



Guía 1

Natalie Julian

Septiembre

Pregunta 1

Sea $\mathbf{y} = (y_1, \dots, y_n)$ variables aleatorias independientes e idénticamente distribuidas desde una distribución exponencial $f(y_i|\lambda) = \lambda \exp(-\lambda y_i)$. Sea $\nu = (\nu_1, \dots, \nu_n)$ un vector de variables indicadoras de censura, donde $v_i = 1$ si y_i es tiempo de falla. Considere indexar el valor de λ por una covariable x . Para esto defina:

$$\lambda_i = \exp(\beta_0 + \beta_1 x_i)$$

Donde β_0 y β_1 corresponden a los parámetros a estimar. Utilice el conjunto de datos `sobrevida.csv`, el cual contiene información sobre los tiempos, censura y x .

- Asumiendo que β_1 es conocido, derive el estimador máximo verosímil de β_0 .
- Utilizando la expresión obtenida en a), calcule el estimador máximo verosímil de β_0 cuando $\beta_1 = 1$, $\beta_1 = 1.6$ y $\beta_1 = 2.1$.
- Calcule y grafique el estimador de Kaplan-Meier con sus respectivos intervalos de confianza.

Pregunta 2

Sea $\mathbf{y} = (y_1, \dots, y_n)$ realizaciones de variables aleatorias independientes e idénticamente distribuidas desde la siguiente densidad:

$$f_Y(y) = \frac{\alpha}{(1+y)^{\alpha+1}}, \quad \alpha > 0, \quad Y \in \mathbb{R}_+$$

Sea $\nu = (\nu_1, \dots, \nu_n)$ un vector de variables indicadoras de censura, donde $\nu_i = 0$ si y_i es censurada por la derecha y $\nu_i = 1$ si y_i es un tiempo de falla. Derive expresiones para lo que se le pide a continuación:

- $F(y)$
- $S(y)$
- $h(t)$
- Encuentre el estimador máximo verosímil de α .

Pregunta 3

Sea T una variable aleatoria con soporte en los reales positivos cuya función de riesgo está dada por:

$$h(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{[1 + (t/\alpha)^\beta]}, \quad \alpha > 0, \quad \beta > 0$$

Derive expresiones para:

- a) $H(t)$
- b) $S(t)$
- c) $f(t)$
- d) $F(t)$

Pregunta 4

Los siguientes datos corresponden a los tiempos de supervivencia (en años) para dos grupos en estudio, cada uno con 25 participantes. El Grupo 1 no tiene historia de enfermedad crónica (CHR=0), y el Grupo 2 tiene una historia positiva de enfermedad crónica (CHR=1):

Group1 (CHR=0): 12.3+, 5.4, 8.2, 12.2+, 11.7, 10.0, 5.7, 9.8, 2.6, 11.0, 9.2, 12.1+, 6.6, 2.2, 1.8, 10.2, 10.7, 11.1, 5.3, 3.5, 9.2, 2.5, 8.7, 3.8, 3.0

Group 2(CHR = 1) : 5.8, 2.9, 8.4, 8.3, 9.1, 4.2, 4.1, 1.8, 3.1, 11.4, 2.4, 1.4, 5.9, 1.6, 2.8, 4.9, 3.5, 6.5, 9.9, 3.6, 5.2, 8.8, 7.8, 4.7, 3.9

Utilizando estos datos,

- a) Calcule y grafique para cada grupo el estimador de Kaplan-Meier con sus respectivos intervalos de confianza.
- b) ¿Existen diferencias entre las curvas de supervivencia de los grupos? Justifique.

Pregunta 5

Conteste verdadero o falso a las siguientes afirmaciones:

- a) El análisis de supervivencia es una colección de procedimientos estadísticos para analizar datos en los cuales la variable observada es el tiempo que transcurre hasta que sucede un evento.
- b) En análisis de supervivencia el término evento es sinónimo de falla.
- c) En la práctica, la función de supervivencia es usualmente graficada como una curva suavizada.
- d) La función de supervivencia puede tomar valores entre $[0, \infty]$.
- e) Si conoce la forma de la función de riesgo, entonces puede determinar la correspondiente curva de supervivencia y viceversa.
- f) Si la curva de supervivencia para el grupo 1 cae completamente por encima de la curva de supervivencia para el grupo 2, entonces la media del tiempo de supervivencia para el grupo 2 es mayor que para el grupo 1.
- g) El conjunto que se encuentra en riesgo en seis semanas es el conjunto de individuos cuyos tiempos de supervivencia son menores o iguales a seis semanas.
- h) Si el conjunto que se encuentra en riesgo en seis semanas consiste de 22 personas, 4 personas fallan y 3 son censuradas para la semana 7, entonces el conjunto de riesgo en las 7 semanas consiste de 18 personas.
- i) Si la razón de riesgo que compara el grupo 1 relativo al grupo 2 es igual a 10, entonces el riesgo de falla es 10 veces mayor en el grupo 1 que en el grupo 2.

Pregunta 6

Asuma que los tiempos de vida T siguen la siguiente función de riesgo,

$$h(t|x) = h_0(t)\exp(x\beta)$$

Demuestre que $f(t|x) = h_0(t)\exp(x\beta)[S_0(t)]^{\exp(x\beta)}$.

Pregunta 7

En un ensayo clínico para estudiar la sobrevivencia en pacientes con cáncer de pulmón, los pacientes fueron asignados aleatoriamente a un tratamiento experimental o un tratamiento placebo. La siguiente tabla reporta la sobrevivencia de pacientes en meses, el signo + indica que la observación fue censurada.

	Tiempos de sobrevivencia t_i (meses)								
Placebo	1	1	1+	2	2	3	3	3+	
Tratamiento	1	2	2+	2	3	3	4	5	5+

- Obtenga la estimación de la función de sobrevivencia por Kaplan-Meier para los pacientes con placebo y tratamiento. Reporte la sobrevivencia de los pacientes al tiempo 2.
- Para determinar la efectividad del tratamiento, ajuste un modelo asumiendo que los tiempos de sobrevivencia siguen un modelo exponencial. Para ello, cree la variable indicadora trt que tome el valor 0 para el placebo y el valor 1 para el tratamiento experimental. Interprete.
- Ahora asuma para estos datos un modelo de regresión de Cox. Escriba el modelo, ajústelo e interprete. Estudie el supuesto de riesgos proporcionales.

Pregunta 8

Considere el modelo de riesgos proporcionales de Cox:

- Demuestre que la razón de riesgos para dos sujetos diferentes no depende del tiempo.
- Considere un modelo de Cox con tres predictores, sexo (codificado como: 1 para hombre y 0 para mujeres), peso del paciente en kilos y tratamiento (codificado como: 1 para los que siguen tratamiento, 0 para grupo control). Sea $\hat{\beta}_1 = 0.3$, $\hat{\beta}_2 = 0.77$ y $\hat{\beta}_3 = -0.18$ los coeficientes para el sexo, el peso y el tratamiento respectivamente. Calcule la razón de riesgo para cada coeficiente e interprete en términos de riesgos relativos.
- ¿Es cierto que en el modelo de Cox, las funciones de riesgo están relacionadas de manera aditiva? Justifique.
- ¿Es cierto que la razón de riesgo es constante sobre el tiempo de sobrevivencia? Justifique.

Pregunta 9

Un grupo de hombres con cáncer de pulmón avanzado e inoperable fue aleatorizado en dos grupos uno con la terapia estándar y otro con quimioterapia de prueba. Se registró el tiempo que sobrevivieron en días. Los predictores considerados fueron: Therapy (type of therapy: standard o test), Cell (type of tumor cell: adeno, large, small or squamous), Prior (prior therapy: 0=no, 1=yes), Age (age in years), Duration (months from diagnosis to randomization), y Kps (Karnofsky performance scale). Este último, es un valor entre 0 y 100, donde el valor 100 se le asigna a una persona completamente sana sin evidencias de enfermedad y el valor 0 a una persona muerta.

Las siguientes tablas presentan los resultados de haber ajustado el modelo de Cox a dichos datos (Las salidas corresponden al software SAS).

The PHREG Procedure
Class Level Information

Class	Value	Design Variables		
Prior	no	0		
	yes	1		
Cell	adeno	1	0	0
	large	0	0	0
	small	0	1	0
	squamous	0	0	1
Therapy	standard	0		
	test	1		

Wald Tests of Individual Effects

Joint Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Kps	1	35.5051	<.0001
Duration	1	0.1159	0.7335
Age	1	1.9772	0.1597
Cell	3	18.5339	0.0003
Prior	1	2.5296	0.1117
Therapy	1	5.2349	0.0221
Prior*Therapy	1	4.1528	0.0416

Parameters Estimates with Reference Coding

Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Label
Kps		1	-0.03300	0.00554	35.5051	<.0001	Karnofsky performance scale
Duration		1	0.00323	0.00949	0.1159	0.7335	months from diagnosis to randomization
Age		1	-0.01353	0.00962	1.9772	0.1597	age in years
Cell	adeno	1	0.78356	0.30382	6.6512	0.0099	cell type adeno
Cell	small	1	0.48230	0.26537	3.3032	0.0691	cell type small
Cell	squamous	1	-0.40770	0.28363	2.0663	0.1506	cell type squamous
Prior	yes	1	0.45914	0.28868	2.5296	0.1117	prior therapy yes
Therapy	test	1	0.56662	0.24765	5.2349	0.0221	type of treatment test
Prior*Therapy	yes test	1	-0.87579	0.42976	4.1528	0.0416	prior therapy yes * type of treatment test

Responda las siguientes preguntas:

- ¿Qué predictores tienen un efecto significativo sobre la sobrevivencia de los pacientes?
- ¿Cuál es el riesgo relativo del grupo con terapia (test) en relación a terapia (standard)?
- ¿El Kps es un factor de riesgo o un factor protector? Justifique.
- Calcule el riesgo relativo del término de interacción.
- Calcule los riesgos relativos del tipo de célula. Interprete, ¿cuáles son factores protectores y cuáles factores de riesgo?
- ¿Recomienda o no la quimioterapia test? ¿Bajo qué circunstancias? Argumente.

Pregunta 10

Considere la siguiente parametrización de la distribución Weibull:

$$f(y) = \frac{k}{\lambda} \left(\frac{y}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{y}{\lambda}\right)^k\right); \quad y > 0, \lambda > 0, k > 0$$

- Reparametrice la distribución con $\gamma = \left(\frac{1}{\lambda}\right)^k$
- Por conveniencia, considere $\phi = \log \gamma$ y obtenga las expresiones usuales ($F(y)$, $S(y)$, etcétera).
- Suponga que se tiene las observaciones $y_i = (T_i, \delta_i)$ con $i = 1, \dots, n$ independientes e idénticamente distribuidas Weibull(α, ϕ) y con δ_i indicatriz de censura por la derecha. Asuma α y ϕ desconocidos con priori para α una distribución Gamma con parámetros b_0 y b_1 y para ϕ una distribución normal con parámetros μ_0 y σ_0 .
- Obtenga la posteriori conjunta y las condicionales completas.
- Escriba un algoritmo en R para muestrear desde las posteriores completas (Metropolis-Hastings). Para la implementación utilice la base de datos **Fan** y fije $b_0 = b_1 = 0.01$, $\mu_0 = 0$ y $\sigma_0^2 = 1000$ y $c = 0.1$ para las distribuciones para obtener candidatos.
- Interprete los resultados obtenidos (calcule bandas de credibilidad y estadísticas de resumen, etc).
- Grafique la función de riesgo y la función de sobrevivencia.

Pregunta 11

Considere n individuos sin censura con tiempos de falla distribuidos exponencialmente y función de log verosimilitud $n \log \rho - \rho \sum x_i$, pruebe que el estadístico de razón W para testear $H_0 : \rho = \rho_0$ es:

$$W = 2(n \log n - n \log \sum x_i - n - n \log \rho_0 + \rho_0 \sum x_i)$$

Y pruebe que el valor esperado bajo la hipótesis nula es:

$$E(W) = 1 + (6n)^{-1} + O(n^{-2})$$

Pregunta 12

Asuma que la censura posee un componente aleatorio y considere la extensión de riesgos constantes por tramos:

$$h_c(t) = \lambda_j \quad (a_{j-1} \leq t < a_j)$$

Y defina $b_j = a_j - a_{j-1}$ el ancho del intervalo. Considere que no se tiene el tiempo de sobrevivencia y la censura de manera exacta, sino que de manera agrupada, es decir, sólo se conoce d_j (el número de fallas) y m_j (el número de censuras) en el intervalo $[a_{j-1}, a_j)$.

- Derive expresiones para las probabilidades condicionales en cada uno de los casos que se presentan.
- Determine una expresión para la contribución de cada intervalo a la log verosimilitud conjunta.
- Encuentre los estimadores de máxima verosimilitud para λ y ρ .
- Cree una función en R que ajuste un modelo exponencial por tramos, la función debe tener como argumentos los datos y los puntos de corte. Pruebe la función (puede utilizar la base de datos **Fan** y ajústela con distintas particiones).

Pregunta 13

Sean T y U variables aleatorias independientes no negativas y defina $X = \min(T, U)$ que corresponde a una observación censurada del tiempo de fallo de la variable T y sea $\delta = \mathbf{1}_{\{T \leq U\}}$ la variable indicatriz de censura (que toma el valor 1 si $X \leq t$ y cero en el otro caso). Defina el proceso de conteo $N(t) : t \geq 0$ dado por:

$$N(t) = \mathbf{1}_{\{X \leq t, \delta=1\}} = \delta \mathbf{1}_{\{T \leq t\}}$$

Muestre que:

$$\lambda(t) \Delta t \approx E[N((t + \Delta t) -) - N(t -) | T \geq t, U \geq t]$$

Interprete la función de riesgo $\lambda(t)$ en este contexto.

1 Soluciones

Pregunta 1

a) **Solución**

La función de verosimilitud está dada por:

$$L(\lambda|y, \nu) = \prod_{i=1}^n h(y_i|\lambda_i)^{\nu_i} S(y_i|\lambda_i)$$

Donde $h(y|\lambda)$ representa la función de riesgo y $S(y|\lambda)$ la función de supervivencia. Para este caso, se tiene:

$$\begin{aligned} S(y) &= 1 - F(y) \\ &= 1 - \int_0^y \lambda e^{-\lambda u} du \\ &= 1 - (1 - e^{-\lambda y}) \\ &= e^{-\lambda y} \end{aligned}$$

$$\begin{aligned} h(y) &= \frac{f(y)}{S(y)} \\ &= \frac{\lambda e^{-\lambda y}}{e^{-\lambda y}} \\ &= \lambda \end{aligned}$$

Así, reemplazando, la función de verosimilitud está dado por:

$$L(\lambda|y, \nu) = \prod_{i=1}^n \lambda_i^{\nu_i} e^{-\lambda_i y_i}$$

Reemplazando $\lambda_i = \exp(\beta_0 + \beta_1 x_i)$ se tiene:

$$L(\lambda|y, \nu) = e^{\sum_{i=1}^n \nu_i (\beta_0 + \beta_1 x_i)} e^{-\sum_{i=1}^n y_i e^{(\beta_0 + \beta_1 x_i)}}$$

Y la función de log-verosimilitud:

$$l(\beta_0|y, \nu, \beta_1, x) = \sum_{i=1}^n \nu_i (\beta_0 + \beta_1 x_i) - \sum_{i=1}^n y_i e^{(\beta_0 + \beta_1 x_i)}$$

Al igualar a cero la función score, se obtiene:

$$\hat{\beta}_0 = \log\left(\sum_{i=1}^n \nu_i\right) - \log\left(\sum_{i=1}^n y_i e^{\beta_1 x_i}\right)$$

Se comprueba que es el estimador máximo verosímil pues la segunda derivada de la función de log-verosimilitud resulta ser negativa.

b) **Solución**

```
datosobrevida<-read.csv('sobrevida.csv',h=T)

head(datosobrevida,4)

##      tiempos censura      x
## 1 6.7382748      1 0.3474116
## 2 6.4010317      1 0.7626581
## 3 0.1236848      1 0.9002416
## 4 1.3070063      1 0.6682708

beta0hat<-function(beta1,datos){
  nu<-datos$censura
  y<-datos$tiempos
  x<-datos$x

  beta0<-log(sum(nu))-log(sum(y*exp(x*beta1)))

  return(beta0)}

beta0hat(1,datosobrevida)

## [1] -1.725132

beta0hat(1.6,datosobrevida)

## [1] -2.020935

beta0hat(2.1,datosobrevida)

## [1] -2.290013
```

c) Solución

```
#install.packages("survival")
library(survival)
my.surv.object<-Surv(time=datosobrevida$tiempos,event = datosobrevida$censura, type = "right")

survfit(my.surv.object ~ 1,conf.type="plain", conf.int=0.90)

## Call: survfit(formula = my.surv.object ~ 1, conf.type = "plain", conf.int = 0.9)
##
##      n  events  median  0.9LCL  0.9UCL
## 1000.00  916.00    2.21    2.06    2.40

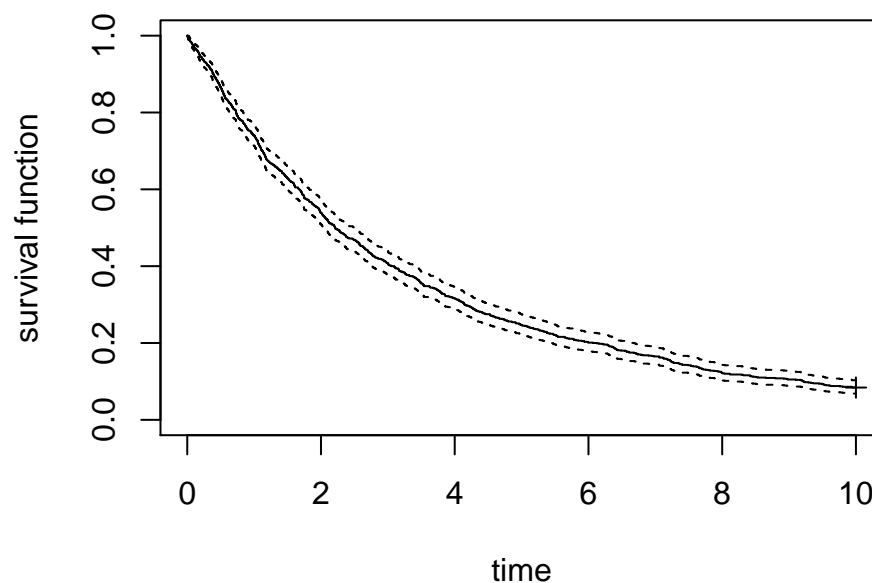
my.fit<-survfit(my.surv.object ~ 1)
str(my.fit)

## List of 17
##  $ n      : int 1000
##  $ time    : num [1:917] 0.00436 0.00893 0.00949 0.01052 0.01277 ...
##  $ n.risk   : num [1:917] 1000 999 998 997 996 995 994 993 992 991 ...
##  $ n.event  : num [1:917] 1 1 1 1 1 1 1 1 1 1 ...
##  $ n.censor : num [1:917] 0 0 0 0 0 0 0 0 0 0 ...
##  $ surv     : num [1:917] 0.999 0.998 0.997 0.996 0.995 0.994 0.993 0.992 0.991 0.99 ...
##  $ std.err  : num [1:917] 0.001 0.00142 0.00173 0.002 0.00224 ...
##  $ cumhaz   : num [1:917] 0.001 0.002 0.003 0.00401 0.00501 ...
##  $ std.chaz : num [1:917] 0.001 0.00141 0.00173 0.002 0.00224 ...
##  $ start.time: num 0
##  $ type     : chr "right"
##  $ logse    : logi TRUE
##  $ conf.int  : num 0.95
##  $ conf.type: chr "log"
##  $ lower     : num [1:917] 0.997 0.995 0.994 0.992 0.991 ...
##  $ upper     : num [1:917] 1 1 1 1 0.999 ...
##  $ call      : language survfit(formula = my.surv.object ~ 1)
##  - attr(*, "class")= chr "survfit"

#summary(my.fit)

plot(my.fit, main="Kaplan-Meier estimate",
      xlab="time", ylab="survival function",mark.time=T)
```

Kaplan–Meier estimate



```
print(survfit(my.surv.object ~ 1), print.rmean=TRUE)

## Call: survfit(formula = my.surv.object ~ 1)
##
##      n      events   *rmean *se(rmean)   median  0.95LCL
```



```
## 1.00e+03 9.16e+02 3.35e+00 9.67e-02 2.21e+00 2.05e+00
## 0.95UCL
## 2.49e+00
## * restricted mean with upper limit = 10
```

Pregunta 2

a) Solución

$$F(y) = 1 - \frac{1}{(1+y)^\alpha}$$

b) Solución

$$S(y) = \frac{1}{(1+y)^\alpha}$$

c) Solución

$$h(y) = \frac{\alpha}{1+y}$$

d) Solución

$$\hat{\alpha}_{EMV} = \frac{\sum_{i=1}^n \nu_i}{\sum_{i=1}^n \ln(1+y_i)}$$

Pregunta 3

a) Solución

$$H(t) = \ln(1 + (t/\alpha)^\beta)$$

b) Solución

$$S(t) = [1 + (t/\alpha)^\beta]^{-1}$$

c) Solución

$$f(t) = \frac{(\beta/\alpha)(t/\alpha)^{\beta-1}}{[1 + (t/\alpha)^\beta]^2}$$

d) Solución

$$F(t) = 1 - [1 + (t/\alpha)^\beta]^{-1}$$

Pregunta 4

a) Solución

```
GRUPO1<-rbind(c(12.3,0),c(5.4,1),c(8.2,1),c(12.2,0),c(11.8,1),c(10,1),c(5.7,1),c(9.8,1),
              c(2.6,1),c(11,1),c(9.2,1),c(12.1,0),c(6.6,1),c(2.2,1),c(1.8,1),c(10.2,1),
              c(10.7,1),c(11.1,1),c(5.3,1),c(3.5,1),c(9.2,1),c(2.5,1),c(8.7,1),c(3.8,1),c(3,1))

GRUPO2<-cbind(c(5.8,2.9,8.4,8.3,9.1,4.2,4.1,1.8,3.1,11.4,2.4,1.4,5.9,1.6,2.8,4.9,3.5,6.5,
              9.9,3.6,5.2,8.8,7.8,4.7,3.9),rep(1,25))

#GRUPO 1

my.surv.object<-Surv(time=GRUPO1[,1],event = GRUPO1[,2], type = "right")

survfit(my.surv.object ~ 1,conf.type="plain", conf.int=0.90)

## Call: survfit(formula = my.surv.object ~ 1, conf.type = "plain", conf.int = 0.9)
##
##      n events median 0.9LCL 0.9UCL
## 25.0   22.0    8.7    5.4   10.0

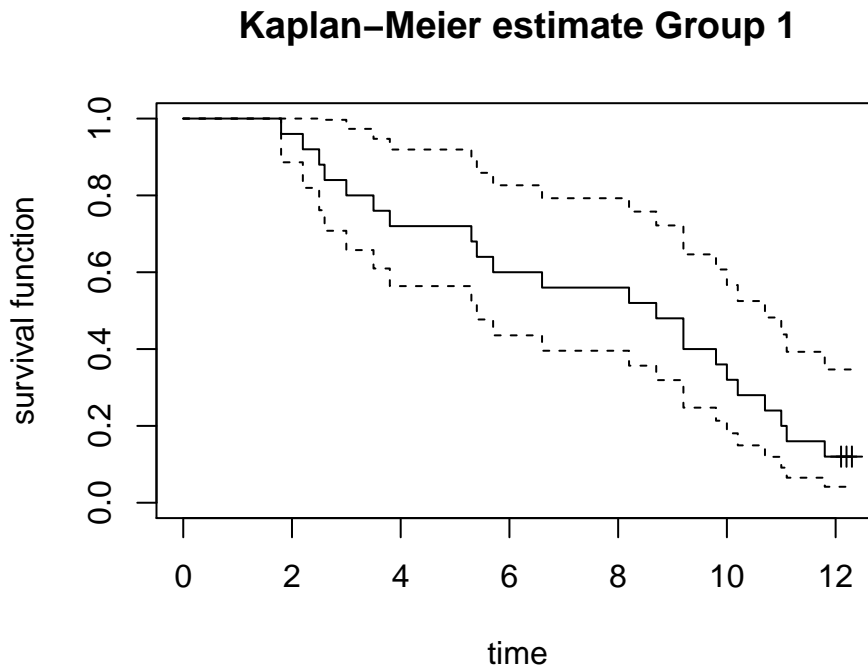
my.fit1<-survfit(my.surv.object ~ 1)
str(my.fit1)

## List of 17
## $ n      : int 25
## $ time   : num [1:24] 1.8 2.2 2.5 2.6 3 3.5 3.8 5.3 5.4 5.7 ...
## $ n.risk  : num [1:24] 25 24 23 22 21 20 19 18 17 16 ...
## $ n.event : num [1:24] 1 1 1 1 1 1 1 1 1 1 ...
## $ n.censor: num [1:24] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv    : num [1:24] 0.96 0.92 0.88 0.84 0.8 0.76 0.72 0.68 0.64 0.6 ...
## $ std.err  : num [1:24] 0.0408 0.059 0.0739 0.0873 0.1 ...
## $ cumhaz   : num [1:24] 0.04 0.0817 0.1251 0.1706 0.2182 ...
## $ std.chaz  : num [1:24] 0.04 0.0578 0.0723 0.0854 0.0978 ...
## $ start.time: num 0
## $ type     : chr "right"
```

```
## $ logse      : logi TRUE
## $ conf.int   : num 0.95
## $ conf.type  : chr "log"
## $ lower      : num [1:24] 0.886 0.82 0.761 0.708 0.658 ...
## $ upper      : num [1:24] 1 1 1 0.997 0.973 ...
## $ call       : language survfit(formula = my.surv.object ~ 1)
## - attr(*, "class")= chr "survfit"
```

```
#summary(my.fit1)
```

```
plot(my.fit1, main="Kaplan-Meier estimate Group 1",
      xlab="time", ylab="survival function", mark.time=T)
```



```
print(survfit(my.surv.object ~ 1), print.rmean=TRUE)
```

```
## Call: survfit(formula = my.surv.object ~ 1)
##
##          n      events      *rmean *se(rmean)   median  0.95LCL
##    25.000    22.000      7.568      0.718     8.700    5.400
##    0.95UCL
##    10.700
##    * restricted mean with upper limit = 12.3
```

```
#GRUPO 2
```

```
my.surv.object<-Surv(time=GRUPO2[,1],event = GRUPO2[,2], type = "right")
```

```
survfit(my.surv.object ~ 1,conf.type="plain", conf.int=0.90)
```

```
## Call: survfit(formula = my.surv.object ~ 1, conf.type = "plain", conf.int = 0.9)
```

```
##
##          n events median 0.9LCL 0.9UCL
##    25.0   25.0   4.7    3.6    5.9
```

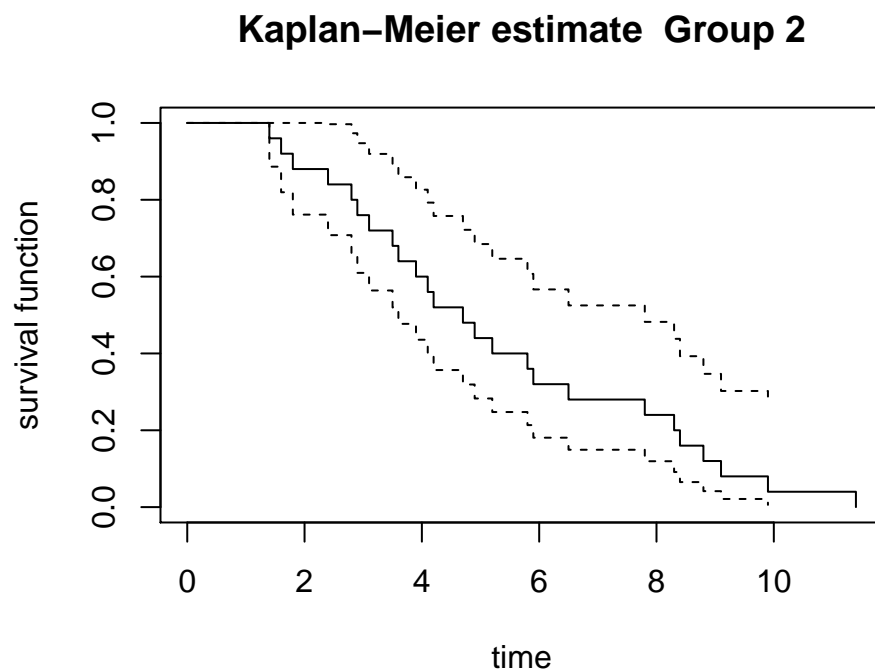
```
my.fit2<-survfit(my.surv.object ~ 1)
str(my.fit2)
```

```
## List of 17
## $ n      : int 25
## $ time   : num [1:25] 1.4 1.6 1.8 2.4 2.8 2.9 3.1 3.5 3.6 3.9 ...
## $ n.risk : num [1:25] 25 24 23 22 21 20 19 18 17 16 ...
## $ n.event : num [1:25] 1 1 1 1 1 1 1 1 1 1 ...
## $ n.censor : num [1:25] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv     : num [1:25] 0.96 0.92 0.88 0.84 0.8 0.76 0.72 0.68 0.64 0.6 ...
## $ std.err  : num [1:25] 0.0408 0.059 0.0739 0.0873 0.1 ...
## $ cumhaz   : num [1:25] 0.04 0.0817 0.1251 0.1706 0.2182 ...
## $ std.chaz : num [1:25] 0.04 0.0578 0.0723 0.0854 0.0978 ...
## $ start.time : num 0
## $ type      : chr "right"
```

```
## $ logse      : logi TRUE
## $ conf.int   : num 0.95
## $ conf.type  : chr "log"
## $ lower      : num [1:25] 0.886 0.82 0.761 0.708 0.658 ...
## $ upper      : num [1:25] 1 1 1 0.997 0.973 ...
## $ call       : language survfit(formula = my.surv.object ~ 1)
## - attr(*, "class")= chr "survfit"
```

```
#summary(my.fit2)
```

```
plot(my.fit2, main="Kaplan-Meier estimate Group 2",
      xlab="time", ylab="survival function", mark.time=T)
```



```
print(survfit(my.surv.object ~ 1), print.rmean=TRUE)
```

```
## Call: survfit(formula = my.surv.object ~ 1)
##
##      n      events    *rmean *se(rmean)   median  0.95LCL
## 25.000    25.000     5.280    0.551     4.700    3.600
## 0.95UCL
## 7.800
## * restricted mean with upper limit = 11.4
```

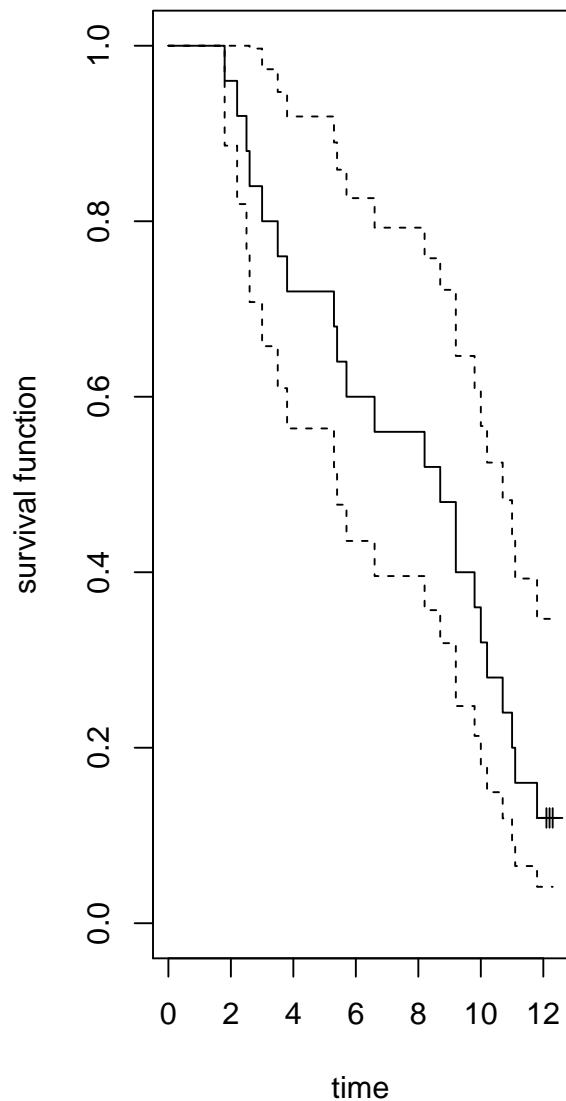
b) Solución

```
par(mfrow=c(1,2))

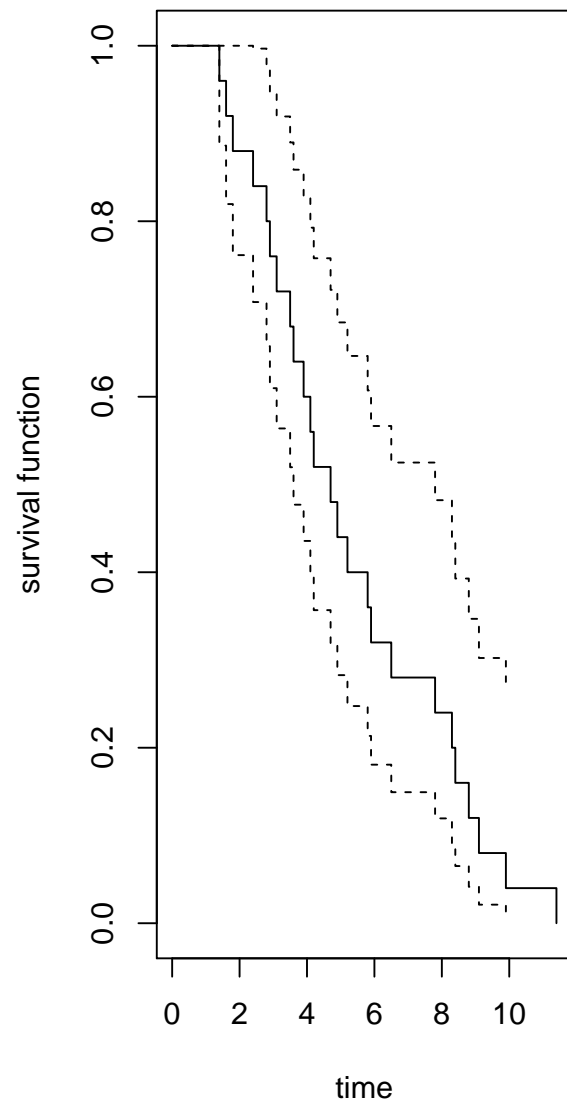
plot(my.fit1, main="Kaplan-Meier estimate Group 1",
      xlab="time", ylab="survival function", mark.time=T)

plot(my.fit2, main="Kaplan-Meier estimate Group 2",
      xlab="time", ylab="survival function", mark.time=T)
```

Kaplan–Meier estimate Group 1



Kaplan–Meier estimate Group 2



```
DATA<-rbind(cbind(GRUPO1,rep(3,25)),cbind(GRUPO2,rep(2,25)))
colnames(DATA)<-c("tiempos","censura","trat")
DATA<-data.frame(DATA)
survdif(Surv(tiempos, censura)~trat, data=DATA)

## Call:
## survdiff(formula = Surv(tiempos, censura) ~ trat, data = DATA)
##
##      N Observed Expected (O-E)^2/E (O-E)^2/V
## trat=2 25      25     16.2    4.76    7.99
## trat=3 25      22     30.8    2.51    7.99
##
##  Chisq= 8  on 1 degrees of freedom, p= 0.005

#Las curvas difieren. (investigar test)
```

Pregunta 5

Solución

- a) Verdadero.
- b) Verdadero.
- c) Falso. El Kaplan-Meier posee una gráfica escalonada usualmente.

- d) Falso. Está entre 0 y 1.
- e) Verdadero.
- f) Falso. La media del tiempo de supervivencia para el grupo 1 es mayor que para el grupo 2.
- g) Falso. Cuyos tiempos de supervivencia son mayores a seis semanas.
- h) Falso. Serían 13 personas.
- i) Verdadero.

Pregunta 7

Solución

a) Solución

Para placebo:

t_k	d_k	q_k	n_k	$\hat{S}(t)$
0	0	0	8	1
1	2	1	8	$(1 - \frac{2}{8})$
2	2	0	5	$(1 - \frac{2}{8})(1 - \frac{2}{5})$
3	2	1	5	$(1 - \frac{2}{8})(1 - \frac{2}{5})(1 - \frac{2}{5})$

Para tratamiento:

t_k	d_k	q_k	n_k	$\hat{S}(t)$
0	0	0	9	1
1	1	0	9	$(1 - \frac{1}{9})$
2	2	1	8	$(1 - \frac{1}{9})(1 - \frac{2}{8})$
3	2	0	5	$(1 - \frac{1}{9})(1 - \frac{2}{8})(1 - \frac{2}{5})$
4	1	0	3	$(1 - \frac{1}{9})(1 - \frac{2}{8})(1 - \frac{2}{5})(1 - \frac{1}{3})$
5	1	1	2	$(1 - \frac{1}{9})(1 - \frac{2}{8})(1 - \frac{2}{5})(1 - \frac{1}{3})(1 - \frac{1}{2})$

O también:

```
datos<-cbind(c(1,1,1,2,2,3,3,3,1,2,2,2,3,3,4,5,5),
             c(1,1,0,1,1,1,1,0,1,1,0,1,1,1,1,0),c(rep(0,8),rep(1,9)))

head(datos)

##      [,1] [,2] [,3]
## [1,]    1    1    0
## [2,]    1    1    0
## [3,]    1    0    0
## [4,]    2    1    0
## [5,]    2    1    0
## [6,]    3    1    0

library(survival)
Sp<-Surv(datos[1:8,1],datos[1:8,2])
Sp

## [1] 1  1  1+ 2  2  3  3  3+

St<-Surv(datos[9:17,1],datos[9:17,2])
St

## [1] 1  2  2+ 2  3  3  4  5  5+

spfit<-survfit(Sp~1,)
stfit<-survfit(St~1,)

summary(spfit)

## Call: survfit(formula = Sp ~ 1)
##
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##      1      8      2     0.75  0.153     0.5027     1.000
##      2      5      2     0.45  0.188     0.1982     1.000
##      3      3      2     0.15  0.138     0.0248     0.906

summary(stfit)
```

```
## Call: survfit(formula = St ~ 1)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   1      9      1    0.889   0.105   0.7056   1.000
##   2      8      2    0.667   0.157   0.4200   1.000
##   3      5      2    0.400   0.174   0.1707   0.938
##   4      3      1    0.267   0.159   0.0829   0.858
##   5      2      1    0.133   0.123   0.0218   0.817

summary(spfit,time=2)

## Call: survfit(formula = Sp ~ 1)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   2      5      4    0.45   0.188   0.198   1

summary(stfit,time=2)

## Call: survfit(formula = St ~ 1)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   2      8      3    0.667   0.157   0.42   1
```

b) Solución

```
sexp<-survreg(Surv(datos[,1],datos[,2])~datos[,3],dist="exponential")
sexp

## Call:
## survreg(formula = Surv(datos[, 1], datos[, 2]) ~ datos[, 3],
##   dist = "exponential")
##
## Coefficients:
## (Intercept)  datos[, 3]
##   0.9808293   0.3690975
##
## Scale fixed at 1
##
## Loglik(model)= -28.3   Loglik(intercept only)= -28.6
##   Chisq= 0.43 on 1 degrees of freedom, p= 0.51
## n= 17

summary(sexp)

##
## Call:
## survreg(formula = Surv(datos[, 1], datos[, 2]) ~ datos[, 3],
##   dist = "exponential")
##           Value Std. Error    z    p
## (Intercept)  0.981      0.408 2.40 0.016
## datos[, 3]   0.369      0.556 0.66 0.507
##
## Scale fixed at 1
##
## Exponential distribution
## Loglik(model)= -28.3   Loglik(intercept only)= -28.6
##   Chisq= 0.43 on 1 degrees of freedom, p= 0.51
## Number of Newton-Raphson Iterations: 3
## n= 17

#El valor p es igual a 0.5, utilizando un 5%, no sería significativo el tratamiento
#para explicar los tiempos de sobrevivencia
AIC(sexp)

## [1] 60.66893
```

c) Solución

```
fit_0<-coxph(Surv(datos[,1],datos[,2])~datos[,3])
summary(fit_0)

## Call:
## coxph(formula = Surv(datos[, 1], datos[, 2]) ~ datos[, 3])
##
##   n= 17, number of events= 13
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## datos[, 3] -0.6671    0.5132  0.6090 -1.095  0.273
```

```
##
##          exp(coef) exp(-coef) lower .95 upper .95
## datos[, 3]    0.5132      1.949    0.1556    1.693
##
## Concordance= 0.603 (se = 0.085 )
## Likelihood ratio test= 1.2 on 1 df,   p=0.3
## Wald test           = 1.2 on 1 df,   p=0.3
## Score (logrank) test = 1.24 on 1 df,   p=0.3

#Al ajustar el modelo de riesgos proporcionales, se obtiene un valor-p para el
#tratamiento de 0.27, es decir, aún no es significativo.
#la exp(coeftrat)=0.5132, es decir, cuando paso del placebo al tratamiento, los
#riesgos disminuyen un 48% aproximadamente.
#El intervalo de confianza al 95% es (0.1556,1.693), el cual contiene al 1,
#por lo tanto, no se puede rechazar que el efecto del tratamiento sea nulo.

cox.zph(fit_0, transform="km", global=TRUE)

##          rho chisq      p
## datos[, 3] 0.0172 0.0038 0.951

#Obtenemos un valor-p de 0.951, es decir, no rechazamos H0 (que los riesgos son proporcionales)
```

Pregunta 8

- a) **Solución**
Trabajar la expresión $HR(t, x_1, x_2)$ y llegar a que no depende del tiempo (se cancelan las funciones que dependen de t).

- b) **Solución**

$$HR_{sexo}(t, x_1 = 1, x_2 = 0) = e^{\hat{\beta}_1} = e^{-0.3} = 1.35$$

Los hombres tienen un 35 por ciento más de riesgo en comparación a las mujeres.

$$HR_{peso}(t, x_1 = 70, x_2 = 69) = e^{\hat{\beta}_2} = e^{0.77} = 2.16$$

Por cada kilo de aumento en el peso, el riesgo aumenta un 116 por ciento.

$$HR_{tratamiento}(t, x_1 = 1, x_2 = 0) = e^{\hat{\beta}_3} = e^{-0.18} = 0.83$$

El tratamiento baja el riesgo en un 17 por ciento.

- c) **Solución**
Falso. Las funciones de riesgo están relacionadas de manera multiplicativa.
- d) **Solución**
Verdadero. La razón de riesgo es constante sobre el tiempo de sobrevida.

Pregunta 9

- a) **Solución**
Los predictores con un efecto significativo sobre la sobrevivencia de los pacientes son: Karnofsky performance scale, cell type, therapy y la interacción de Prio therapy con la therapy actual. Esto se deduce de los resultados del test de Wald.
- b) **Solución**
 $e^{0.56662} = 1.7623$. Los pacientes en el grupo de terapia test un 76 por ciento más de riesgo de morir en relación a los que reciben la terapia estándar.
- c) **Solución**
 $e^{-0.033} = 0.9675$ es un factor protector, de hecho el riesgo de morir disminuye en un 3.2 por ciento por cada unidad de aumento en el kps.
- d) **Solución**
 $e^{-0.87579} = 0.4165$. Los pacientes que recibieron terapia previa y ahora están recibiendo la quimioterapia de prueba, tienen un 58.35 por ciento menos de riesgos de morir.
- e) **Solución**
 $e^{0.78356} = 2.189252$. Los pacientes con celulas tipo adeno tienen un 118 por ciento más de riesgo de morir en relación a los que tienen células large. (Factor de riesgo)
Los pacientes con células small tienen un 61.98 por ciento más de riesgo de morir en relación a los que tienen células large. (Factor de riesgo)
 $e^{-0.4077} = 0.6652$. Los pacientes con células tipo suqmous tienen un 33.48 por ciento menos de riesgo de morir en relación a los que tienen células large. (Factor protector)
- f) **Solución**
Se recomienda aplicar la quimioterapia test sólo a los pacientes que recibieron una terapia previa.

Pregunta 10

Solución

```

Fan<-read.table('Fan.txt')
colnames(Fan)<-c("lifetime","censor")
Fan[,1]<-Fan$lifetime/1000

attach(Fan)

library(truncnorm)

# Calcula cantidades fijas
nu=(censor==0)*1; y=lifetime; n=length(nu)
sum_nu=sum(nu)

# Fija hyper-parámetros
mu0=0; sigma02=1000; alpha0=0.01; kappa0=0.01; c=0.1

# Fija valores iniciales
alpha=5; lambda=0.1
alphaante=alpha; lambdaante=lambda
nscan=50000

alphaM=vector(mode="numeric", length=nscan)
lambdaM=vector(mode="numeric", length=nscan)
alphaM[1]=alpha
lambdaM[1]=lambda

for(i in 2:nscan)
{
  # Actualización de alpha
  alphaaste=rtruncnorm(n=1, a=0, b=Inf, mean=alphaante, sd=c)
  logPaste=(alpha0+sum_nu-1)*log(alphaaste)+ sum((alphaaste-1)*nu*log(y)-exp(lambda)*y^alphaaste)-alphaaste*kappa0
  logPante=(alpha0+sum_nu-1)*log(alphaante)+ sum((alphaante-1)*nu*log(y)-exp(lambda)*y^alphaante)-alphaante*kappa0
  logJtaste=log(dtruncnorm(x=alphaaste, a=0, b=Inf, mean=alphaante, sd=c))
  logJtante=log(dtruncnorm(x=alphaante, a=0, b=Inf, mean=alphaaste, sd=c))
  lnrl=logPaste-logJtaste-logPante+logJtante
  p1=exp(min(lnrl,0))
  u1=runif(1)
  if(p1>u1){alpha=alphaaste}
  if(p1<=u1){alpha=alphaante}
  alphaante=alpha

  # Actualización de lambda
  lambdaaste=rnorm(n=1, mean=lambdaante, sd=c)
  logPaste=lambdaaste*sum_nu-exp(lambdaaste)*sum(y^alpha)-(0.5*(lambdaaste-mu0)^2)/sigma02
  logPante=lambdaante*sum_nu-exp(lambdaante)*sum(y^alpha)-(0.5*(lambdaante-mu0)^2)/sigma02
  lnrl2=logPaste-logPante
  p2=exp(min(lnrl2,0))
  u2=runif(1)
  if(p2>u2){lambda=lambdaaste}
  if(p2<=u2){lambda=lambdaante}
  lambdaante=lambda

  alphaM[i]=alpha
  lambdaM[i]=lambda
}

plot(alphaM, type="l")
plot(lambdaM, type="l")

summary(alphaM)
summary(lambdaM)

# Muestras desde la posteriori
indices=seq(2000,50000,48)
alpham=alphaM[indices]
lambdam=lambdaM[indices]
acf(alpham)
acf(lambdam)

hist(alpham)
hist(lambdam)
mean.alpha=mean(alpham)
mean.lambda=mean(lambdam)
exp(mean(lambdam))

p025.alpha=quantile(alpham, probs=0.025)
p0975.alpha=quantile(alpham, probs=0.975)

p025.lambda=quantile(lambdam, probs=0.025)
p0975.lambda=quantile(lambdam, probs=0.975)

# Función de supervivencia
yy=seq(0,2,0.01)
Sy=exp(-exp(mean.lambda)*yy^(mean.alpha))
Sy025=exp(-exp(p025.lambda)*yy^(p025.alpha))
Sy0975=exp(-exp(p0975.lambda)*yy^(p0975.alpha))
plot(yy, Sy, type="l")
lines(yy, Sy025, lty=2, col=2)
lines(yy, Sy0975, lty=2, col=2)

```



```
# Función de riesgo
hy=mean.alpha*yy^(mean.alpha-1)*exp(mean.lambda)
hy025=p025.alpha*yy^(p025.alpha-1)*exp(p025.lambda)
hy0975=p0975.alpha*yy^(p0975.alpha-1)*exp(p0975.lambda)
plot(yy, hy, type="l")
lines(yy, hy025, lty=2, col=2)
lines(yy, hy0975, lty=2, col=2)
```

Pregunta 11

Se puede definir de manera general:

$$W(\omega_0) = W = 2[l(\hat{w}, \hat{\lambda}) - l(\omega_0, \hat{\lambda}_{\omega_0})]$$

Donde $l(\hat{w}, \hat{\lambda})$ corresponde a la estimación de máxima verosimilitud usual y $l(\omega_0, \hat{\lambda}_{\omega_0})$ corresponde a la estimación de máxima verosimilitud restringida a la hipótesis nula ($\omega = \omega_0$). Bajo la hipótesis nula, el estadístico $W(\omega_0) = W$ tiene aproximadamente una distribución chi-cuadrado, con $p_\omega = \dim(\omega)$ grados de libertad. Si la distribución asintótica fuera exacta, se tendría que $E(W) = p_\omega$, sin embargo, se tiene por el *Factor de corrección de Barlett* que:

$$E(W) = p_\omega \left[1 + \frac{c}{n} + o\left(\frac{1}{n}\right) \right]$$

En este caso, $p_\omega = 1$ y c se estima de manera consistente.

Pregunta 12

a) Solución

Existen tres casos en este problema.

i) $r_{j-1} - d_j - m_j$ sujetos que sobrevivieron durante el intervalo y que no están censurados, la probabilidad condicional en este caso es $\exp(-b(\rho + \lambda))$.

ii) d_j sujetos que fallaron en el intervalo, su probabilidad condicional es $\frac{\rho}{\rho + \lambda} (1 - \exp[-b(\rho + \lambda)])$.

iii) m_j sujetos censurados en el intervalo, su probabilidad condicional es $\frac{\lambda}{\lambda + \rho} (1 - \exp[-b(\rho + \lambda)])$

b) Solución

La contribución en cada intervalo es:

$$l_j(p_j, \lambda_j) = -(r - d - m)b(\rho + \lambda) + d \log(\rho/(\rho + \lambda)) + m \log(\lambda/(\rho + \lambda)) + (d + m) \log(1 - \exp[-b(\rho + \lambda)])$$

c) Solución

Los estimadores de máxima verosimilitud son:

$$\hat{\rho} = -\frac{d}{b(d + m)} \log((r - d - m)/r)$$

$$\hat{\lambda} = -\frac{m}{b(d + m)} \log((r - d - m)/r)$$

d) Solución

```
exptramos <- function(data, timevar, deathvar, bounds) {
  # pwe: expands an S data frame for piece-wise exponential survival
  # G. Rodriguez, Nov 29, 1992
  #
  # Check arguments: time and death must be variables in the data frame
  # and boundaries must be non-negative and strictly increasing
  if(!is.data.frame(data)) stop("First argument must be a data frame")
  if(is.na(match(tn <- deparse(substitute(timevar)), names(data))))
    stop(paste("\n\tSurvival time", tn,
               "must be a variable in the data frame"))
  if(is.na(match(dn <- deparse(substitute(deathvar)), names(data))))
    stop(paste("\n\tDeath indicator", dn,
               "must be a variable in the data frame"))
  width <- diff(bounds)
  if(any(bounds < 0) | any(width <= 0)) stop(paste(
    "Invalid interval boundaries in", deparse(substitute(
      bounds)))) #
  # Expand the data frame creating one pseudo-observation for each
  # interval visited, add interval number, events and exposure time
  # (existing variables with these names will be overwritten)
  n <- cut(data[, tn], bounds)
  data <- data[rep(seq(along = n), n), ]
  i <- NULL
  for(k in 1:length(n))
    i <- c(i, 1:n[k])
  data$events <- ifelse(data[, tn] > bounds[i + 1], 0, data[, dn])
  data$exposure <- ifelse(data[, tn] > bounds[i + 1], width[i], data[, tn
    ] - bounds[i])

  data$interval <- i
  attr(data$interval, "levels") <- attr(n, "levels")
  data
}

#Ejemplo:
exptramos(Fan, lifetime, censor, c(0, 44, 100, 150))
```

Pregunta 13

Solución

Se sabe que:

$$\lambda(t) = -d[\log S(t)]/dt$$

Por independencia entre T y U :

$$\begin{aligned}\lambda(t) &= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\{t \leq T < t + \Delta t\} [P\{T \geq t\}]^{-1} \\ &= \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\{t \leq T < t + \Delta t | T \geq t, U \geq t\}\end{aligned}$$

Luego,

$$P\{t \leq T < t + \Delta t | T \geq t, U \geq t\} = \lambda(t)\Delta t + o(\Delta t)$$

Considere $N(t-) = \lim_{s \uparrow t} N(s)$. Luego,

$$\lambda(t)\Delta t \approx P\{N((t + \Delta t)-) - N(t-) = 1 | T \geq t, U \geq t\}$$

Dada la naturaleza de la variable $N((t + \Delta t)-) - N(t-)$ (variable aleatoria 0-1), se tiene que:

$$\lambda(t)\Delta t \approx E\{N((t + \Delta t)-) - N(t-) = 1 | T \geq t, U \geq t\}$$

La función de hazard da una tasa promedio del cambio condicional del cambio en N sobre $[t, t + \Delta t]$, dado que ambos, la censura y los tiempos de falla exceden o igualan t , y por eso, indirectamente especifica la tasa condicional a la que N salta en pequeños intervalos.