  de *tiempos medios de permanencia en cada estado transitorio  $j$*  habiendo partido de 0 en el instante  $t = 0$  en el intervalo  $[0, t]$ , se obtiene a partir de las componentes de la matriz  $N(t)$  aplicando esperanzas condicionadas y a partir del vector de probabilidades iniciales estimado en el modelo. Nótese que se utilizarán únicamente las dos primeras filas de dicha matriz, teniendo en cuenta que cada paciente ingresa al sistema en el estado 0. Por otro lado dada la estimación del vector de probabilidades  $\alpha_0$ , en particular, se usará únicamente la primera fila de  $N(t)$ , cuya dimensión es  $11 \times 11$ :

$$m_0(t) = 1.0000(n_{1,1}(t) + n_{1,2}(t))$$

$$m_1(t) = 1.0000(n_{1,3}(t) + n_{1,4}(t) + n_{1,5}(t) + n_{1,6}(t))$$

$$m_2(t) = 1.0000(n_{1,7}(t) + n_{1,8}(t) + n_{1,9}(t))$$

$$m_3(t) = 1.0000(n_{1,10}(t) + n_{1,11}(t))$$

El ajuste entre la función  $m_j$  y los valores de tiempos medios muestrales ha sido muy bueno cuando  $j = 0$ , lo que puede apreciarse en la Figura 2.81. A pesar de que en cada estado transitorio se cuenta con un tamaño muestral adecuado y de que los ajustes de las distribuciones en cada caso son muy buenos, no habrá similitud entre los tiempos medios de permanencia en los estados 1, 2 y 3 y los tiempos medios muestrales correspondientes. Debemos tener en cuenta que los promedios muestrales se verán notablemente sesgados por

valores extremos elevados como los de salida del sistema sin progresión, aspecto que ya se había mencionado con anterioridad desde el análisis estadístico descriptivo de las variables. Los tiempos medios teóricos calculados a partir del generador, presuponen la homogeneidad de la variable respuesta analizada dentro del estrato estudiado y como se ha reseñado, esto no ocurre con nuestras observaciones.

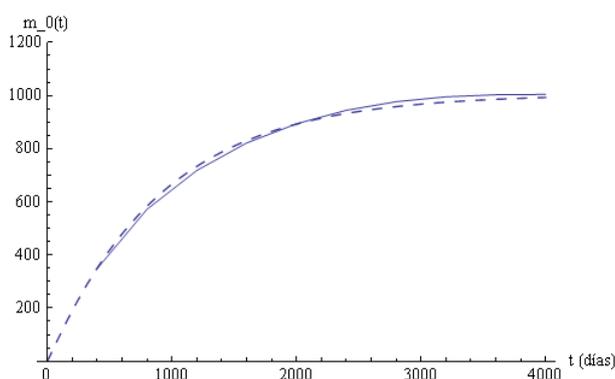


Figura 2.81. Tiempos esperados (- -) y tiempos medios muestrales, de permanencia en el estado 0 en el intervalo  $[0, t]$ , habiendo partido en  $t = 0$  del estado 0. Muestra de  $n=795$  pacientes que salen del sistema sin progresión

En la Tabla 2.19 se presentan los tiempos medios muestrales de permanencia en los distintos estados transitorios, en distintos intervalos  $[0, t]$ .

$t$ (en días)	$m_0(t)$	$m_1(t)$	$m_2(t)$	$m_3(t)$	$m_{Sist,emp}(t)$	$m_{Sist,est}(t)$
400	349.06	71.57	15.32	0.00	378.94	377.66
800	571.91	216.86	106.83	57.91	684.18	686.87
1200	718.03	346.49	206.21	206.79	922.62	929.36
1600	821.66	441.52	306.47	372.79	1108.49	1114.14
2000	893.76	520.26	386.81	546.50	1253.76	1251.08
2400	944.57	584.90	459.99	693.15	1366.79	1350.07
2800	976.94	626.73	523.55	832.89	1449.79	1420.11
3200	995.36	654.50	574.86	942.97	1506.42	1468.77
3600	1002.80	636.42	602.57	1028.37	1540.71	1502.05
4000	1004.23	686.68	622.01	1070.98	1557.05	1524.52

Tabla 2.19. Tiempos medios de permanencia en cada estado y en el sistema con estado absorbente NP, hasta cierto instante  $t$ .

Si se analizan las dos últimas columnas de la tabla anterior, es claro que tiempos medios muestrales y estimados a partir del modelo que hemos creado se ajustan muy bien. Para visualizarlo más claramente, hemos considerado las variables  $X = m_{Sist,emp}(t)$  e  $Y = m_{Sist,est}(t)$ , realizando un análisis de la correlación entre ambas y planteado un modelo lineal que ha presentado excelente ajuste:  $Y = aX + b + \varepsilon$ , con un valor para  $R_a^2 \cong 0.9999$ . El mejor ajuste lineal, se ha obtenido para  $\hat{a} \cong 0.99$  y  $\hat{b} \cong 0$ , no habiendo rechazado la hipótesis nula en el contraste:  $H_0 : b = 0$ ,  $H_1 : b \neq 0$ . En el contraste  $H_0 : a = 0$ ,  $H_1 : a \neq 0$  se ha rechazado la hipótesis nula a partir de un  $p$ -valor igual a  $2 \times 10^{-16}$ , claramente significativo.

## 2.6. Costes medios en función del tiempo

A partir de los tiempos medios de permanencia en cada estado transitorio, ya sea en el sistema que representa el proceso de la enfermedad para los pacientes que derivan en progresión ó en el sistema que representa el proceso para los pacientes que salen sin progresión, es posible calcular los costes medios en función del tiempo (expresado en días) teniendo en cuenta los costes directos e indirectos implicados en tal caso:

- Se entiende por coste directo, el consumo de recursos del sistema de atención sanitaria, incluyendo en ellos los correspondientes a la atención hospitalaria, tanto en régimen de hospitalización como de atención ambulatoria en consultas externas; los de la atención primaria y de los fármacos aspecto que puede ampliarse en [27] (López Bastida, J., et. al., 2002).
- Los costes indirectos consisten en la disminución o pérdida de productividad debida a la mortalidad prematura y a la incapacidad laboral atribuible a la enfermedad.

Sean los parámetros cuyos valores pueden ser estimados a partir de estudios estadísticos previos, en determinada población:

- Coste directo por unidad de tiempo (en días) por permanecer en el estado transitorio  $j$ :

$$C_{d,j}, \text{ con } j \in \{0, 1, 2, 3\}.$$

- Coste indirecto por unidad de tiempo (en días) por permanecer en el estado transitorio  $j$ :

$$C_{i,j}, \text{ con } j \in \{0, 1, 2, 3\}.$$

- El tiempo medio (en días) de permanencia en el estado  $j$ , hasta el instante  $t$ :

$$m_j(t), \text{ con } j \in \{0, 1, 2, 3\}.$$

A su vez, si se tiene en cuenta que el coste directo por unidad de tiempo en cada estado tiene una componente fija que corresponde al coste por el seguimiento en la etapa inicial que incluye los gastos por el diagnóstico y tratamiento inicial, además de un coste variable por unidad de tiempo, es posible descomponerlo de la siguiente forma:

$$C_{d,j}(t) = CF_{d,j} + CV_{d,j} \cdot m_j(t), \forall j \in \{0, 1, 2, 3\}$$

Luego se tiene que el *coste medio* en función de los tiempos medios de permanencia en cada uno de los estados transitorios para esta enfermedad, viene dado por:

$$CM(t) = \sum_{j=0}^{j=3} \{ [CF_{d,j} + CV_{d,j} \cdot m_j(t)] + C_{i,j} \cdot m_j(t) \}$$

La estimación específica de los costos exige conocer costes específicos en determinado instante de las distintas componentes de cada tipo de coste. Esto requiere de un estudio más profundo, sin embargo, se deja el planteo para futuras líneas de investigación en este sentido.

---

## **La supervivencia a la enfermedad en distintos grupos de riesgo**

---

En este capítulo presentamos una generalización del modelo inicial a partir de una base de datos cuyas características principales han sido reseñadas en la metodología (Capítulo 1) y que a pesar de su menor tamaño cuenta con información adicional sobre distintas variables de interés en relación al diagnóstico inicial y al proceso de la enfermedad de los pacientes. Con el objetivo de discutir los efectos de las covariables en el generador del modelo y, por lo tanto, en las principales medidas de supervivencia, fue necesario adecuar el MAM. Según se ha constatado en la literatura relacionada a este tema, estos efectos suelen estudiarse en forma aislada dentro de los distintos estratos, lo que limita la comparación posterior entre los distintos grupos de riesgo.

### **3.1. El modelo espacio - estados**

El sistema multiestado que representa las distintas etapas de la enfermedad para esta base de datos, se presenta en la Figura 3.1 y muestra todas las transiciones que han realizado

los  $n = 122$  pacientes que se han observado durante un intervalo de tiempo, con los correspondientes recuentos.

El estado 0 representa el ingreso al sistema de control y seguimiento de la enfermedad. En el instante  $t = 0$  se fija el comienzo de dicho seguimiento individual y en todos los casos corresponde a la primera intervención quirúrgica del paciente. Los estados 1 y 2 corresponden a la primera y segunda recidiva local (RL) sin metástasis. Un paciente ingresa al estado 3 si presenta metástasis pero sin RL e ingresa a 4 si presenta ambas cosas. Los estados  $S$  y  $M$  representan la salida del sistema con vida y muerte respectivamente.

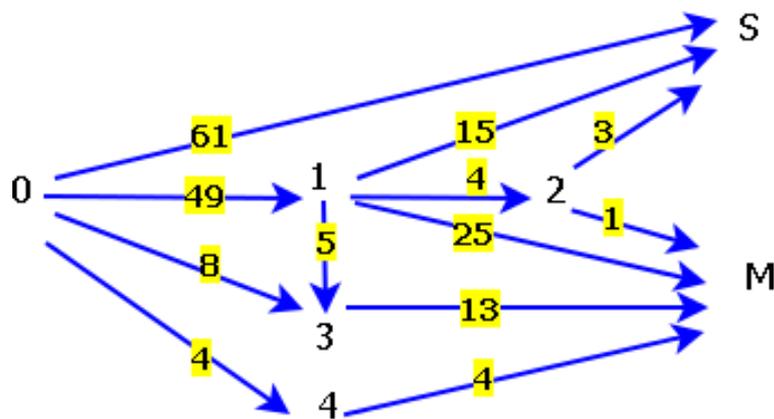


Figura 3.1. Sistema multi-estado que representa las distintas etapas de la enfermedad de los  $n = 122$  pacientes.

En un primer nivel de simplificación, con el objetivo de obtener los recuentos mínimos adecuados para hacer inferencias poblacionales y en relación a las distintas transiciones entre estados, podemos plantearnos el siguiente diagrama multi-estados que se observa en la Figura 3.2, donde el estado 1 representa una o más RL sin metástasis y el estado 2 representa la constatación de metástasis, con o sin RL, simplificando el sistema original de forma considerable aunque sin perder de vista las características esenciales de las transiciones de cada paciente:

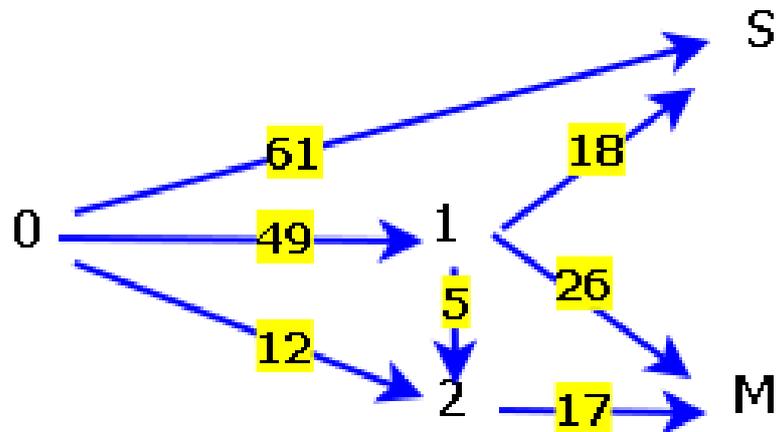


Figura 3.2. Primera simplificación del sistema multi-estado que representa las distintas etapas de la enfermedad de los  $n = 122$  pacientes.

Para el abordaje del problema ha sido necesario realizar una nueva simplificación del modelo anterior, con el objetivo de tener suficientes recuentos para todos los tipos de transiciones y teniendo en cuenta la gran cantidad de observaciones censuradas en esta base de datos. Puede verse dicho modelo en la Figura 3.3, donde en este caso el estado 1 representa una o más recidivas con o sin metástasis,  $M$  es el estado absorbente que representa el fallecimiento del paciente y  $V$  representa la finalización del seguimiento médico del paciente, quien hasta dicho instante ha sobrevivido a la enfermedad. Nótese que en este último caso lo que concluye el tiempo de observación es la finalización del período de estudio y que desconocemos cómo será el proceso posterior de la enfermedad. Es por eso que dichos registros de tiempo se han considerado como datos con censura aleatoria y por derecha.

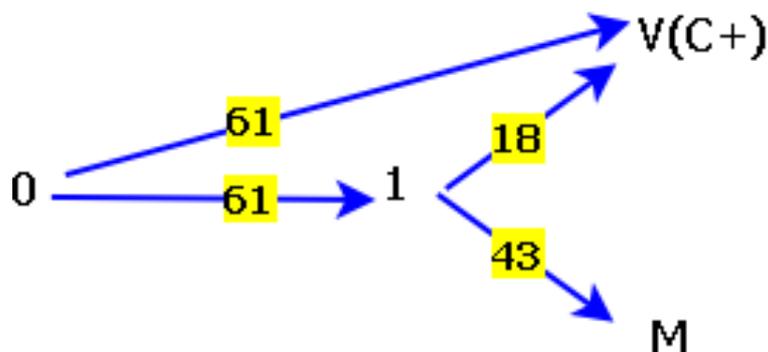


Figura 3.3. Diagrama del sistema multi-estado simplificado, para los  $n = 122$  pacientes.

### 3.2. Determinación de los distintos grupos de riesgo

En el gráfico de la Figura 3.4 se aprecia que hay aproximadamente un 70% de personas que superan los 60 años, la tabla de la Figura 3.5 permite destacar que más del 90% de las observaciones corresponden a pacientes hombres. Los recuentos de la tabla de la Figura 3.6 muestran un seguimiento de pacientes con más de un 30% de casos en los dos niveles superiores de riesgo de la variable *Estadío*. A su vez, el cruzamiento de las variables *Situación* y *Estadío* genera los recuentos de la Figura 3.7 con valores no nulos en todas las celdas. En virtud de lo anterior será posible realizar un estudio de la enfermedad y de las principales medidas de interés, diferenciado por subgrupos de distinto riesgo en relación al diagnóstico inicial. Los recuentos de dicha tabla ponen de manifiesto que el grupo de pacientes con menor riesgo de absorción por parte del estado M, según el diagnóstico inicial en relación a la variable *Estadío*, es el de los que se sitúan en sus dos primeros niveles T2 o T2-3 y lo constituyen 82 pacientes. Este análisis pretende ser un ejemplo de cómo es posible abordar en forma más completa el estudio de cualquier enfermedad a partir del estudio de los efectos de covariables que se hayan identificado como relevantes en el contexto del

problema, en particular si se cuenta con suficiente información a partir de bases de datos más detalladas y/o con mayor número de casos observados.

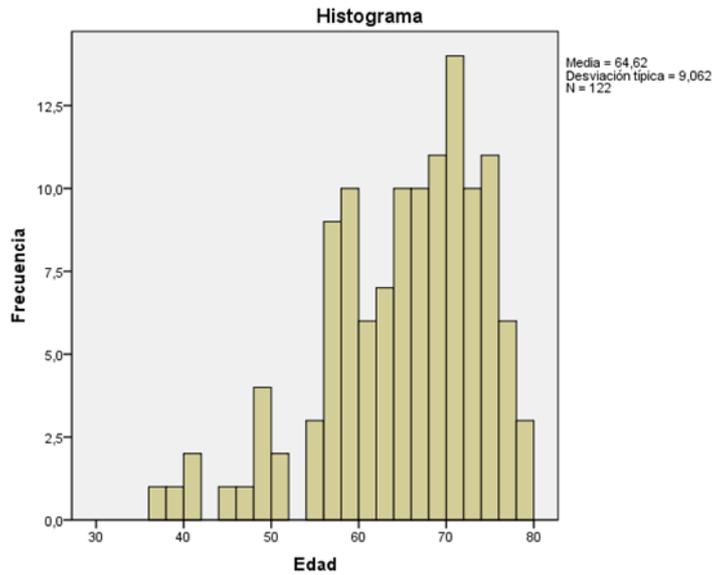


Figura 3.4. Histograma de frecuencias absolutas para la variable *Edad*.

		Sexo			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Varón	110	90,2	90,2	90,2
	Mujer	12	9,8	9,8	100,0
	Total	122	100,0	100,0	

Figura 3.5. Tabla de frecuencias para la variable *Sexo*.

		Estadio			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	T2	56	45,9	45,9	45,9
	T2_3	26	21,3	21,3	67,2
	T3	32	26,2	26,2	93,4
	T4	8	6,6	6,6	100,0
	Total	122	100,0	100,0	

Figura 3.6. Tabla de frecuencias para la variable *Estadio*.

		Estadio				Total
		T2	T2_3	T3	T4	
Situación	Vivo	47	16	14	2	79
	Muerto	9	10	18	6	43
Total		56	26	32	8	122

Figura 3.7. Tabla de contingencia para las variables *Situación* y *Estadio*.

El análisis paramétrico y no paramétrico de correlación entre *Situación* (V o M) y *Estadio*, como puede apreciarse ya a partir de los recuentos de la última tabla, es significativo rechazándose que el coeficiente de correlación valga 0 (ver Figura 3.8). También se ha estudiado la correlación entre *Situación* y *Edad* resultando no significativa, como puede verse en la Figura 3.9.

Correlaciones			
		Situación	Estadio
Situación	Correlación de Pearson	1	,413**
	Sig. (bilateral)		,000
	N	122	122
Estadio	Correlación de Pearson	,413**	1
	Sig. (bilateral)	,000	
	N	122	122

\*\* La correlación es significativa al nivel 0,01 (bilateral).

Correlaciones				
			Situación	Estadio
Tau_b de Kendall	Situación	Coefficiente de correlación	1,000	,384**
		Sig. (bilateral)		,000
		N	122	122
	Estadio	Coefficiente de correlación	,384**	1,000
		Sig. (bilateral)	,000	
		N	122	122
Rho de Spearman	Situación	Coefficiente de correlación	1,000	,412**
		Sig. (bilateral)		,000
		N	122	122
	Estadio	Coefficiente de correlación	,412**	1,000
		Sig. (bilateral)	,000	
		N	122	122

\*\* La correlación es significativa al nivel 0,01 (bilateral).

Figura 3.8. Análisis de correlación para las variables *Situación* y *Estadio*.

Correlaciones				
		Edad	Situación	
Edad	Correlación de Pearson	1	,141	
	Sig. (bilateral)		,121	
	N	122	122	
Situación	Correlación de Pearson	,141	1	
	Sig. (bilateral)	,121		
	N	122	122	

Correlaciones				
			Edad	Situación
Tau_b de Kendall	Edad	Coefficiente de correlación	1,000	,115
		Sig. (bilateral)		,130
		N	122	122
	Situación	Coefficiente de correlación	,115	1,000
		Sig. (bilateral)	,130	
		N	122	122
Rho de Spearman	Edad	Coefficiente de correlación	1,000	,138
		Sig. (bilateral)		,130
		N	122	122
	Situación	Coefficiente de correlación	,138	1,000
		Sig. (bilateral)	,130	
		N	122	122

Figura 3.9. Análisis de correlación para las variables *Situación* y *Edad*.

Como se ha constatado, la variable *Estadío* condiciona la supervivencia a esta enfermedad. A modo de ejemplo y sin pérdida de generalidad, se ha tomado la decisión de incluir únicamente esta covariable ya que si bien el modelo será susceptible de ser aplicado a un vector de covariables de mayor dimensión, es necesario contar con un mínimo de casos para todos los subgrupos de distintos niveles de riesgo, en relación a cada una de las covariables que lo componen y también para los recuentos de las transiciones entre cada par de estados del sistema. La variable *Edad* no presenta correlación significativa con la Situación final del paciente, sin embargo, podría afectar al tiempo de permanencia en algún estado del sistema o en el sistema global. A pesar de lo antes dicho no se incluirá esta covariable en el modelo, no por carecer de interés sino por el restringido número de casos

en el subgrupo de pacientes de menor riesgo cuando se considera el cruce de la *Edad* y *Estadío*.

El siguiente diagrama representa las transiciones de los pacientes que constituyen el grupo de menor riesgo en relación a la covariable *Estadío*. Como se verá posteriormente, es a partir de este grupo que se obtendrán las matrices de *riesgo base* y donde la covariable *Estadío* tendrá valor asignado 0 (ver Figura 3.10).

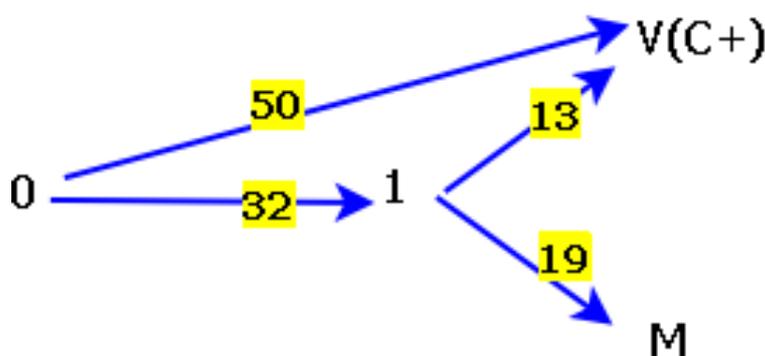


Figura 3.10. Diagrama del sistema multiestado que representa las distintas etapas de la enfermedad de los  $n = 82$  pacientes del grupo de menor riesgo en la covariable *Estadío*.

En la Tabla 3.1 se presentan los recuentos de las transiciones desde cada estado transitorio al resto de los estados del modelo simplificado. Análogamente para los dos subgrupos de riesgo mínimo en la Tabla 3.2 y de riesgo máximo en la Tabla 3.3, según la variable *Estadío*.

Estados Transitorios	0	1	M	C
0	0	61	0	61
1	0	0	43	18

Tabla 3.1. Transiciones entre los estados del modelo.

Estados Transitorios	0	1	M	C
0	0	32	0	50
1	0	0	19	13

Tabla 3.2. Transiciones entre los estados del modelo para el grupo de riesgo base o mínimo.

Estados Transitorios	0	1	M	C
0	0	29	0	11
1	0	0	24	5

Tabla 3.3. Transiciones entre los estados del modelo para el grupo de riesgo máximo.

### 3.3. Análisis de los tiempos de transición entre estados

El resumen estadístico descriptivo de los tiempos de permanencia en cada estado transitorio se presenta en la Tabla 3.4. Cabe destacar el amplio período que abarca en algunos casos más de 13 años de seguimiento para algunos de los pacientes que no presentan recidiva. De forma similar con un mínimo de 2 meses a más de 8 años en ciertos casos, es el tiempo de seguimiento que se ha hecho a aquellos pacientes que luego de una primera recidiva o metástasis dejan de ser observados en el estudio, con información censurada por derecha. Nótese que el tiempo medio de permanencia en el estado 0 antes de alcanzar una primera recidiva o metástasis es aproximadamente el triple del tiempo medio para el tiempo de permanencia en el estado 1, antes del fallecimiento. No obstante, la desviación típica para el tiempo de permanencia en ambos estados transitorios es similar.

Transición	n	Mínimo	Máximo	Media	Desviación Típica
0-1	61	3	53	19.10	10.14
0-C	61	6	165	63.44	44.34
1-M	43	1	65	6.35	10.21
1-C	18	2	104	35.39	34.38

Tabla 3.4. Descripción estadística de los tiempos (en meses) en cada estado transitorio del sistema.

Las correlaciones entre los tiempos de estadía en 0 y 1 (estados transitorios) ha resultado no significativa, tanto en el contraste paramétrico como en el no paramétrico para un nivel  $\alpha = 0.05$  de significación, como puede apreciarse en las tablas de las Figuras 3.11 y 3.12:

**Correlaciones**

		Tiempo hasta_1oC_ RL_Mt_RLy Mt_o_Salida	Tiempo entre_1_M
Tiempo_hasta_1oC_RL_ Mt_RLyMt_o_Salida	Correlación de Pearson	1	-,038
	Sig. (bilateral)		,807
	N	43	43
Tiempo_entre_1_M	Correlación de Pearson	-,038	1
	Sig. (bilateral)	,807	
	N	43	43

Figura 3.11. Análisis paramétrico de la correlación entre los tiempos de permanencia en los estados transitorios 0 y 1.

Correlaciones			Tiempo hasta 1 <sup>o</sup> RL M <sub>0</sub> y Salida	Tiempo entre 1 <sup>o</sup> M
Tau_b de Kendall	Tiempo hasta 1 <sup>o</sup> RL M <sub>0</sub> y Salida	Coefficiente de correlación	1,000	,062
		Sig. (bilateral)	.	,580
		N	43	43
	Tiempo entre 1 M	Coefficiente de correlación	,062	1,000
		Sig. (bilateral)	,580	.
		N	43	43
Rho de Spearman	Tiempo hasta 1 <sup>o</sup> RL M <sub>0</sub> y Salida	Coefficiente de correlación	1,000	,104
		Sig. (bilateral)	.	,505
		N	43	43
	Tiempo entre 1 M	Coefficiente de correlación	,104	1,000
		Sig. (bilateral)	,505	.
		N	43	43

Figura 3.12. Análisis no paramétrico de correlación entre los tiempos de permanencia en los estados transitorios 0 y 1.

A partir del algoritmo EM, se ha realizado un ajuste con distribución PH con  $p = 2$  para las 122 observaciones correspondientes a los tiempos de permanencia en el sistema, antes de la salida por el estado que representa la muerte del paciente. Téngase en cuenta que en este caso se trata de 43 tiempos de vida y 79 casos de censura, en que los pacientes salen del sistema por haber finalizado el tiempo del estudio.

Se obtuvo las siguientes estimaciones del vector  $\alpha$ , de la matriz  $A$  y de la matriz columna de absorción  $B$ , con las que se completa el generador por bloques sin la incorporación de covariables:

$$\hat{\alpha} = (0.2446, 0.7554), \hat{A} = \begin{bmatrix} -0.0978 & 0.0651 \\ 0.0159 & -0.0563 \end{bmatrix} \text{ y } \hat{B} = \begin{bmatrix} 0.0327 \\ 0.0404 \end{bmatrix}$$

En la Figura 3.13 se presenta la salida que ofrece el programa Mathematica del gráfico de la función de supervivencia estimada en el sistema global, a partir del algoritmo EM (para  $p = 2$  en la distribución PH) considerando los tiempos de vida junto a los tiempos censurados para los  $n = 122$  casos, frente a la estimación de la función de supervivencia que surge del estimador de Kaplan-Meier. Por otro lado, la Figura 3.14 presenta la

función de supervivencia estimada a través del primer modelo frente a la función empírica de supervivencia construida para los 43 tiempos hasta la muerte (no censurados). Una comparación entre ambas figuras evidencia que el procedimiento empleado a partir de las estimaciones en el presente modelo e incorporando datos censurados, genera mejor ajuste que el modelo que emplea el estimador de Kaplan-Meier.

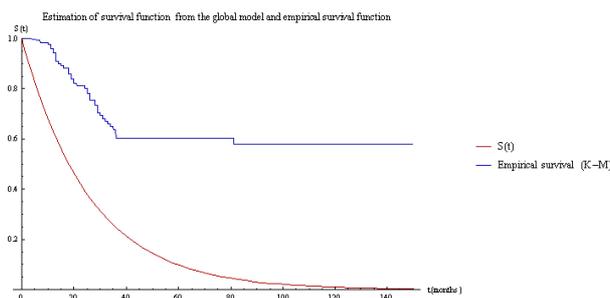


Figura 3.13. Gráfico de la función de supervivencia estimada en el sistema global a partir del algoritmo EM considerando censura y  $n = 122$  casos, frente a la que surge de aplicar la estimación por Kaplan-Meier.

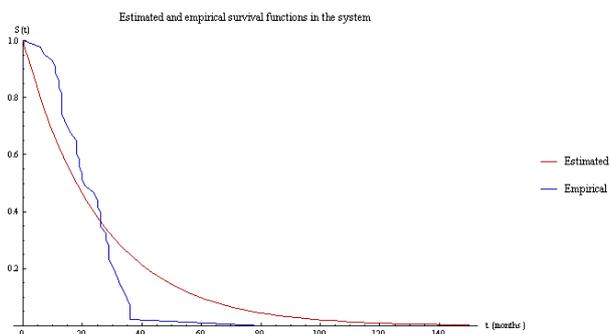


Figura 3.14. Gráfico de la función de supervivencia estimada en el sistema global, a partir del algoritmo EM considerando censura y  $n = 122$  casos, frente a la función empírica de supervivencia.

Para este último caso, el valor del estadístico de Kolmogorov-Smirnov es aproximadamente  $D = 0.2762$ , que resulta ser mayor que los dos valores  $D_{\alpha=0.05} = 0.2029$  y  $D_{\alpha=0.01} = 0.2432$ . Sin embargo, nótese que de  $n = 122$  casos hay únicamente 43 valores de tiempos de fallo sin censura y un total de  $n = 79$  tiempos censurados por derecha, lo

que representa casi un 65 % de las observaciones. Esto distorsiona bastante lo que a partir de ellos podemos obtener, a pesar de lo cual el ajuste aunque no satisfactorio, evidencia su bondad si se tiene en cuenta el peso de la censura en la base de datos.

### 3.4. Ajuste máximo verosímil del generador del sistema a partir del ingreso de covariables

Teniendo en cuenta que el modelo de Cox expuesto en los preliminares del Capítulo 1 no es aplicable directamente debido a que estamos frente a un modelo multidimensional, haremos una extensión del mismo donde  $\lambda(t|z)$  estará representada por la matriz de riesgos o razones de transición  $Q(z)$ , la función  $\lambda_0(t)$  será considerada como constante e igual a la matriz de riesgo o de razones de transición condicionada al conjunto de pacientes de menor riesgo, de acuerdo al análisis estadístico inicial, donde se asignará a la covariable valor 0. Este es el caso, como ya se ha argumentado antes, de los sujetos para los que la variable *Estadío* corresponde a alguna de las dos categorías  $T2$  o  $T2 - 3$ , en un conjunto de  $n = 82$  pacientes con *Estadío* en uno de los dos primeros niveles. Se deduce que se asignará valor 1 a la covariable en el estrato de pacientes para los que la variable *Estadío* corresponde a alguna de las dos categorías superiores  $T3$  o  $T4$ .

El procedimiento se ha efectuado para una sola covariable y en relación a la última simplificación del sistema multi-estados, dadas las restricciones de la base de datos de la que se dispone. Sin embargo, vale la pena destacar cómo se construiría el generador del modelo en función de un vector de covariables y de un sistema de mayores dimensiones; aunque luego no podamos completar las estimaciones en este caso, debido a las limitaciones de la información disponible. El objetivo es evidenciar la aplicabilidad de los MAM y la facilidad que ofrecen para el ingreso de las covariables de interés, en sistemas multidimensionales

tan complejos como sea necesario para representar las situaciones en su contexto.

**En relación a la primera simplificación del diagrama de transiciones que se ha presentado en la Figura 3.2**

El generador  $Q_0$  para ese conjunto de pacientes que constituyen el grupo de menor riesgo, viene dado por bloques y corresponde a la siguiente matriz:

$$Q_0 = \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix},$$

donde  $A$  es la matriz de razones de transición para los estados transitorios de  $\{0, 1, 2\}$  y  $B$  la matriz columna que contiene las razones de absorción por parte del estado  $M = 3$ . El tiempo de permanencia en cada estado transitorio  $i$  sigue una distribución  $PH(\alpha_i, A_i)$ , con vector de razones de absorción  $A_i^0, i = 0, 1, 2$ . Las probabilidades de transición entre dos estados transitorios distintos  $i$  y  $j$  se representan por  $p_{ij}$ :

$$A = \begin{pmatrix} A_0 & p_{01}A_0^0\alpha_1 & p_{02}A_0^0\alpha_2 \\ 0 & A_1 & p_{12}A_1^0\alpha_2 \\ 0 & 0 & A_2 \end{pmatrix} \quad B = \begin{pmatrix} 0 \\ p_{13}A_1^0 \\ A_2^0 \end{pmatrix}$$

En este contexto, el generador del sistema que depende del vector de covariables  $Z$  viene dado a partir del generador en el grupo de riesgo base antes expuesto, al que se han incorporado las covariables como sigue:

$$Q(Z) = \begin{pmatrix} A(Z) & B(Z) \\ 0 & 0 \end{pmatrix}$$

siendo:

$$A(Z) = \begin{pmatrix} A_0 [p_{01} \exp(\beta_{01}Z) + p_{02} \exp(\beta_{02}Z)] & p_{01}A_0^0 \alpha_1 \exp(\beta_{01}Z) & p_{02}A_0^0 \alpha_2 \exp(\beta_{02}Z) \\ 0 & A_1 [p_{12} \exp(\beta_{12}Z) + p_{13} \exp(\beta_{13}Z)] & p_{12}A_1^0 \alpha_2 \exp(\beta_{12}Z) \\ 0 & 0 & A_2 \exp(\beta_{23}Z) \end{pmatrix}$$

$$B(Z) = \begin{pmatrix} 0 \\ p_{13}A_1^0 \exp(\beta_{13}Z) \\ A_2^0 \exp(\beta_{23}Z) \end{pmatrix}$$

Por otro lado el vector de probabilidades de entrada en el sistema para los estados transitorios, teniendo en cuenta las fases de cada distribución PH específica, viene dado por  $\alpha = (\alpha_0, 0, 0, 0, 0)$ .

Cabe destacar la generalidad del método y la versatilidad en la adaptación a otras muchas situaciones. Si se trata de un sistema más complejo, simplemente se definen los bloques no diagonales de  $Q(Z)$  y se tienen en cuenta luego para la construcción de los diagonales, a partir de las propiedades del generador  $Q_0$  en el estrato de pacientes de menor riesgo para todas las covariables involucradas (donde cada una tiene asignado valor 0).

**En relación a la segunda simplificación del diagrama de transiciones que se ha presentado en la Figura 3.3**

Como ya lo hemos adelantado anteriormente, todas las estimaciones se deben realizar a partir de una base de datos en la que se pueda disponer de un recuento mínimo de transiciones entre todos los pares de estados que componen  $E$ . Es así que se ha tomado

la decisión de ejemplificar los cálculos de las estimaciones de las medidas de interés a partir de la última simplificación del sistema multi-estado.

Es en el estrato de pacientes donde la variable  $Z$ , que representa el *Estadio* de la enfermedad según el diagnóstico inicial toma valor 0, donde se han calculado los bloques de *riesgo base* que componen el generador inicial:

$$Q_0 = \left[ \begin{array}{cc|c} A_0 & A_0^0 \cdot \alpha_1 & 0_{2 \times 1} \\ 0_{2 \times 2} & A_1 & A_1^0 \\ \text{---} & \text{---} & \text{---} \\ 0_{1 \times 2} & 0_{1 \times 2} & 0 \end{array} \right]$$

La incorporación de las covariables al modelo nos permite obtener:

$$Q(Z) = \left[ \begin{array}{cc|c} A_0 \cdot \exp(\beta'_{01} \cdot Z) & A_0^0 \cdot \alpha_1 \cdot \exp(\beta'_{01} \cdot Z) & 0_{2 \times 1} \\ 0_{2 \times 2} & A_1 \exp(\beta'_{1M} \cdot Z) & A_1^0 \cdot \exp(\beta'_{1M} \cdot Z) \\ \text{---} & \text{---} & \text{---} \\ 0_{1 \times 2} & 0_{1 \times 2} & 0 \end{array} \right]$$

Las estimaciones de los bloques principales de  $Q_0$  para la subpoblación de pacientes en riesgo mínimo en relación a la covariable *Estadio* han podido obtenerse a partir de la sucesiva aplicación del algoritmo EM. Todos los modelos han presentado excelente ajuste para  $p = 2$  y 1000 pasos.

$$\alpha_0 = (1.0000, 0.0000) \quad A_0 = \begin{bmatrix} -0.1017 & 0.1017 \\ 0.0000 & -0.1017 \end{bmatrix} \quad A_0^0 = \begin{bmatrix} 0.0000 \\ 0.1017 \end{bmatrix}$$

$$\alpha_1 = (1.0000, 0.0000) \quad A_1 = \begin{bmatrix} -0.8103 & 0.8103 \\ 0.0131 & -0.3372 \end{bmatrix} \quad A_1^0 = \begin{bmatrix} 0.0000 \\ 0.3242 \end{bmatrix}$$

Los parámetros  $\beta_{01}$  y  $\beta_{1M}$ , han sido estimados a partir de la función *log-verosimilitud parcial* dado que según se ha argumentado en la metodología, es idónea para su aplicación cuando se dispone de una base de datos de supervivencia con casos censurados. Para la estimación de parámetros, se ha optimizado el logaritmo de la función log-verosimilitud parcial propuesta por Breslow debido a múltiples coincidencias en los tiempos de supervivencia.

En la Tabla 3.5 puede apreciarse en cada caso la estimación de cada parámetro a partir del algoritmo de Nelder-Mead adaptado a la función matricial implicada (ver Apéndice IV), el número de iteraciones del algoritmo y una aproximación al *error cuadrático medio* (MCE):

Parámetro	Estimación	k iteraciones	MCE	$exp(\beta)$
$\beta_{01}$	0.8821	15	$9.14 \times 10^{-9}$	2.4159
$\beta_{1M}$	0.5134	15	$2.91 \times 10^{-9}$	1.6710

Tabla 3.5. Estimación de los parámetros a partir de la función *Log-verosimilitud parcial*.

Nótese que la estimación de  $\beta_{01}$  pone de manifiesto un incremento de un 141.59% en las intensidades de transición entre las fases virtuales del estado 0 y desde dicho estado a las del estado 1, en el grupo de mayor riesgo ( $Z = 1$ ) respecto a las correspondientes intensidades de transición en el grupo de menor riesgo ( $Z = 0$ ). Análogamente puede observarse que la estimación de  $\beta_{1M}$  implica un incremento de un 67.10% en las intensidades de transición entre las fases virtuales del estado 1 y desde dicho estado al estado  $M$ , en el grupo de mayor riesgo ( $Z = 1$ ) respecto a las correspondientes intensidades de transición en el grupo de menor riesgo ( $Z = 0$ ).

Si se particulariza el generador del sistema  $Q(Z)$  para los distintos valores de la covariable, nos permitirá comparar las medidas de interés para cada uno de los subgrupos de riesgo. Siendo:

$$Q(Z) = \begin{bmatrix} A(Z) & B(Z) \\ 0 & 0 \end{bmatrix}$$

A partir de las correspondientes estimaciones, los bloques principales de los generadores en ambos grupos de riesgo son:

$$A(Z = 0) = \begin{bmatrix} -0.1017 & 0.1017 & 0 & 0 \\ 0 & -0.1017 & 0.1017 & 0 \\ 0 & 0 & -0.8103 & 0.8103 \\ 0 & 0 & 0.0131 & -0.3372 \end{bmatrix}$$

$$B(Z = 0) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.3242 \end{bmatrix}$$

$$A(Z = 1) = \begin{bmatrix} -0.2458 & 0.2458 & 0. & 0 \\ 0 & -0.2458 & 0.2458 & 0 \\ 0 & 0 & -1.3540 & 1.3540 \\ 0 & 0 & 0.0218 & -0.5635 \end{bmatrix}$$

$$B(Z = 1) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.5417 \end{bmatrix}$$

Las funciones de supervivencia para ambos grupos, se han obtenido como es habitual en este modelo a partir de la expresión:

$$S(t, Z) = 1 - \alpha B(t, Z)$$

Teniendo en cuenta que todo paciente ingresa al sistema en el estado 0 para  $t = 0$  y teniendo en cuenta el vector de probabilidades válido para el grupo de menor riesgo, es que fijamos el vector de distribución de probabilidades iniciales en  $\alpha = (1, 0, 0, 0)$ .

El gráfico comparativo entre las funciones de supervivencia estimadas en ambos grupos de riesgo, es el que se presenta en la Figura 3.15:

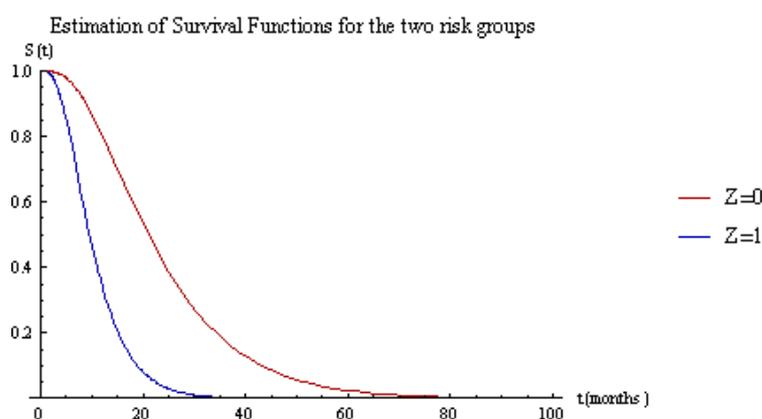


Figura 3.15. Comparación de funciones de supervivencia para ambos grupos de riesgo, en relación a la covariable *Estadio*.

El gráfico comparativo de las funciones de supervivencia estimada y empírica, para el subgrupo de riesgo mínimo (cuando la covariable *Estadio* toma valor  $Z = 0$ ), es el que se registra en la Figura 3.16. El valor del estadístico de Kolmogorov-Smirnov es aproximadamente  $D = 0.1746$ , que resulta ser menor que los dos valores  $D_{\alpha=0.05} = 0.3015$  y  $D_{\alpha=0.01} = 0.3614$ . Al no pertenecer el valor de  $D$  a las regiones críticas con extremos inferiores 0.3015 y 0.3614 admitimos que el ajuste logrado entre la supervivencia empírica y la correspondiente función de supervivencia estimada por el modelo es satisfactorio para ambos niveles de significación, a pesar de la cantidad de datos censurados que representan un 77% del subgrupo de pacientes.

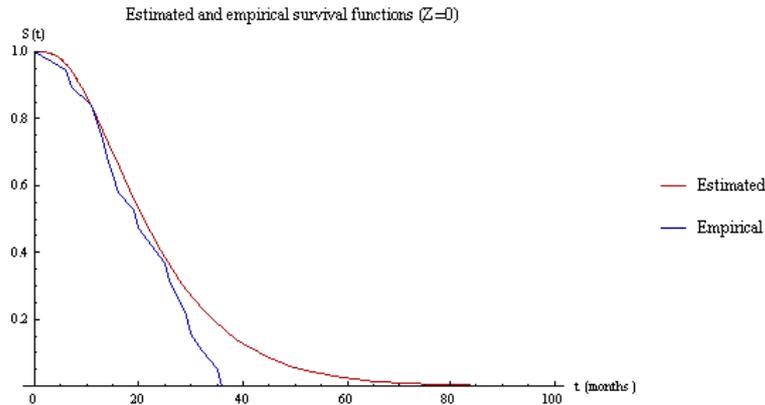


Figura 3.16. Comparación de funciones de supervivencia empírica y estimada para el grupo de riesgo base en relación a la covariable *Estadío* ( $Z = 0$ ).

Si consideramos el intervalo de tiempo correspondiente a diez años,  $[0, T = 120]$  (en meses), el tiempo medio hasta la muerte se ha estimado como:

$$\widehat{MTTD}(T) = \int_0^T S(t, Z = 0) dt = 24.02$$

El valor anterior (en meses) se ha obtenido a partir de un algoritmo de aproximaciones sucesivas sobre particiones cada vez más finas, en  $i = 41$  iteraciones y una diferencia entre los dos últimos términos de la sucesión estimado en  $\Delta = 9.45 \times 10^{-6}$ .

La distorsión en el ajuste entre la función de supervivencia estimada por el modelo y la empírica para el subgrupo de pacientes de riesgo máximo, cuando  $Z = 1$  (ver Figura 3.17), se da por razones vinculadas al tamaño de la muestra en dicho caso y a la proporción de censura. Nótese que en este subgrupo se cuenta solamente con 40 datos, menos de la mitad de casos que en la estimación anterior. A su vez, de estos, poco más de la mitad son no censurados (24 casos). Si se contase con un mayor tamaño muestral dentro del estrato, se podría conseguir mucho mejor ajuste e inferencias más precisas en la población, como ya se vio en el caso del subgrupo de menor riesgo.

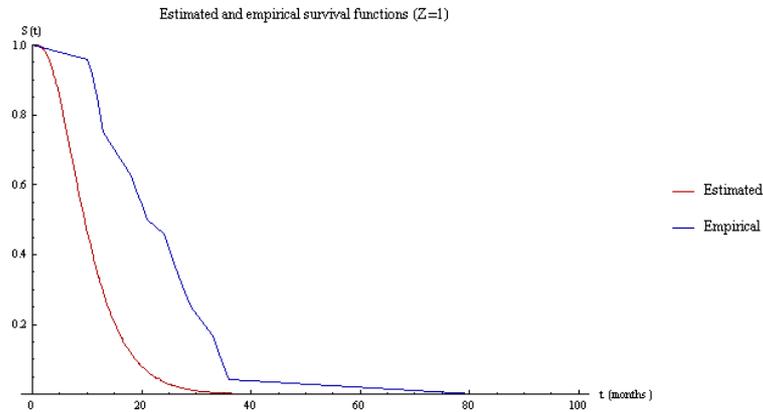


Figura 3.17. Comparación de funciones de supervivencia empírica y estimada para el grupo de riesgo máximo y en relación a la covariable *Estadío* ( $Z = 1$ ).

El tiempo medio hasta la muerte estimado en el intervalo de tiempo  $[0, T = 120]$  (en meses) para pacientes en el grupo de riesgo máximo, ha sido calculado como:

$$\widehat{MTTD}(T) = \int_0^T S(t, Z = 1) dt = 10.75$$

El valor anterior (en meses), se ha obtenido a partir de un algoritmo de aproximaciones sucesivas en  $i = 65$  iteraciones y una diferencia entre los dos últimos términos de la sucesión estimado en  $\Delta = 9.98 \times 10^{-6}$ . Este valor no alcanza ni siquiera a un año de vida, mientras que en el grupo de riesgo mínimo se obtuvo aproximadamente dos años como promedio.

---

## Estimaciones con procesos semi-Markovinos

---

En este capítulo presentamos un modelo multi-estado alternativo que responde a la caracterización de las etapas de la enfermedad según la información recogida en la base de datos inicial, constituida por las observaciones realizadas durante un período fijo de quince años a 960 pacientes. A pesar del gran volumen de información censurada se obtienen buenos ajustes de las principales medidas de supervivencia, a partir de las estimaciones con procesos semi-Markovianos, gracias a las propiedades de consistencia de los estimadores construidos.

En la Figura 4.1, se presenta el diagrama de transición entre los estados de  $E = E_0 \cup D$ , donde  $E_0 = \{0, 1, 2, 3\}$  es el conjunto de los estados transitorios y  $D = \{4\}$  es el conjunto unitario con un estado absorbente que representa la progresión de la enfermedad ( $P$ ). La salida del sistema sin progresión nuevamente se representa con  $NP$ , pero según se ha argumentado en los preliminares,  $NP$  no ha sido considerado un estado sino que permite identificar los valores de tiempos de supervivencia censurados.

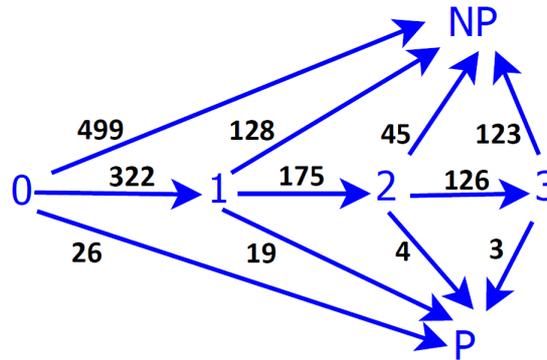


Figura 4.1. Diagrama de transiciones entre los estados del sistema.

#### 4.1. Construcción de los estimadores de las principales medidas de supervivencia

El proceso estocástico  $(S_n, J_n)$  define un proceso de renovación de Markov o, equivalentemente, un proceso semi-Markoviano en el espacio de estados  $E$  dado que es posible suponer que se verifica:

$$\begin{aligned} P(J_{n+1} = j, S_{n+1} - S_n \leq t \mid J_0, \dots, J_n, S_1, \dots, S_n) \\ = P(J_{n+1} = j, S_{n+1} - S_n \leq t \mid J_n) \quad (a.s.), \end{aligned}$$

donde  $j \in E$  y  $t \in R^+$ . El proceso  $Z = (Z_t, t \in R^+)$  se dice que es un *proceso semi-Markoviano* (SMP).

Recordando que:

$$Q_{ij}(t) = P(J_{n+1} = j, S_{n+1} - S_n \leq t \mid J_n = i),$$

con  $i, j \in E$  y  $t \geq 0$ .

La matriz  $Q(t) = (Q_{ij}(t))_{(i,j) \in E \times E}$ , para nuestro modelo será:

$$Q(t) = \left[ \begin{array}{cccc|c} 0 & Q_{01}(t) & 0 & 0 & Q_{04}(t) \\ 0 & 0 & Q_{12}(t) & 0 & Q_{14}(t) \\ 0 & 0 & 0 & Q_{23}(t) & Q_{24}(t) \\ 0 & 0 & 0 & 0 & Q_{34}(t) \\ \hline 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

Donde la matriz que corresponde al conjunto de estados transitorios es:

$$Q_0(t) = \left[ \begin{array}{cccc} 0 & Q_{01}(t) & 0 & 0 \\ 0 & 0 & Q_{12}(t) & 0 \\ 0 & 0 & 0 & Q_{23}(t) \\ 0 & 0 & 0 & 0 \end{array} \right]$$

Supondremos que  $(J_n)$  es una cadena de Markov homogénea en el tiempo con el espacio de estados  $E$  y probabilidades de transición

$$p_{ij} = Q_{ij}(+\infty) = P(J_{n+1} = j | J_n = i) \quad (4.1)$$

Siendo  $F_{ij}(t) = P(X_{n+1} \leq t \mid J_{n+1} = j, J_n = i)$  la función de distribución condicional para el tiempo entre saltos, se tiene:

$$\begin{aligned} F_{ij}(t) &= \frac{P(X_{n+1} \leq t, J_{n+1} = j \mid J_n = i)}{P(J_{n+1} = j \mid J_n = i)} \\ &= \frac{Q_{ij}(t)}{p_{ij}} \end{aligned}$$

para  $p_{ij} \neq 0$ , y

$$Q_{ij}(t) = p_{ij}F_{ij}(t), \quad i, j \in E, \quad t \in \mathbb{R}^+$$

A su vez la función de distribución para el tiempo de permanencia en  $i$  del proceso  $Z$ , como se argumentó en los preliminares, es:

$$H_i(t) = \sum_{j \in E} Q_{ij}(t), \quad i \in E, \quad t \geq 0$$

La función de supervivencia se vincula con las funciones antes expuestas, a partir de la definición de la convolución de funciones, de la siguiente forma:

$$\begin{aligned} S(t) &= \alpha [I - Q_0(t)]^{*(-1)} * [I - H_0(t)] \mathbf{1} \\ &= \alpha [I - Q_0(t)]^{*(-1)} * \bar{H}_0(t) \end{aligned}$$

donde  $\bar{H}_0(t)$  es una matriz columna cuyo  $i$ -ésimo elemento es:

$$\bar{H}_i(t) = 1 - H_i(t) = 1 - \sum_{j \in E} Q_{ij}(t), \quad i \in E_0$$

Como en esta aplicación el vector de probabilidades inicial es  $\alpha = (1, 0, 0, 0)$ ,

$$\bar{H}_0(t) = \begin{bmatrix} 1 - (Q_{01}(t) + Q_{04}(t)) \\ 1 - (Q_{12}(t) + Q_{14}(t)) \\ 1 - (Q_{23}(t) + Q_{24}(t)) \\ 1 - Q_{34}(t) \end{bmatrix}$$

y

$$[I - Q_0(t)]^{*(-1)} = \begin{bmatrix} 1 & Q_{01}(t) & Q_{01}(t) * Q_{12}(t) & Q_{01}(t) * Q_{12}(t) * Q_{23}(t) \\ 0 & 1 & Q_{12}(t) & Q_{12}(t) * Q_{23}(t) \\ 0 & 0 & 1 & Q_{23}(t) \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Entonces la función de supervivencia será estimada a partir de las estimaciones de los núcleos y de la relación:

$$\begin{aligned} S(t) &= 1 - [Q_{04}(t) + Q_{01}(t) * Q_{14}(t) \\ &+ Q_{01}(t) * Q_{12}(t) * Q_{24}(t) \\ &+ Q_{01}(t) * Q_{12}(t) * Q_{23}(t) * Q_{34}(t)] \end{aligned}$$

### **Estimación de las principales medidas de supervivencia, teniendo en cuenta las observaciones censuradas**

Estamos interesados en la estimación de  $Q_{ij}(t)$  a partir de una muestra de  $K$  trayectorias individuales, cada una de ellas consideradas como un SMP  $Z^{(k)}$ ,  $k = 1, \dots, K$ , i.i.d. con un núcleo común  $Q(t)$  y un vector inicial de probabilidades  $\alpha$ .

Como no hay razón para asumir que el tiempo de permanencia en el estado  $i$  depende del estado siguiente, es posible escribir:

$$Q_{ij}(t) = p_{ij}F_i(t), \quad i, j \in E, \quad t \in \mathbb{R}^+,$$

donde:

$$F_i(t) = P(X_{n+1} \leq t \mid J_n = i).$$

Si  $T$  es la longitud del intervalo de observación en el estudio, para cada proceso individual  $k$  se define:

- El número total de saltos antes de  $T$ :

$$N_T^k = \sup \left\{ n \in \mathbb{N} : S_n^k \leq T \right\}$$

- El número total de visitas al estado  $i$  antes de  $T$ :

$$N_T^k(i) = \sum_{h=1}^{N_T^k} \mathbf{1}_{\{J_{h-1}^k = i\}}, \quad i \in E$$

- El número total de saltos desde  $i$  a  $j$  antes de  $T$ :

$$N_T^k(i, j) = \sum_{h=1}^{N_T^k} \mathbf{1}_{\{J_{h-1}^k = i, J_h^k = j\}}, \quad i, j \in E$$

Nótese que en nuestra aplicación, dado que el proceso de cada paciente  $k$  es ascendente en el sistema multi-estado, el número total de visitas al estado  $i$  antes de  $T$  de ese paciente es 1 o 0. De igual forma el número de saltos desde  $i$  a  $j$  antes de  $T$ . No obstante, el planteo pretende generalizar estimadores para otros casos.

Para los  $K$  procesos, el total del número de visitas al estado  $i$  y el número total de saltos desde  $i$  a  $j$  antes de  $T$ , que coinciden en este caso con el total de pacientes que arriban a  $i$  y que transitan entre  $i$ - $j$ , son respectivamente:

$$N_T(i, K) = \sum_{k=1}^K N_T^k(i), \quad i \in E$$

$$N_T(i, j, K) = \sum_{k=1}^K N_T^k(i, j), \quad i, j \in E$$

Es posible aplicar un filtro a nuestra base de datos para seleccionar los pacientes que visitan el estado  $i$ , luego será posible estimar las probabilidades de transición:

$$\hat{p}_T(i, j, K) = \frac{N_T(i, j, K)}{N_T(i, K)}$$

Para cada estado transitorio  $i$ , teniendo en cuenta que

$$\hat{F}_i(t) = 1 - \hat{F}_i(t), \quad i \in E_0$$

y que además a partir de una base de datos con observaciones censuradas  $\hat{F}_i(t)$  puede ser estimada a través del estimador de Kaplan-Meier:

$$\hat{F}_i(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{r_j}\right), \quad i \in E_0$$

siendo  $0 < t_1 < \dots < t_m$  los tiempos de arribo al estado absorbente 4 desde el estado  $i$ , con  $r_j$  pacientes en riesgo justo antes del instante  $t_j$  y  $d_j$  fallos (progresiones) en cada

tiempo de supervivencia  $t_j$ , donde  $j \in \{1, \dots, m\}$ .

A partir de los tiempos de permanencia en cada estado y sus correspondientes códigos de censura, es posible automatizar todos los cálculos involucrados en el estimador de Kaplan-Meier, evitando errores de cálculo manual. La definición de  $\widehat{F}_i(t)$  como función continua y escalonada constituye el paso previo a obtener las entradas de la matriz  $Q$ . Posteriormente a partir de la estimación de las correspondientes convoluciones, se obtendrá la estimación de la función de supervivencia.

Finalmente, la estimación del tiempo medio de progresión de la enfermedad, hasta  $T$ , podrá realizarse a través de la aproximación a la integral definida para la función anterior en el intervalo  $[0, T]$ , a partir de sucesivas aproximaciones con funciones escalonadas en particiones cada vez más finas del intervalo de tiempo.

$$\widehat{MTTP}_{K,T} = \int_0^T \widehat{S}_{K,T}(t) dt$$

## 4.2. Estimaciones a partir de la base de datos

Siguiendo los procedimientos que han sido fundamentados en los preliminares (Capítulo 1), se han realizado las estimaciones de los núcleos correspondientes al SMP, de la función de supervivencia y del tiempo medio hasta la progresión de la enfermedad.

Las probabilidades condicionadas que corresponden a las transiciones entre los estados del sistema han sido organizadas en la matriz siguiente:

$$\widehat{P}(T, K) = \begin{bmatrix} 0 & \widehat{p}_{01} & 0 & 0 & \widehat{p}_{04} & \widehat{p}_{05} \\ 0 & 0 & \widehat{p}_{12} & 0 & \widehat{p}_{14} & \widehat{p}_{15} \\ 0 & 0 & 0 & \widehat{p}_{23} & \widehat{p}_{24} & \widehat{p}_{25} \\ 0 & 0 & 0 & 0 & \widehat{p}_{34} & \widehat{p}_{35} \end{bmatrix}$$

Las estimaciones de los valores no nulos de la matriz anterior, se presentan en la Tabla 4.1.

$\widehat{p}_{01} = \frac{322}{847} \simeq 0.3802$	$\widehat{p}_{04} = \frac{26}{847} \simeq 0.0307$	$\widehat{p}_{05} = \frac{499}{847} \simeq 0.5891$
$\widehat{p}_{12} = \frac{175}{322} \simeq 0.5435$	$\widehat{p}_{14} = \frac{19}{322} \simeq 0.0590$	$\widehat{p}_{15} = \frac{128}{322} \simeq 0.3975$
$\widehat{p}_{23} = \frac{126}{175} \simeq 0.7200$	$\widehat{p}_{24} = \frac{4}{175} \simeq 0.0229$	$\widehat{p}_{25} = \frac{45}{175} \simeq 0.2571$
	$\widehat{p}_{34} = \frac{3}{126} \simeq 0.0238$	$\widehat{p}_{35} = \frac{123}{126} \simeq 0.9762$

Tabla 4.1. Estimación de las probabilidades condicionadas de transición

Las estimaciones de las entradas no nulas del núcleo del proceso  $Q(t)$  se representan gráficamente en las Figuras 4.2 y 4.3:

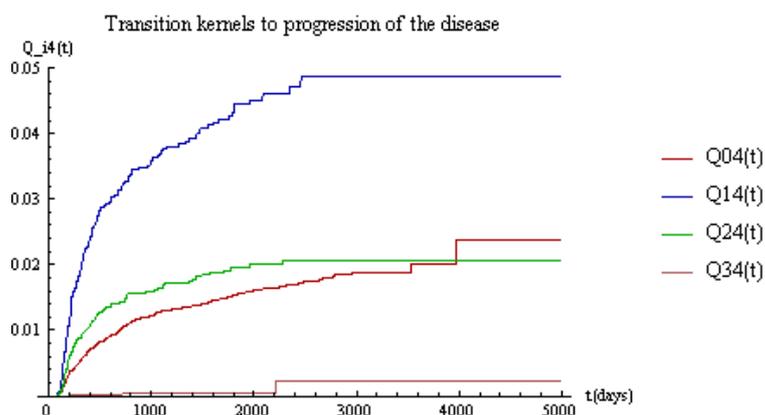


Figura 4.2. Estimaciones de las componentes de  $Q(t)$ , en relación al estado absorbente 4.

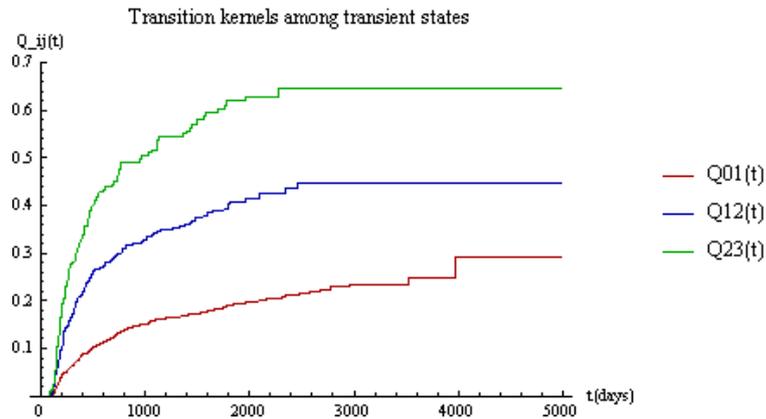


Figura 4.3. Estimaciones de las componentes de  $Q(t)$ , en relación a los estados transitorios.

La estimación de la función de supervivencia ha sido calculada a partir del procedimiento descrito en la metodología y representada en la Figura 4.4:

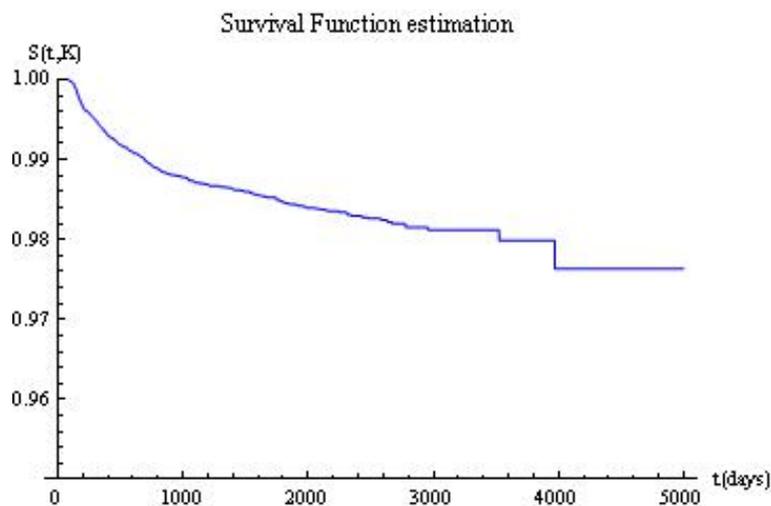


Figura 4.4. Representación de la estimación de la función de supervivencia.

Algunos valores de tiempo y sus correspondientes valores en la función de supervivencia se presentan en la Tabla 4.2. Podemos observar su comportamiento a lo largo del tiempo, lo que confirma la elevada supervivencia a la enfermedad.

t (days)	$\widehat{S}(t)$	t (days)	$\widehat{S}(t)$
100	0.9998	2600	0.9824
200	0.9966	2700	0.9820
300	0.9949	2800	0.9815
400	0.9930	2900	0.9815
500	0.9918	3000	0.9812
600	0.9909	3100	0.9812
700	0.9898	3200	0.9812
800	0.9888	3300	0.9812
900	0.9881	3400	0.9812
1000	0.9878	3500	0.9812
1100	0.9872	3600	0.9799
1200	0.9869	3700	0.9799
1300	0.9866	3800	0.9799
1400	0.9863	3900	0.9799
1500	0.9860	4000	0.9764
1600	0.9856	4100	0.9764
1700	0.9853	4200	0.9764
1800	0.9846	4300	0.9764
1900	0.9843	4400	0.9764
2000	0.9840	4500	0.9764
2100	0.9837	4600	0.9764
2200	0.9836	4700	0.9764
2300	0.9834	4800	0.9764
2400	0.9830	4900	0.9764
2500	0.9826	5000	0.9764

Tabla 4.2. Estimación de los valores en la función supervivencia en relación al tiempo (en días, sin partición).

La estimación del tiempo medio hasta la progresión de la enfermedad en el intervalo  $[0, T]$ , corresponde al siguiente cálculo integral:

$$\widehat{MTTP}_{K,T} = \int_0^T \widehat{S}_{K,T}(t) dt$$

donde hemos considerado  $T = 5000$  días, resultando  $\widehat{MTTP}_{K,T} = 4918.17$  (días). El cálculo ha sido realizado usando el procedimiento que se puede consultar en el apéndice V. En

este caso se requirió de  $i = 1549$  pasos y la correspondiente integral fue calculada por aproximaciones sucesivas con el cálculo de áreas de regiones escalonadas sobre particiones cada vez más finas y un error menor que  $2.31 \times 10^{-7}$ . El error ha sido calculado en cada paso teniendo en cuenta la convergencia de la sucesión de las áreas de las funciones escalonadas a la integral definida en  $[0, T]$  y resulta del valor absoluto de la diferencia de los últimos dos términos.

### 4.3. Análisis de la consistencia de los estimadores propuestos

Analizamos a continuación la consistencia fuerte de los estimadores empleados anteriormente y definidos en la metodología de trabajo, cuando el número de pacientes  $K \rightarrow +\infty$  para un valor fijo  $T > 0$ .

#### Estimadores de las probabilidades de transición

El estimador de las probabilidades de transición de un proceso semi-Markoviano es fuertemente consistente, es decir,

$$\hat{P}_T(i, j, K) = \frac{\sum_{k=1}^K N_T^k(i, j)}{\sum_{k=1}^K N_T^k(i)} = \frac{N_T(i, j, K)}{N_T(i, K)} \xrightarrow[K \rightarrow +\infty]{a.s.} p_{ij}$$

En el contexto de procesos semi-Markovianos se tiene:

$$p_{ij} = P(J_n = j | J_{n-1} = i), \quad i, j \in E$$

Sean  $X^k = N_T^k(i, j) - p_{ij}N_T^k(i)$ ,  $k = 1, \dots, K$ , con  $E[X^{(k)}] = 0$ .

Estamos interesados en este valor porque sabemos que de acuerdo con la Ley Fuerte de los Grandes Números, si  $K \rightarrow \infty$ , se da la siguiente convergencia casi segura:

$$\bar{X}_K = \frac{1}{K} \sum_{k=1}^K \left( N_T^k(i, j) - p_{ij} N_T^k(i) \right) \xrightarrow[K \rightarrow +\infty]{a.s.} a$$

Demostraremos que  $a = 0$  a partir de lo que obtendremos:

$$\frac{\sum_{k=1}^K N_T^k(i, j)}{\sum_{k=1}^K N_T^k(i)} = \frac{N_T(i, j, K)}{N_T(i, K)} \xrightarrow[K \rightarrow +\infty]{a.s.} p_{ij}$$

A partir del estado  $l \in E$ , donde  $l = J_0$ :

$$\begin{aligned} E_l \left[ N_T^k(i, j) \right] &= l \left[ \sum_{h=1}^{N_T^k} \mathbb{1}_{\{J_{h-1}^k=i, J_h^k=j\}} \right] \\ &= E_l \left[ E_l \left[ \sum_{h=1}^{N_T^k} \mathbb{1}_{\{J_{h-1}^k=i, J_h^k=j\}} \mid N_T^k \right] \right] \\ &= \sum_{n \geq 1} E_l \left[ \sum_{h=1}^{N_T^k} \mathbb{1}_{\{J_{h-1}^k=i, J_h^k=j\}} \mid N_T^k = n \right] P_l \left( N_T^k = n \right) \\ &= \sum_{n \geq 1} E_l \left[ \sum_{h=1}^n \mathbb{1}_{\{J_{h-1}^k=i, J_h^k=j\}} \right]_l \left( N_T^k = n \right) \\ &= \sum_{n \geq 1} \sum_{h=1}^n P_l \left( J_{h-1}^k = i, J_h^k = j \right) P_l \left( N_T^k = n \right) \\ &= \sum_{n \geq 1} \sum_{h=1}^n P_l \left( J_{h-1}^k = i \right) P_l \left( J_h^k = j \mid J_{h-1}^k = i \right) P_l \left( N_T^k = n \right) \\ &= p_{ij} \sum_{n \geq 1} \sum_{h=1}^n P^{h-1}(l, i) P_l \left( N_T^k = n \right) \end{aligned}$$

Con similar procedimiento podemos plantear:

$$\begin{aligned}
E_l \left[ N_T^k(i) \right] &= E_l \left[ \sum_{h=1}^{N_T^k} \mathbb{1}_{\{J_{h-1}^k=i\}} \right] \\
&= E_l \left[ E_l \left[ \sum_{h=1}^{N_T^k} \mathbb{1}_{\{J_{h-1}^k=i\}} \mid N_T^k \right] \right] \\
&= \sum_{n \geq 1} E_l \left[ \sum_{h=1}^{N_T^k} \mathbb{1}_{\{J_{h-1}^k=i\}} \mid N_T^k = n \right] P_l \left( N_T^k = n \right) \\
&= \sum_{n \geq 1} E_l \left[ \sum_{h=1}^n \mathbb{1}_{\{J_{h-1}^k=i\}} \right] P_l \left( N_T^k = n \right) \\
&= \sum_{n \geq 1} \sum_{h=1}^n P_l \left( J_{h-1}^k = i \right) P_l \left( N_T^k = n \right) \\
&= \sum_{n \geq 1} \sum_{h=1}^n P^{h-1}(l, i) P_l \left( N_T^k = n \right)
\end{aligned}$$

Finalmente:

$$E \left[ N_T^k(i, j) - p_{ij} N_T^k(i) \right] = 0$$

### El estimador producto límite de Kaplan-Meier

Sean  $X_1, \dots, X_n$  variables aleatorias positivas e independientes con una función de distribución continua común  $F$ . Con independencia de estas variables sean además  $U_1, \dots, U_n$  variables aleatorias independientes y positivas con distribución común pero que puede ser no continua  $G$ .

Gill en [19] y Borgan en [9], analizan el problema concerniente a las inferencias no paramétricas de  $F$ , basadas en observaciones censuradas  $(C_i, \delta_i), i = 1, \dots, n$ , siendo:

$$C_i = \min \{X_i, U_i\}, \quad \delta_i = \mathbf{1}_{\{X_i \leq U_i\}}$$

$H$  es la distribución de las variables  $C_i$  dada por:

$$1 - H = (1 - F)(1 - G)$$

$F$  es estimada habitualmente a partir del estimador propuesto por Kaplan y Meier en 1958, definiendo procesos  $N$  e  $Y$  en  $[0, +\infty)$  de la siguiente forma:

$$N(t) = \#\{i : C_i \leq t, \delta_i = 1\} \quad \text{y} \quad Y(t) = \#\{i : C_i \geq t\}$$

Entonces:

$$\begin{aligned} \hat{F}(t) &= 1 - \hat{F}(t) \\ &= 1 - \prod_0^t \left(1 - \frac{\Delta N(s)}{Y(s)}\right), \end{aligned}$$

En la ecuación anterior  $\Delta N(s) = N(s) - N(s-)$ , donde  $N(s-)$  denota el límite por la izquierda de la función  $N(t)$  en  $s$ , es decir, el límite de  $N(t)$  cuando  $t \rightarrow s-$ . Entonces,  $\Delta N(s)$  define el número de sucesos (habitualmente muertes) en el instante  $s$ .

El lema 2.8 de Gill en [19] nos asegura que:

Para todo  $\tau$  tal que  $H(\tau^-) < 1$ :

$$\sup_{t \leq \tau} |\widehat{F}(t) - F(t)| \xrightarrow[n \rightarrow +\infty]{a.s.} 0$$

### Estimadores de los núcleos y de la función de supervivencia

A partir de las dos últimas propiedades concluimos que:

El estimador  $\widehat{Q}_{ij}(t, K)$  es fuertemente consistente, es decir,

$$\widehat{Q}_{ij}(t, K) \xrightarrow[K \rightarrow +\infty]{a.s.} Q_{ij}(t)$$

La convolución de los estimadores de los núcleos es además fuertemente consistente:

$$\left( \widehat{Q}_{ij} * \widehat{Q}_{jh} \right) (t, K) \xrightarrow[K \rightarrow +\infty]{a.s.} (Q_{ij} * Q_{jh})(t), \quad i, j, h \in E$$

Sabemos que  $|\widehat{Q}_{ij}(t, K)| \leq 1$  y  $|\widehat{Q}_{ij}(t, K)\widehat{Q}_{jh}(t, K)| \leq 1$ . A partir de la proposición anterior, sabemos que

$$\widehat{Q}_{ij}(t, K) \xrightarrow[K \rightarrow +\infty]{a.s.} Q_{ij}(t)$$

Es así que:

$$\widehat{Q}_{ij}(s, K)\widehat{Q}_{jh}(t-s, K) \xrightarrow[K \rightarrow +\infty]{a.s.} Q_{ij}(s)Q_{jh}(t-s)$$

Finalmente por el teorema de la convergencia dominada de Lebesgue, podemos concluir la demostración.

A partir de la definición de la función de supervivencia para esta aplicación y en virtud de todas las propiedades previas, obtenemos el siguiente enunciado:

El estimador  $\widehat{S}_{K,T}(t)$  es fuertemente consistente, es decir,

$$\widehat{S}_{K,T}(t) \xrightarrow[K \rightarrow +\infty]{a.s.} S(t)$$

### Estimador del tiempo medio hasta la progresión de la enfermedad

El MTTP en el intervalo  $[0, T]$  es

$$MTTP_T = \int_0^T S(t) dt$$

Dicho valor ha sido estimado por:

$$\widehat{MTTP}_{K,T} = \int_0^T \widehat{S}_{K,T}(t) dt$$

El estimador  $\widehat{MTTP}_{K,T}$  es fuertemente consistente, es decir,

$$\widehat{MTTP}_{K,T} \xrightarrow[K \rightarrow +\infty]{a.s.} MTTP_T$$

Sabemos que:

- $|\widehat{S}_{K,T}(t)| \leq 1$ , lo que resulta evidente dada su definición.

- $\widehat{S}_{K,T}(t) \xrightarrow[K \rightarrow +\infty]{a.s.} S(t)$ , de acuerdo a la proposición anterior.

Nuevamente estamos en las condiciones de hipótesis del teorema de la convergencia dominada de Lebesgue, a partir de lo que podemos concluir que:

$$\int_0^T \widehat{S}_{K,T}(t) dt \xrightarrow[K \rightarrow +\infty]{a.s.} \int_0^T S(t) dt$$

Es a partir de las proposiciones antes expuestas que podemos ver la bondad del ajuste realizado, incluso a pesar de la introducción de un gran volumen de observaciones censuradas y de haber relajado las hipótesis de independencia entre los tiempos de estancia en los distintos estados del sistema, ya que no son necesarias para la aplicación de procesos semi-Markovianos. Esta aplicación se constituye en un antecedente para su implementación en el estudio de muchas otras enfermedades y a otros estudios en diferentes contextos.

---

# Conclusiones

---

## Conclusions

The analysis of the evolution of diseases from experimental data faces the difficulty of working on databases in which information is not as large as we would like and, a part of it, often derives in censored data. In this sense, we perform two approaches to analyze the evolution of the bladder cancer: a multidimensional Markov process and a semi-Markov process.

In general, and this is the case, the sojourn times in states are not exponential, and the Markov property is not satisfied. This difficulty has been avoided in the first model considering phase-type distributions as the basic tool in the study of sojourn times, that are approximated by these distributions, and the Markovian structure of the stochastic process governing the system is preserved. In this way, the number of exponential phases enlarges the orders of the involved matrices. On the other hand, the use of phase-type distributions in Markovian processes enables a good approximation since they are a set dense in the set of distributions defined in the positive half real line, with the advantage of using the matrix-analytic methods. Moreover, the expressions are presented in an algorithmic form for being computational implemented. The fits between the empirical and estimated involved functions have been right.

The study is completed when covariates are considered. For diseases with successive recurrences the Cox model is not applicable, a more flexible model is the state-space one, characterized by the identification of different levels of the disease over time (states). In fact, one of the main contributions of this research is the construction of a stochastic model using the matrix-analytic methods to study the evolution of the bladder cancer under the effects of certain covariates and comparing the survival in different risk groups. The covariates are selected from the database by taking into account the evolution of the different group of patients. They are introduced in the system similarly to the Cox model, but considering the matrix structure of the process. The survival measures are calculated and compared with the corresponding empirical ones. The parameters of the covariates are estimated by using the Nelder-Mead algorithm to optimize the partial likelihood, which allows us to enter censored data. This is carried out extending the parameter calculations to the matrix functions.

A different methodology is performed via the semi-Markov process. In the Markov processes, the future evolution of the process only depends on the current state. Applying them requires verifying this hypothesis as well as we have done it. The semi-Markov processes relax this hypothesis assuming that the future evolution depends on both the occupied state and the time elapsed since the last transition. Furthermore, the sojourn time in each state can have an arbitrary distribution. The semi-Markov model that we have created for the disease is suitable in this case and enables the estimation of its kernels, that is, the conditional distribution function in each transient state until the process jumps to another one and from that, other interesting measures of survival. Also, each estimation has involved procedures that have been generalized and they will be able to be used in other applications.

The model we have developed by a semi-Markov process shows a remarkable goodness of fit to the corresponding real process as a consequence of its properties of strong consistency in spite of that and its relaxed assumptions. These ones, are very important for the applicability of the models in other medical contexts, so this justifies our extension.

From all models constructed, relevant achievements have been the estimation of the main survival measures which provide useful information to medical experts for the creation of new treatments and preventive actions, specially when the survival time is lengthy, as in this disease.

An immediate way to continue this first stage of research is the study of others properties of the estimators that we used, such as asymptotic convergences. These desirable properties would allow us to make parametric inferences, as the estimation through of confidence intervals.

These methodologies we have proposed will be able to apply to groups of patients with different diagnosis or treatments. It could be specially interesting to measure the effectiveness of treatments to compare them, in order to improve the wellness of patients. It is our interest to continue the study incorporating these new elements to the state-space model and to perform a deeper study of cancer. The procedure can be appropriate also in other diseases.

We have seen in biomedical literature that the costs deserve special attention. The treatment of diseases is an important point for governments, patients, and society. Particularly, bladder cancer has a long treatment. We have presented how the costs can be estimated and applied to data, when it can be available. However, this approximation can be improved too by a study of the specific costs in the health care.

## **Conclusiones**

El análisis de la evolución de enfermedades a partir de datos experimentales se enfrenta a la dificultad de trabajar con bases de datos no tan grandes como se desea y con una elevada cantidad de datos censurados. En este trabajo hemos desarrollado distintos

modelos para analizar la evolución del cáncer de vejiga, a partir de un proceso de Markov multidimensional y procesos semi-Markovianos.

En general, y en este caso, los tiempos de permanencia en los estados del sistema no son exponenciales y la propiedad de Markov no se satisface. Esta dificultad se ha subsanado en el primer modelo considerando distribuciones tipo-fase como la herramienta básica. Los tiempos de estancia en cada estado y en el propio sistema son aproximados por estas distribuciones y la estructura Markoviana es la que lo gobierna. Aunque el número de fases exponenciales amplía los órdenes de las matrices involucradas no es un problema, dada la versatilidad del método algorítmico. El uso de distribuciones tipo-fase en los procesos de Markov permite buenas estimaciones, ya que constituyen un conjunto denso en el de las distribuciones definidas en la recta real positiva, con la ventaja de permitirnos utilizar los métodos analítico-matriciales. En este sentido, las expresiones se han presentado en forma algorítmica haciendo posible su implementación computacional. La bondad de los ajustes entre las medidas empíricas y las funciones estimadas involucradas, ha sido adecuada.

El estudio se ha completado a través de la inclusión de covariables. Para enfermedades con recurrencias sucesivas el modelo de Cox no es aplicable. Los modelos basados en un espacio multi-estado resultan ser suficientemente flexibles para adaptarse a cada caso, requiriéndose la identificación de diferentes etapas de la enfermedad a lo largo del tiempo. Uno de los principales aportes de esta investigación ha sido la construcción de un modelo estocástico utilizando los métodos analíticos-matriciales con el objetivo de estudiar, a modo de ejemplo, la evolución del cáncer de vejiga bajo los efectos de ciertas covariables y comparar la supervivencia en diferentes grupos de riesgo. Las covariables han sido seleccionadas a partir de la base de datos teniendo en cuenta la evolución de diferentes grupos de pacientes. Se han introducido en el sistema en forma similar a como se hace en el modelo de Cox, pero considerando la estructura matricial del proceso. Las medidas de supervivencia una vez estimadas, son comparadas con las empíricas correspondientes.

Los parámetros de las covariables se han estimado utilizando el algoritmo de Nelder-Mead para optimizar la verosimilitud parcial, lo que ha permitido ingresar datos censurados. Esto ha sido llevado a cabo ampliando los cálculos de los parámetros a partir de funciones matriciales.

Una metodología diferente ha sido aplicada con un modelo que involucra procesos semi-Markovianos. En los procesos de Markov, la evolución futura del proceso sólo depende del estado actual. Aplicarlos requiere verificar esta hipótesis así como lo hemos hecho en el primer modelo. Los procesos semi-Markovianos relajan esta hipótesis asumiendo que su evolución futura depende tanto del estado ocupado en cierto instante, como del tiempo transcurrido desde la última transición. Además, el tiempo de estancia en cada estado puede tener una distribución arbitraria. El modelo semi-Markoviano que hemos creado para esta enfermedad ha evidenciado ser adecuado y ha permitido la estimación de sus núcleos, es decir, de las funciones de distribución condicionales en cada estado transitorio hasta que el proceso salte desde ese a otro estado, así como otras medidas interesantes de supervivencia. Cada estimación ha involucrado procedimientos que se han generalizado y podrán utilizarse en otras aplicaciones. Se ha evidenciado una notable bondad de ajuste con el correspondiente proceso real, como consecuencia de sus propiedades de fuerte consistencia, a pesar de sus supuestos iniciales no demasiado estrictos. Esta característica es importante para la aplicabilidad de estos modelos en otros contextos médicos, justificando ampliamente nuestra extensión.

Cabe destacar que a partir de la estimación de las principales medidas de supervivencia en los modelos propuestos, se ha proporcionado información útil a los expertos médicos para la creación de nuevos tratamientos y acciones preventivas; especialmente importante cuando el tiempo de supervivencia es largo como en esta enfermedad, siendo este el objetivo principal de nuestro trabajo.

Una forma inmediata de continuar esta primera etapa de investigación es a partir del estudio de otras propiedades de los estimadores que se han involucrado, así como las convergencias asintóticas. Estas propiedades deseables nos permitirán hacer inferencias paramétricas, así como la estimación a través de intervalos de confianza.

La metodología que hemos propuesto se podrá aplicar a grupos de pacientes con diferentes diagnósticos o tratamientos. Podría ser especialmente interesante para medir la eficacia de los tratamientos, con el fin de mejorar el bienestar de los pacientes. Es de nuestro interés continuar el estudio incorporando estos nuevos elementos en el modelo espacio-estados para realizar un estudio más profundo del cáncer. Sin embargo, la posible ampliación del modelo construido con un número mayor de covariables es evidente, a partir de una base de datos que nos lo permita. El procedimiento puede ser apropiado también en otras enfermedades.

Hemos visto en la literatura biomédica que los costos merecen especial atención. El tratamiento de enfermedades es un punto importante para los gobiernos, los pacientes y sociedad en general. Particularmente, el cáncer de vejiga tiene un tratamiento prolongado. En este estudio se ha presentado cómo estimar los costos a partir de los datos, cuando estén disponibles. Sin embargo, esta aproximación también puede mejorarse mediante un estudio actualizado de los costes específicos en la asistencia sanitaria.

---

## Apéndice I

---

Este apéndice contiene las ejecuciones en *SPSS* para completar los valores de las variables incorporadas a la base original.



```
GET
FILE='C:\UGR\ANTECEDENTES-UPV\Copia de TVStimes.sav'.
DATASET NAME Conjunto_de_datos1 WINDOW=FRONT.

SAVE OUTFILE='C:\UGR\ANTECEDENTES-UPV\Copia de TVStimes-CON VARIABLES NUEVAS Y
CORREGIDO.sav'
/COMPRESSED.
SAVE OUTFILE='C:\UGR\ANTECEDENTES-UPV\Copia de TVStimes-CON VARIABLES NUEVAS Y
CORREGIDO.sav'
/COMPRESSED.
IF (enum=3) estado_ant=2.
EXECUTE.
IF (enum=3 & status=0 & status.prog=0) estado_act=5.
EXECUTE.
IF (enum=3 & status=0 & status.prog=0) estado_sig=5.
EXECUTE.
IF (enum=3 & status=1 & status.prog=1) estado_sig=4.
EXECUTE.
IF (enum=3 & status=1 & status.prog=1) estado_act=4.
EXECUTE.
IF (enum=3 & status=1 & status.prog=0) estado_act=3.
EXECUTE.
IF (enum=3 & status=1 & status.prog=0) estado_sig=3.
EXECUTE.
```

```
IF (enum>3) estado_ant=3.
EXECUTE.
IF (enum>3) estado_act=3.
EXECUTE.
IF (enum>3) estado_sig=3.
EXECUTE.
IF (enum<=2) estado_ant=enum-1.
EXECUTE.
IF (enum<=2 & status=0 & status.prog=0) estado_act=5.
EXECUTE.
IF (enum<=2 & status=0 & status.prog=0) estado_sig=5.
EXECUTE.
IF (enum<=2 & status=1 & status.prog=1) estado_act=4.
EXECUTE.
IF (enum<=2 & status=1 & status.prog=1) estado_sig=4.
EXECUTE.
IF (enum<=2 & status=1 & status.prog=0) estado_act=enum.
EXECUTE.
DESCRIPTIVES VARIABLES=estado_sig
/STATISTICS=MEAN STDDEV MIN MAX.
```

---

## Apéndice II

---

Este apéndice contiene las ejecuciones en *SPSS* para analizar la correlación entre los tiempos de permanencia en los estados.

Se presenta a continuación, en la figura .7 las ejecuciones de los filtrados correspondientes para obtener los valores de las variables *tiempo de permanencia en 0* y *tiempo de permanencia en 1*, a partir de la base original de datos.

```
DATASET COPY para_correl_tiempos_en_0_y_1.
DATASET ACTIVATE para_correl_tiempos_en_0_y_1.
FILTER OFF.
USE ALL.

SELECT IF (estado_ant=0 & estado_act=1 & (estado_sig=2 |
estado_sig=4)).

EXECUTE.
DATASET ACTIVATE Conjunto_de_datos1.
DATASET ACTIVATE para_correl_tiempos_en_0_y_1.

SAVE OUTFILE='F:\Alicia\UGR\ANTECEDENTES-
UPV\para_correlacion_tiempos_en_0.sav'
/COMPRESSED.
DATASET ACTIVATE Conjunto_de_datos1.
DATASET CLOSE para_correl_tiempos_en_0_y_1.
DATASET COPY para_correl_tiempos_en_1.
DATASET ACTIVATE para_correl_tiempos_en_1.
FILTER OFF.
USE ALL.

SELECT IF ((estado_ant=1 & estado_act=2 & (estado_sig=3 | estado_sig=4
| estado_sig=5)) | (estado_ant=1 & estado_act=4 & estado_sig=4)).

EXECUTE.
DATASET ACTIVATE Conjunto_de_datos1.
DATASET ACTIVATE para_correl_tiempos_en_1.

SAVE OUTFILE='F:\Alicia\UGR\ANTECEDENTES-
UPV\para_correlacion_tiempos_en_1.sav'
/COMPRESSED.
DATASET ACTIVATE Conjunto_de_datos1.
DATASET CLOSE para_correl_tiempos_en_1.
```

Figura 7. Procedimiento de filtrado para obtener los valores de las variables tiempo de permanencia en 0 y tiempo de permanencia en 1.

Se presenta a continuación, en las figuras 8 y 9, las ejecuciones de los filtrados en SPSS para obtener los valores de las variables *tiempo de permanencia en 1* y *tiempo de permanencia en 2*, a partir de la base original de datos.

```
DATASET COPY para_correl_tiempos_1_2_tiempos_en_1.
DATASET ACTIVATE para_correl_tiempos_1_2_tiempos_en_1.
FILTER OFF.
USE ALL.

SELECT IF (estado_ant=1 & estado_act=2 & (estado_sig=3 |
estado_sig=4)).
EXECUTE.
DATASET ACTIVATE Conjunto_de_datos1.
```

Figura 8. Procedimiento de filtrado para obtener los valores de la variable *tiempo de permanencia en 1*, con el objetivo de analizar la correlación con la variable *tiempo de permanencia en 2*.

```
DATASET COPY para_correl_tiempos_1_2_tiempos_en_2.
DATASET ACTIVATE para_correl_tiempos_1_2_tiempos_en_2.
FILTER OFF.
USE ALL.

SELECT IF ((estado_ant=2 & estado_act=3) | (estado_ant=2 &
estado_act=4)).
EXECUTE.
DATASET ACTIVATE Conjunto_de_datos1.

DATASET ACTIVATE para_correl_tiempos_1_2_tiempos_en_2.

SAVE OUTFILE='F:\Alicia\UGR\ANTECEDENTES-
UPV\para_correlacion_tiempos_1_2_tiempos_en_2.sav'
/COMPRESSED.

DATASET ACTIVATE Conjunto_de_datos1.
DATASET CLOSE para_correl_tiempos_1_2_tiempos_en_2.
```

Figura .9. Procedimiento de filtrado para obtener los valores de la variable *tiempo de permanencia en 2*, con el objetivo de analizar la correlación con la variable *tiempo de permanencia en 1*.



---

## Apéndice III

---

Este apéndice contiene el algoritmo a ejecutar en Matlab, contenido en el archivo PHplot.m, modificado.

```

% PHplot.m %
clear;
n=460; %CAMBIADO%
k=1.1;
l=1;
load phases;
load inputdistr;
load sample; %AGREGADO%
a=phases(:,1)';
p=length(a);
T=phases(:, 2:p+1);

t = -T*ones(size(a'));
vv = -a*inv(T)*ones(size(a'));
var = 2*a*inv(T*T)*ones(size(a'))-vv*vv;
std=sqrt(var);
truncpoint = vv+4*sqrt(var);
dt = truncpoint/n;
for i=1:n+1,
    y(i)=dt*(i-1);
    x(i)= sample(i,2) ; %AGREGADO%
    F(i)=1-a*expm(T*y(i))*ones(size(a'));
    S(i)=1-F(i);
    f(i)=a*expm(T*y(i))*t;
    r(i)=f(i)/S(i);
    F_teorica(i)=1-a*expm(T*x(i))*ones(size(a')); %AGREGADO%
    f_teorica(i)=a*expm(T*x(i))*t; %AGREGADO%

```

```

S_teorica(i)=1-F_teorica(i); %AGREGADO%
r_teorica(i)=f_teorica(i)/S_teorica(i); %AGREGADO%
end
%-----%

F_emp(1)=1/(n+1); %AGREGADO%
S_emp(1)=1-1/(n+1);
for i=2:n+1,
    F_emp(i)=1/(n+1)+F_emp(i-1);
    S_emp(i)=1-F_emp(i);
end

%-----%
for i=1:length(inputdistr(:,1)),
    x1(i)=inputdistr(i,1);
    F1(i)=inputdistr(i,2);
    S1(i)=1-F1(i);
    f1(i)=inputdistr(i,3);
    r1(i)=inputdistr(i,4);
end
%-----%
while 1
    disp(' ');
    disp('1. Mean and standard-deviation');
    disp('2. Display survival function');
    disp('3. Display distribution function');
    disp('4. Display density');
    disp('5. Display failure rate');
    disp('6. Print graph');

```

```

disp('7. Save graph on file "PHgraph"');
disp('8. Load new estimates of pi and T');
disp('9. Values of distribution function');%AGREGADO%
disp('10. Display fitted PH distribution and empirical distribution');
disp('11. Display fitted survival function and empirical survival');
disp('12. Quit');
q=input('Select (1-12): ');

if q==1
    disp(' ');
    disp('The fitted PH-distribution has mean ='), disp(vv);
    disp('and standard-deviation ='), disp(std);
end
if q==2
    plot(y,S,'--',x1,S1,'-');
    title('Survival function');
    xlabel('- input, - - fitted PH');
    ylabel(' ');
    set(gca,'xlim',[0,y(n)],[0,1]),'ylim',[0,1]);
end
if q==3
    plot(y,F,'--',x1,F1,'-');
    title('Distribution function');
    xlabel('- input, - - fitted PH');
    ylabel(' ');
    set(gca,'xlim',[0,y(n)],[0,1]),'ylim',[0,1]);
end
if q==4
    plot(x,f_teorica,'--');
    title('Density');

```

```
xlabel(' - - fitted PH');
ylabel('f');
set(gca,'xlim',[0,x(n+1)],'ylim',[0,max(f_teorica)]);
end
if q==5
    plot(x,r_teorica,'--');
    title('Failure rate');
    xlabel(' - - fitted PH');
    ylabel('h');
    set(gca,'xlim',[0,x(n+1)],'ylim',[0,max(r_teorica)]);
end
if q==6
    print;
end
if q==7
    print PHgraph -dps;
end
if q==8
    clear a T t;
    clear y F S f r;
    load phases;
    a=phases(:,1)';
    p=length(a);
    T=phases(:, 2:p+1);

    t = -T*ones(size(a));
    vv = -a*inv(T)*ones(size(a));
    var = 2*a*inv(T*T)*ones(size(a))-vv*vv;
    std=sqrt(var);
    truncpoint = vv+4*sqrt(var);
```

```

dt = truncpoint/n;
for i=1:n+1
    y(i)=dt*(i-1);
    F(i)=1-a*expm(T*y(i))*ones(size(a'));
    S(i)=1-F(i);
    f(i)=a*expm(T*y(i))*t;
    r(i)=f(i)/S(i);
end
end
if q==9 %AGREGADO PARA CALCULAR EL ESTADÍSTICO DEL CONTRASTE DE KOLMOGOROV-SMIRNOV%
    disp(' ');
    disp('The distribution function values are'),
    F=F_teorica';
    disp(F);
end

if q==10 %AGREGADO PARA OBTENER EL GRÁFICO DE AMBAS FUNCIONES EN EL MISMO SISTEMA DE COORDENADAS%
    plot(x,F_teorica,'--',x,F_emp,'-');
    title('Fitted PH distribution and empirical distribution');
    xlabel('-- input sample, - - fitted PH');
    ylabel(' F ');
    set(gca,'xlim',[0,x(n+1)],'ylim',[0,1]);
end

if q==11 %AGREGADO PARA OBTENER EL GRÁFICO DE AMBAS FUNCIONES EN EL MISMO SISTEMA DE COORDENADAS%
    plot(x,S_teorica,'--',x,S_emp,'-');
    title(' Fitted Survival function and empirical survival');
    xlabel('-- input sample, - - fitted PH');
    ylabel('S');
    set(gca,'xlim',[0,x(n+1)],'ylim',[0,1]);

```

```
end
    if q==12 %CAMBIADO%
        l=0;
    end
end
end
```



---

## Apéndice IV

---

Este apéndice contiene:

- El procedimiento para el ingreso de los tiempos de supervivencia en cada estado, codificados según censura, los correspondientes niveles de la covariable de interés y la construcción de la función log-verosimilitud parcial.
- la adaptación y aplicación del algoritmo de Nelder Mead, utilizado en la optimización de las funciones log-verosimilitudes parciales, para el modelo multiestado con covariables. Se ha utilizado lenguaje Wolfram y ha sido ejecutado para obtener la matriz de transiciones entre estados  $Q(Z)$  en Mathematica.

**Creación de la función log-verosimilitud parcial a partir de los valores de tiempos de supervivencia y/o censurados por derecha** Ejemplo para el estado  $i = 0$ :

```
deathtimes0 = Table[0, {122}];
death0 = Table[0, {122}];
risk0 = Table[0, {122}];
s0 = Table[0, {122}];
h0 = 0;
r0 = 122;
aux0 = 0;
For[k = 1, k <= 122, k++, {
  If[code0[[k]] == 1, deathtimes0[[1]] = times0[[k]]; death0[[1]] = 1;
    s0[[1]] = z0[[k]]; risk0[[1]] = r0; r0 = r0 - 1; h0 = h0 + 1;
    aux0 = k; Break[], r0 = r0 - 1];
}];
```

```

For[j = (aux0 + 1), j <= 122, j++, {
  If[code0[[j]] == 1 && times0[[j]] == deathtimes0[[h0]], r0 = r0 - 1;
    death0[[h0]] = death0[[h0]] + 1; s0[[h0]] = s0[[h0]] + z0[[j]];
  If[code0[[j]] == 1 && times0[[j]] > deathtimes0[[h0]], h0 = h0 + 1;
    risk0[[h0]] = r0; r0 = r0 - 1; deathtimes0[[h0]] = times0[[j]];
    death0[[h0]] = 1; s0[[h0]] = z0[[j]];
  If[ code0[[j]] == 0, r0 = r0 - 1];
}];
numeradores0 = Table[0, {30}];
denominadores0 = Table[0, {30}];
For[k = 1, k <= 30, k++, {
  numeradores0[[k]] = Exp[b*s0[[k]]];
}];
For[k = 1, k <= 30, k++, {
  sumaux = 0;
  For[j = (122 - risk0[[k]] + 1), j <= 122 , j++, {
    sumaux = sumaux + Exp[b*z0[[j]]];
  }];
  denominadores0[[k]] = sumaux^(death0[[k]]);}];
verosim0 = numeradores0[[1]]/denominadores0[[1]];
For[k = 2, k <= 30, k++, {
  verosim0 = verosim0*numeradores0[[k]]/denominadores0[[k]];
}];
logverosim0[b_] = Log[verosim0];
f[b_] = logverosim0[b];

```

### Adaptación del algoritmo de Nelder-Mead:

```

n := 1;
p := 1;
x := 2;
g := 0.5;
s := 0.5;
X = {-2, 2};
Y = {f[X[[1]]], f[X[[2]]]};

For[k = 1, k <= 1000, k++, {
  For[h = 1, h <= n, h++, {
    For[i = 1, i <= h, i++, {

```

```

If[Y[[h + 1]] > Y[[i]],
  Y2 = Table[0, {j, (h - i + 1)}];
  X2 = Table[0, {j, (h - i + 1)}];
  For[j = 1, j <= (h - i + 1), j++, {
    Y2[[j]] = Y[[i + j - 1]];
    X2[[j]] = X[[i + j - 1]];
  }];
  Y[[i]] = Y[[h + 1]];
  X[[i]] = X[[h + 1]];
  For[k = 1, k <= (h - i + 1), k++, {
    Y[[i + k]] = Y2[[k]];
    X[[i + k]] = X2[[k]];
  }];
  Break[]
]
}]
}];

xbar = (1/n)*Sum[X[[i]], {i, 1, n, 1}];
xrefle = xbar + p*(xbar - X[[2]]);
yrefle = f[xrefle];
max = f[X[[1]]];
min = f[X[[2]]];
xexpan = xbar + x*(xrefle - xbar);
yexpan = f[xexpan];
xconf = xbar + g*(xrefle - xbar);
yconf = f[xconf];
xcond = xbar + g*(X[[2]] - xbar);
ycond = f[xcond];
xencog = X[[1]] + s*(X[[2]] - X[[1]]);
yencog = f[xencog];
If[(yrefle > max) && (yexpan > yrefle), X[[2]] = xexpan];
If[(yrefle > max) && (yexpan <= yrefle), X[[2]] = xrefle];
If[(yrefle <= max) && (yrefle > min) && (yconf > min),
  X[[2]] = xconf];
If[(yrefle <= max) && (yrefle > min) && (yconf <= min),
  X[[2]] = xencog];
If[(yrefle <= max) && (yrefle <= min) && (ycond > min),
  X[[2]] = xcond];
If[(yrefle <= max) && (yrefle <= min) && (ycond <= min),
  X[[2]] = xencog];
Y[[2]] = f[X[[2]]];
ybar = (1/(n + 1))*Sum[Y[[i]], {i, 1, (n + 1), 1}];

```

```
mce = Sqrt[(1/(n + 1))*Sum[(Y[[i]] - ybar)^2, {i, 1, (n + 1), 1}]];
If[mce < 1*10^(-8), Print["El número de iteraciones es k= ", k];
  Print["El error cuadrático medio es mce= ", mce];
  Print["El valor estimado de b es b= ", X[[1]]]; Break[]];
};
```

---

## Apéndice V

---

En este apéndice, se muestra a través de un ejemplo (para el estado transitorio 0) los procedimientos para ingresar los vectores de tiempos de supervivencia de los pacientes en cada estado transitorio, su codificación de censura y para calcular automáticamente paso a paso, los números de sujetos en riesgo y las frecuencias de "fallos". Los algoritmos que se detallan, han permitido en primer lugar construir un estimador de la función de distribución para el tiempo de permanencia en cada estado transitorio, a partir del estimador de Kaplan-Meier. Luego, se han calculado las estimaciones de los *Núcleos*, de la función de *Supervivencia* y del *Tiempo medio hasta la progresión de la enfermedad* hasta  $T$ , tiempo durante el que se efectúa el seguimiento de los pacientes.

### **Cálculos iniciales a partir del ingreso de las observaciones:**

Los tiempos de fallo se han ordenado en el vector *times0* con sus correspondientes códigos de censura en *code0*, sin cuantificar empates. Luego, el vector *deathtimes0* que contiene los tiempos reales de ocurrencia del suceso recidiva o muerte, con sus correspondientes frecuencias absolutas en *death0* y el vector con los correspondientes números de pacientes en riesgo justo antes de cada instante de ocurrencia del suceso *risk0*, han sido calculados a través de los siguientes procedimientos:

```
deathtimes0 = Table[0, {847}];
```

```
death0 = Table[0, {847}];
```

```

risk0 = Table[0, {847}];
h0 = 0;
r0 = 847;
aux0 = 0;
For[k = 1, k <= 847, k++, {
  If[code0[[k]] == 1, deathtimes0[[1]] = times0[[k]]; death0[[1]] = 1;
    risk0[[1]] = r0; r0 = r0 - 1; h0 = h0 + 1; aux0 = k; Break[],
    r0 = r0 - 1];
}]
For[j = (aux0 + 1), j <= 847, j++, {
  If[code0[[j]] == 1 && times0[[j]] == deathtimes0[[h0]], r0 = r0 - 1;
    death0[[h0]] = death0[[h0]] + 1];
  If[code0[[j]] == 1 && times0[[j]] > deathtimes0[[h0]], h0 = h0 + 1;
    risk0[[h0]] = r0; r0 = r0 - 1; deathtimes0[[h0]] = times0[[j]];
    death0[[h0]] = 1];
  If[ code0[[j]] == 0, r0 = r0 - 1];
}]

```

Otro algoritmo, ha permitido obtener las funciones escalonadas  $\widehat{F}_0(t)$ , como paso previo a la obtención de los *Núcleos*  $Q_{0j}$ , donde  $j \in \{1, 2, 3\}$ . Posteriormente, otros procedimientos que implican la estimación de ciertas convoluciones han permitido estimar la función de Supervivencia.

```

auxfbar0 = Table[1, {282}];
auxfbar0[[1]] = 1 - death0[[1]]/risk0[[1]];
For[k = 2, k <= 282, k++, {

```

```

    auxfbar0[[k]] = auxfbar0[[k - 1]]*(1 - death0[[k]]/risk0[[k]]);
  }];
fbar0[t_] := Module[{k, f}, For[k = 1, k <= 281, k++, {
  If[deathtimes0[[k]] <= t && t < deathtimes0[[k + 1]] ,
    f = auxfbar0[[k]]];
}];
If[0 <= t && t < deathtimes0[[1]] , f = 1];
If[deathtimes0[[282]] <= t, f = auxfbar0[[282]]]; f]

```

### Estimadores de los Núcleos:

```

p01 = 322/847;
p04 = 26/847;
p12 = 175/322;
p14 = 19/322;
p23 = 126/175;
p24 = 4/175;
p34 = 3/126;
ker01[t_] := (1 - fbar0[t])*p01;
ker04[t_] := (1 - fbar0[t])*p04;
ker12[t_] := (1 - fbar1[t])*p12;
ker14[t_] := (1 - fbar1[t])*p14;
ker23[t_] := (1 - fbar2[t])*p23;
ker24[t_] := (1 - fbar2[t])*p24;
ker34[t_] := (1 - fbar3[t])*p34;

```

### Convoluciones:

```
f1[t_] = ker01[t];
f2[t_] = ker14[t];
s1[t_] = 0;
For[k = 1, k <= 100, k++, {
  s1[t_] =
    s1[t] + f1[t - (k*t/100)]*(f2[k*t/100] - f2[(k - 1)*t/100]);
}];
f3[t_] = ker01[t];
f4[t_] = ker12[t];
s2[t_] = 0;
For[k = 1, k <= 100, k++, {
  s2[t_] =
    s2[t] + f3[t - (k*t/100)]*(f4[k*t/100] - f4[(k - 1)*t/100]);
}];
f5[t_] = s2[t];
f6[t_] = ker24[t];
s3[t_] = 0;
For[k = 1, k <= 100, k++, {
  s3[t_] =
    s3[t] + f5[t - (k*t/100)]*(f6[k*t/100] - f6[(k - 1)*t/100]);
}];
f7[t_] = ker23[t];
f8[t_] = ker34[t];
s4[t_] = 0;
For[k = 1, k <= 100, k++, {
  s4[t_] =
```

```

    s4[t] + f7[t - (k*t/100)]*(f8[k*t/100] - f8[(k - 1)*t/100]);
  }];
f9[t_] = s2[t];
f10[t_] = s4[t];
s5[t_] = 0;
For[k = 1, k <= 100, k++, {
  s5[t_] =
    s5[t] + f9[t - (k*t/100)]*(f10[k*t/100] - f10[(k - 1)*t/100]);
}];

```

#### **Estimación de la función de Supervivencia:**

```
stheoretical[t_] := 1 - (ker04[t] + s1[t] + s3[t] + s5[t]);
```

#### **Estimación del Tiempo Medio de progresión de la enfermedad (MTTP) en $[0, T]$ :**

```

area = 0;
T = 5000;
delta = 0;
For[i = 1, i <= 10000, i++, {
  aux = 0;
  For[j = 1, j <= i, j++, {
    aux = aux + stheoretical[(j*(T/i) + (j - 1)*(T/i))*1/2]*(T/i);
  }];
  delta = Abs[aux - area];
  area = aux;
  If[delta < 1*10^(-5), Print["The step number is i= ", i];
}];

```

```
Print["The diference between the last two terms is delta= ",  
delta]; Print["The estimation of MTTP, by integrate of S(t) in [0,T] is  
MTTP= ", area]; Break[]  
}]
```

---

## Índice de figuras

---

1.1	Diagrama del modelo que representa en forma completa las etapas de la enfermedad observadas en $n=847$ pacientes, con dos estados absorbentes. . .	24
1.2	Sistema multi-estado que representa las distintas etapas de la enfermedad de los $n = 122$ pacientes. . . . .	27
1.3	Diagrama de transiciones entre los estados del sistema. . . . .	30
2.1	Diagrama del modelo que representa en forma completa las etapas de la enfermedad observadas en $n=847$ pacientes, con dos estados absorbentes. . .	36
2.2	Histograma correspondiente a la variable <i>Tiempo de permanencia en 0</i> , a partir de 847 casos. . . . .	38
2.3	Histograma correspondiente a la variable <i>Tiempo de permanencia en 1</i> , a partir de 322 casos. . . . .	39
2.4	Histograma correspondiente a la variable <i>Tiempo de permanencia en 2</i> , a partir de 175 casos. . . . .	39
2.5	Histograma correspondiente a la variable <i>Tiempo de permanencia en 3</i> , a partir de 126 casos. . . . .	40
2.6	Análisis paramétrico de correlación entre las variables <i>Tiempo de permanencia en 0</i> y <i>Tiempo de permanencia en 1</i> , a partir de 322 casos. . . . .	41
2.7	Análisis no paramétrico de correlación entre las variables <i>Tiempo de permanencia en 0</i> y <i>Tiempo de permanencia en 1</i> , a partir de 322 casos. . .	42
2.8	Análisis paramétrico de correlación entre las variables <i>Tiempo de permanencia en 1</i> y <i>Tiempo de permanencia en 2</i> , a partir de 175 casos. . . . .	42

2.9	Análisis no paramétrico de correlación entre las variables <i>Tiempo de permanencia en 1</i> y <i>Tiempo de permanencia en 2</i> , a partir de 175 casos. . .	43
2.10	Análisis paramétrico de correlación entre las variables <i>Tiempo de permanencia en 2</i> y <i>Tiempo de permanencia en 3</i> , a partir de 126 casos. . . . .	43
2.11	Análisis no paramétrico de correlación entre las variables <i>Tiempo de permanencia en 2</i> y <i>Tiempo de permanencia en 3</i> , a partir de 126 casos. . .	44
2.12	Pruebas de bondad de ajuste para la distribución de la variable <i>Tiempo de permanencia en el estado 0</i> . . . . .	45
2.13	Pruebas de bondad de ajuste para la distribución de la variable <i>Tiempo de permanencia en el estado 1</i> . . . . .	45
2.14	Pruebas de bondad de ajuste para la distribución de la variable <i>Tiempo de permanencia en el estado 2</i> . . . . .	46
2.15	Pruebas de bondad de ajuste para la distribución de la variable <i>Tiempo de permanencia en el estado 3</i> . . . . .	46
2.16	Gráfico de la función de distribución tipo fase estimada, para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 847$ casos . . . . .	50
2.17	Gráfico de la función de supervivencia estimada, para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 847$ casos. . . . .	50
2.18	Gráfico de la función densidad estimada, para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 847$ casos. . . . .	51
2.19	Gráfico de la función razón de fallo estimada, para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 847$ casos. . . . .	51
2.20	Gráfico de la función de distribución tipo fase estimada, para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 322$ casos . . . . .	53
2.21	Gráfico de la función de supervivencia estimada, para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 322$ casos. . . . .	53
2.22	Gráfico de la función densidad estimada, para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 322$ casos. . . . .	54
2.23	Gráfico de la función razón de fallo estimada, para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 322$ casos. . . . .	55

2.24	Gráfico de la función de distribución tipo fase estimada, para el <i>Tiempo de permanencia en 2</i> , a partir de la muestra de $n = 175$ casos . . . . .	57
2.25	Gráfico de la función de supervivencia estimada, para el <i>Tiempo de permanencia en 2</i> , a partir de la muestra de $n = 175$ casos. . . . .	57
2.26	Gráfico de la función densidad estimada, para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 175$ casos. . . . .	58
2.27	Gráfico de la función razón de fallo estimada, para el <i>Tiempo de permanencia en 2</i> , a partir de la muestra de $n = 175$ casos. . . . .	58
2.28	Gráfico de la función de distribución tipo fase estimada, para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 126$ casos . . . . .	60
2.29	Gráfico de la función de supervivencia estimada, para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 126$ casos. . . . .	60
2.30	Gráfico de la función densidad estimada, para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 126$ casos. . . . .	61
2.31	Gráfico de la función razón de fallo estimada, para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 126$ casos. . . . .	61
2.32	Gráfico de la función de supervivencia $S(t)$ estimada(- -), en contraste con la supervivencia empírica. . . . .	71
2.33	Gráfico de la función densidad $f(t)$ , para el sistema multiestado. . . . .	72
2.34	Gráfico de la función razón de fallo $h(t)$ , para el sistema multiestado. . . . .	72
2.35	Representación gráfica de los tiempos medios muestrales de permanencia en el estado transitorio 0, para cada intervalo $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante $t = 0$ , en comparación con los tiempos medios estimados (- -). . . . .	79
2.36	Representación gráfica de los tiempos medios muestrales de permanencia en el estado transitorio 1, para cada intervalo $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante $t = 0$ , en comparación con los tiempos medios estimados (- -). . . . .	80

2.37	Gráfico de los tiempos medios muestrales de permanencia en el estado transitorio 2, para cada intervalo $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante $t = 0$ , en comparación con los tiempos medios estimados (- -). . . . .	80
2.38	Gráfico de los tiempos medios muestrales de permanencia en el estado transitorio 3, para cada intervalo $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante $t = 0$ , en comparación con los tiempos medios estimados (- -). . . . .	81
2.39	Gráfico de los tiempos medios muestrales de permanencia en el sistema, para cada intervalo $[0, t]$ , habiendo ingresado al sistema en el estado 0 en el instante $t = 0$ , en comparación con los tiempos medios estimados (- -). . . . .	81
2.40	Diagrama del sistema multiestado, con un estado absorbente P, para los $n=52$ pacientes que derivan en progresión . . . . .	84
2.41	Gráfico de la función de distribución para el sistema multiestado, con un estado absorbente P y $n=52$ pacientes que derivan en progresión . . . . .	85
2.42	Gráfico de la función de densidad para el sistema multiestado, con un estado absorbente P y $n=52$ pacientes que derivan en progresión . . . . .	86
2.43	Gráfico de la función de supervivencia para el sistema multiestado, con un estado absorbente P y $n=52$ pacientes que derivan en progresión . . . . .	86
2.44	Gráfico de la función razón de fallo para el sistema multiestado, con un estado absorbente P y $n=52$ pacientes que derivan en progresión . . . . .	87
2.45	Tiempos medios estimados de permanencia en el sistema multiestado y estado absorbente P (- -), en comparación con los tiempos medios en la muestra de $n=52$ pacientes que derivan en progresión . . . . .	89
2.46	Gráfico de la función de distribución tipo fase estimada para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 52$ casos . . . . .	91
2.47	Gráfico de la función de supervivencia estimada para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 52$ casos. . . . .	91
2.48	Gráfico de la función de distribución tipo fase estimada para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 26$ casos . . . . .	93

2.49	Gráfico de la función de supervivencia estimada para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 26$ casos. . . . .	93
2.50	Gráfico de la función de distribución tipo fase estimada para el <i>Tiempo de permanencia en 2</i> , a partir de la muestra de $n = 7$ casos . . . . .	94
2.51	Gráfico de la función de supervivencia estimada para el <i>Tiempo de permanencia en 2</i> , a partir de la muestra de $n = 7$ casos. . . . .	95
2.52	Gráfico de la función de distribución tipo fase estimada para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 3$ casos . . . . .	96
2.53	Gráfico de la función de supervivencia estimada para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 3$ casos. . . . .	96
2.54	Gráfico de la función de supervivencia estimada en el sistema con estado absorbente P, teniendo en cuenta los tiempos en los estados transitorios (- -) en contraste con la función empírica de supervivencia. . . . .	101
2.55	Gráficos de las funciones de supervivencia estimadas: sin tener en cuenta los tiempos en los estados transitorios (-) y teniéndolos en cuenta (- -). . . .	101
2.56	Tiempos esperados (- -) y tiempos medios muestrales de permanencia en el estado 0 en el intervalo $[0, t]$ , habiendo partido en $t = 0$ del estado 0. Muestra de $n=52$ pacientes que salen del sistema con progresión . . . . .	103
2.57	Diagrama del sistema multiestado con un estado absorbente NP, para los $n=795$ pacientes que salen del sistema sin progresión . . . . .	105
2.58	Gráfico de la función de distribución para el sistema multiestado con un estado absorbente NP y $n=795$ pacientes que salen del sistema sin progresión	107
2.59	Gráfico de la función de densidad para el sistema multiestado, con un estado absorbente NP y $n=795$ pacientes que salen del sistema sin progresión . . .	108
2.60	Gráfico de la función de supervivencia para el sistema multiestado, con un estado absorbente NP y $n=795$ pacientes que salen del sistema sin progresión	108
2.61	Gráfico de la función razón de fallo para el sistema multiestado, con un estado absorbente NP y $n=795$ pacientes que salen del sistema sin progresión	109

2.62	Tiempos medios de permanencia estimados (- -) en el sistema con estado absorbente NP, en comparación con los tiempos medios en la muestra de $n=795$ pacientes que salen del sistema sin progresión . . . . .	111
2.63	Gráfico de la función de distribución tipo fase estimada para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 795$ casos . . . . .	113
2.64	Gráfico de la función de supervivencia estimada para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 795$ casos. . . . .	113
2.65	Gráfico de la función densidad tipo fase estimada para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 795$ casos . . . . .	114
2.66	Gráfico de la función razón de fallo estimada para el <i>Tiempo de permanencia en 0</i> , a partir de la muestra de $n = 795$ casos. . . . .	114
2.67	Gráfico de la función de distribución tipo fase estimada para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 296$ casos . . . . .	116
2.68	Gráfico de la función de supervivencia estimada para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 296$ casos. . . . .	116
2.69	Gráfico de la función densidad tipo fase estimada para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 296$ casos . . . . .	117
2.70	Gráfico de la función razón de fallo estimada para el <i>Tiempo de permanencia en 1</i> , a partir de la muestra de $n = 296$ casos. . . . .	117
2.71	Gráfico de la función de distribución tipo fase estimada para el <i>Tiempo de permanencia en 2</i> , a partir de la muestra de $n = 168$ casos . . . . .	119
2.72	Gráfico de la función de supervivencia estimada para el <i>Tiempo de permanencia en 2</i> , a partir de la muestra de $n = 168$ casos. . . . .	119
2.73	Gráfico de la función densidad tipo fase estimada para el <i>Tiempo de permanencia en 2</i> , a partir de la muestra de $n = 168$ casos . . . . .	120
2.74	Gráfico de la función razón de fallo estimada para el <i>Tiempo de permanencia en 2</i> , a partir de la muestra de $n = 168$ casos. . . . .	120
2.75	Gráfico de la función de distribución tipo fase estimada para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 123$ casos . . . . .	121

2.76	Gráfico de la función de supervivencia estimada para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 123$ casos. . . . .	122
2.77	Gráfico de la función densidad tipo fase estimada para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 123$ casos . . . . .	122
2.78	Gráfico de la función razón de fallo estimada para el <i>Tiempo de permanencia en 3</i> , a partir de la muestra de $n = 123$ casos. . . . .	123
2.79	Gráficos de las funciones de supervivencia estimadas: sin tener en cuenta los tiempos transitorios y a partir de considerarlos (- -). . . . .	127
2.80	Gráfico de la función de supervivencia estimada (- -) en el sistema con estado absorbente NP, en contraste con la función empírica de supervivencia. . . . .	128
2.81	Tiempos esperados (- -) y tiempos medios muestrales, de permanencia en el estado 0 en el intervalo $[0, t]$ , habiendo partido en $t = 0$ del estado 0. Muestra de $n=795$ pacientes que salen del sistema sin progresión . . . . .	130
3.1	Sistema multi-estado que representa las distintas etapas de la enfermedad de los $n = 122$ pacientes. . . . .	134
3.2	Primera simplificación del sistema multi-estado que representa las distintas etapas de la enfermedad de los $n = 122$ pacientes. . . . .	135
3.3	Diagrama del sistema multi-estado simplificado, para los $n = 122$ pacientes. . . . .	136
3.4	Histograma de frecuencias absolutas para la variable <i>Edad</i> . . . . .	137
3.5	Tabla de frecuencias para la variable <i>Sexo</i> . . . . .	137
3.6	Tabla de frecuencias para la variable <i>Estadío</i> . . . . .	138
3.7	Tabla de contingencia para las variables <i>Situación</i> y <i>Estadío</i> . . . . .	138
3.8	Análisis de correlación para las variables <i>Situación</i> y <i>Estadío</i> . . . . .	139
3.9	Análisis de correlación para las variables <i>Situación</i> y <i>Edad</i> . . . . .	140
3.10	Diagrama del sistema multiestado que representa las distintas etapas de la enfermedad de los $n = 82$ pacientes del grupo de menor riesgo en la covariable <i>Estadío</i> . . . . .	141
3.11	Análisis paramétrico de la correlación entre los tiempos de permanencia en los estados transitorios 0 y 1. . . . .	143

3.12	Análisis no paramétrico de correlación entre los tiempos de permanencia en los estados transitorios 0 y 1. . . . .	144
3.13	Gráfico de la función de supervivencia estimada en el sistema global a partir del algoritmo EM considerando censura y $n = 122$ casos, frente a la que surge de aplicar la estimación por Kaplan-Meier. . . . .	145
3.14	Gráfico de la función de supervivencia estimada en el sistema global, a partir del algoritmo EM considerando censura y $n = 122$ casos, frente a la función empírica de supervivencia. . . . .	145
3.15	Comparación de funciones de supervivencia para ambos grupos de riesgo, en relación a la covariable <i>Estadío</i> . . . . .	152
3.16	Comparación de funciones de supervivencia empírica y estimada para el grupo de riesgo base en relación a la covariable <i>Estadío</i> ( $Z = 0$ ). . . . .	153
3.17	Comparación de funciones de supervivencia empírica y estimada para el grupo de riesgo máximo y en relación a la covariable <i>Estadío</i> ( $Z = 1$ ). . . . .	154
4.1	Diagrama de transiciones entre los estados del sistema. . . . .	156
4.2	Estimaciones de las componentes de $Q(t)$ , en relación al estado absorbente 4. . . . .	163
4.3	Estimaciones de las componentes de $Q(t)$ , en relación a los estados transitorios. . . . .	164
4.4	Representación de la estimación de la función de supervivencia. . . . .	164
.5	Procedimiento automático para completar los valores de las variables incorporadas a la base de datos. . . . .	181
.6	Procedimiento automático para completar los valores de las variables incorporadas a la base de datos. . . . .	182
.7	Procedimiento de filtrado para obtener los valores de las variables <i>tiempo de permanencia en 0</i> y <i>tiempo de permanencia en 1</i> . . . . .	185
.8	Procedimiento de filtrado para obtener los valores de la variable <i>tiempo de permanencia en 1</i> , con el objetivo de analizar la correlación con la variable <i>tiempo de permanencia en 2</i> . . . . .	186

- 
- .9 Procedimiento de filtrado para obtener los valores de la variable *tiempo de permanencia en 2*, con el objetivo de analizar la correlación con la variable *tiempo de permanencia en 1*. . . . . 187



---

## Índice de tablas

---

2.1	Transiciones entre los estados del modelo . . . . .	36
2.2	Descripción estadística de los tiempos en cada estado transitorio. . . . .	37
2.3	Comparativa de valores de $S(t)$ , para los distintos estados transitorios y en el sistema. . . . .	63
2.4	Comparativa de valores de $h(t)$ , para los distintos estados transitorios. . . .	64
2.5	Análisis de la bondad de los ajustes logrados en cada distribución, a partir de la distribución PH especificada. . . . .	65
2.6	Estimación de la supervivencia y razón de fallo, en distintos valores de tiempo. . . . .	73
2.7	Comparación de valores de la supervivencia empírica y estimada, en distintos instantes. . . . .	74
2.8	Tiempos medios de permanencia en cada estado y en el sistema, hasta cierto instante $t$ . . . . .	82
2.9	Descripción estadística de los tiempos en cada estado transitorio. . . . .	84
2.10	Estimación de la supervivencia y razón de fallo, en distintos valores de tiempo. . . . .	88
2.11	Estimación de valores de la función supervivencia en cada estado transitorio y en el sistema con estado absorbente P. . . . .	97
2.12	Bondad de los ajustes en el sistema con estado absorbente P. . . . .	102
2.13	Tiempos medios de permanencia en cada estado y en el sistema con estado absorbente P, hasta cierto instante $t$ . . . . .	104
2.14	Descripción estadística de los tiempos en cada estado transitorio del sistema con estado absorbente NP. . . . .	105

2.15	Estimación de la supervivencia y razón de fallo en distintos valores de tiempo, para el sistema con estado absorbente NP . . . . .	110
2.16	Evaluación de las funcines de <i>supervivencia</i> en los estados transitorios del sistema con estado absorbente NP . . . . .	124
2.17	Evaluación de las funcines <i>razón de fallo</i> en los estados transitorios del sistema con estado absorbente NP . . . . .	124
2.18	Análisis de la bondad de los ajustes, en el sistema con estado absorbente NP.	128
2.19	Tiempos medios de permanencia en cada estado y en el sistema con estado absorbente NP, hasta cierto instante $t$ . . . . .	130
3.1	Transiciones entre los estados del modelo. . . . .	141
3.2	Transiciones entre los estados del modelo para el grupo de riesgo base o mínimo. . . . .	142
3.3	Transiciones entre los estados del modelo para el grupo de riesgo máximo. .	142
3.4	Descripción estadística de los tiempos (en meses) en cada estado transitorio del sistema. . . . .	143
3.5	Estimación de los parámetros a partir de la función <i>Log-verosimilitud parcial</i> .	150
4.1	Estimación de las probabilidades condicionadas de transición . . . . .	163
4.2	Estimación de los valores en la función supervivencia en relación al tiempo (en días, sin partición). . . . .	165

---

## Bibliografía

---

- [1] Aalen, O. O., Cook, R.J. y Roysland, K. (2015). *Does Cox analysis of a randomized survival study yield a causal treatment effect?* . Lifetime Data Analysis. 21, 579-593.
- [2] Amorim, L. D., Cai, J. (2015). *Modelling recurrent events: a tutorial for analysis in epidemiology*. International Journal of Epidemiology. 44, 324-333.
- [3] Andersen, P. K., y Gill, R. D. (1982). *Cox' regression model for counting processes: a large sample study*. Annals of Statistics. 10, 1100-1120.
- [4] Andersen, P. K. y Keiding, N. (2002). *Multi-state models for event history analysis..* Statistical Methods in Medical Research. 11, 91-115.
- [5] Asmussen, S., Nerman, O. y Olsson, M. (1996). *Fitting phase-type distribution via the EM algorithm*. Scandinavian Journal of Probability. 23, 419-441.
- [6] Avendaño Garrido, M. (2013). *Modelos generalizados de riesgos proporcionales para el análisis de supervivencia*. Tesis doctoral dirigida por María Carmen Pardo Llorente. Facultad de Ciencias Matemáticas. Departamento de Estadística e Investigación Operativa I. Universidad Complutense de Madrid.
- [7] Avendaño, M.L. y Pardo, M. (2016). *A semiparametric generalized proportional hazards model for right-censored data*. Annals of the Institute of Statistical Mathematics, Springer. The Institute of Statistical Mathematics.. 68(2), 353-384.

- [8] Barroeta Rojo, C. (2016). *Modelos para el análisis de supervivencia en tiempo discreto: aplicación en el área de veterinaria*. Tesis doctoral dirigida por Anna Espinal Berenguer y Olga Juliá de Ferren. Universidad de Barcelona.
- [9] Borgan, O. (2005). *Kaplan-Meier Estimator*. Encyclopedia of Biostatistics, J. Wiley. Disponible en: <https://doi.org/10.1002/9781118445112.stat06033>
- [10] Botteman, M. F., Pashos, C. L., Redaelli, A. Laskin, B. y Hauser, R. (2003). *The Health Economics of Bladder Cancer*. *Pharmacoeconomics*. 7, 1315-1330.
- [11] Breslow, N. (1974). *Covariance analysis of censored survival data*. *Biometrics*. 30, 89–99.
- [12] Cox, D. R. (1972). *Regression models and life tables (with discussion)*. *Journal of the Royal Statistical Society, Series B*. 34, 187- 220.
- [13] Dumitrescu, M., Gámiz, M.L. y Limnios, N. *Minimum divergence estimators for the Radon-Nikodym derivatives of the Semi-Markov kernel*. *Journal of Theoretical Applied Statistics. Statistics*. 50:3, 486-504, DOI: 10.1080/02331888.2015.1060239.
- [14] Gámiz, M.L., Kulasekera, K.B., Limnios, N. y Lindqvist, B.H. (2011). *Applied Nonparametric Statistics in Reliability*. Springer.
- [15] García, B., Rubio, G. y Santamaría, C. (2005). *A Predictive Mathematical Model in the Recurrence of Bladder Cancer*. *Mathematical and Computer Modelling*. 42, 621-334.
- [16] García-Mora, B., Santamaría, C., Rubio, G. y Pontones, J. L. (2008). *Modeling the recurrence-progression process in bladder carcinoma*. *Computers & Mathematics with Applications*. 56, 619–630.
- [17] Geskus, R. B. (2016). *Data Analysis with Competing Risks and Intermediate States*. Chapman & Hall, Boca Raton.
- [18] Gill, R.D. (1981). *Testing with replacement and the product limit estimator*. *The Annals of Statistics*. Vol. 9, No. 4, 853-860.

- [19] Gill, R.D. (1983). *Large sample behaviour of the product limit estimator on the whole line*. The Annals of Statistics. Vol. 2, No. 1, 49-58.
- [20] Kay, R. (1986). *A Markov model for analysing cancer markers and disease states in survival studies*. Biometrics. 42, 855-865.
- [21] Letón Molina, E. (2005). *Generalización y equivalencias de tests no paramétricos para el análisis de datos censurados*. Madrid: Universidad Complutense de Madrid.
- [22] Limnios, N. y Oprisan, G. (2001). *Semi-Markov Processes and Reliability*. New York, USA: Springer Science+Business Media, LLC.
- [23] Limnios, N. y Ouhbi, B. (2004). *Empirical estimators of reliability and related functions for Semi-Markov systems*. Mathematical and Statistical Methods in Reliability, Cap. 30. Bo Lindqvist and Kjell Doksum editors. [Disponible en [https://www.worldscientific.com/doi/abs/10.1142/9789812795250\\_0030](https://www.worldscientific.com/doi/abs/10.1142/9789812795250_0030)].
- [24] Limnios, N. y Barbu, V. (2008). *Semi-Markov Chains and Hidden Semi-Markov models toward Applications. Their Use in Reliability and DNA Analysis*. New York, USA: Springer Science+Business Media, LLC.
- [25] Limnios, N. (2012). *Reliability Measures of Semi-Markov Systems with General State Space*. Methodology and Computing in Applied Probability. Vol. 14, 895-917.
- [26] Limnios, N. y Girardin, V. (2018). *Apply Probability From Random Sequences to Stochastic Processes*. Cham, Switzerland: Springer Nature Switzerland AG.
- [27] López Bastida, J., et. al. (2002). *Los costes socioeconómicos de las enfermedades cardiovasculares y del cáncer en las Islas Canarias en 1998*. [Disponible en <https://www.sciencedirect.com/science/article/pii/S0213911103717300>].
- [28] Luján-Marco, S. (2012). *Modelización matemática de la multirrecidiva y heterogeneidad individual para el cálculo del riesgo biológico de recidiva y progresión del tumor vesical no músculo invasivo*. Tesis doctoral, dirigida

- por Jiménez-Cruz, F. et.al. Departamento de Cirugía. Facultad de Medicina y Odontología. Universidad de Valencia.
- [29] Montoro Cazorla, D., Pérez Ocón, R. y Pereira das Neves Yedig, A. (2021). *A Longitudinal Study of the Bladder Cancer Applying a State-Space Model with Non-Exponential Staying Time in States*. Mathematics. 9(4), 363. Disponible en: <https://doi.org/10.3390/math9040363>
- [30] Neuts, M.F. (1981). *Matrix-Geometric Solutions in Stochastic Models. An Algorithmic Approach*. Dover Publications, INC. New York.
- [31] Norris, J.R. (1997). *Markov Chains*. Cambridge University Press.
- [32] Olsson, M., et. al. (1996). *Fitting phase type distributions via the EM algorithm*. Scand. J. Statist. 23, 419-441.
- [33] Pérez-Ocón, R., Ruiz-Castro, J.E. y Gámiz-Pérez, M.L. (1998). *A Multivariate Model to Measure the Effect of Treatments in Survival to Breast Cancer*. Departamento de Estadística e Investigación Operativa. UGR. Biometrical Journal 40 (1998) 6, 703-715.
- [34] Pérez-Ocón, R., Ruiz-Castro, J.E. y Gámiz-Pérez, M.L. (2000). *Markov models with lognormal transition rates in the analysis or survival times*. Departamento de Estadística e Investigación Operativa. UGR. Sociedad de Estadística e Investigación Operativa. Test. 9 , 2, 353-370.
- [35] Pérez-Ocón, R., Ruiz-Castro, J.E. y Gámiz-Pérez, M.L. (2001). *Non-homogeneous Markov models in the analysis of survival after breast cancer*. Departamento de Estadística e Investigación Operativa. UGR. Appl. Statist. 50, Part 1, 111-124.
- [36] Pérez-Ocón, R., Ruiz-Castro, J.E. y Gámiz-Pérez, M.L. (2001). *A piecewise Markov process for analysing survival from breast cancer in different risk groups*. Departamento de Estadística e Investigación Operativa. UGR. Statistics in Medicine. Statist Med. 20, 109-122.
- [37] Pérez-Ocón, R. y Ruiz-Castro, J.E. (2003) *A Multiple-Absorbent Markov Process in Survival Studies: Application to Breast Cancer*. Departamento

- de Estadística e Investigación Operativa. UGR. Biometrical Journal. 45, 8, 1-15.
- [38] Pérez Ocón, R. (2016). *Evaluación de la Fiabilidad y Mantenimiento de Sistemas de Ingeniería*. Apuntes de curso en el Máster Oficial en Estadística Aplicada. Departamento de Estadística e Investigación Operativa. Universidad de Granada. Curso 2016/17.
- [39] Porta, N., Calle, M. L., Malats, C. y Gómez, G. (2012). *A dynamic model for the risk of bladder cancer progression*. Statistics in Medicine. 31, 287-300.
- [40] Ross, S.M. (2010). *Introduction to probability models*. 10th ed. Academic Press. [Disponible en <http://www.sciencedirect.com/science/book/9780123756862>].
- [41] Rubio, G., García-Mora, B., Santamaría C. y Santonja, F. (2014). *Incorporating Covariates in a flowgraph model for bladder carcinoma*. Instituto de Matemática Multidisciplinar. Universitat Politècnica de Valencia. Universitat de Valencia.
- [42] Ruiz-Castro, J. E. y Pérez-Ocón, R. (2004). *Semi-Markov model in biomedical studies*. Communications in Statistics-Theory and Methods. 33, 437-455.
- [43] Sanguesa, C y Sanz, G (2014). *Cadenas de Markov en tiempo continuo* Apuntes del curso *Procesos Estocásticos y Probabilidad*, del Máster de Modelización e Investigación Matemática, Estadística y Computación. Departamento de Métodos Estadísticos. Universidad de Zaragoza. Curso 2013/2014.
- [44] Santamaría, C., García-Mora, B., Rubio, G. y Navarro, E. (2009). *A Markov model for analyzing the evolution of bladder carcinoma*. Mathematical and Computer Modelling. 50, 726-732.
- [45] Torres, I. (2000). *Distribuciones tipo fase en el análisis de la fiabilidad de sistemas reparables*. Tesis doctoral, dirigida por Rafael Pérez-Ocón. Departamento de Estadística. Facultad de Ciencias. Universidad de Granada.