

Estimación de la Función de Supervivencia con Datos de Riesgos en Competencia Bivariados Dependientes

Un Estudio Comparativo Utilizando el Estimador Cópula Gráfico

Por:

Hugo Alberto Brango García



UNIVERSIDAD NACIONAL DE COLOMBIA

FACULTAD DE CIENCIAS

ESCUELA DE ESTADÍSTICA

MEDELLÍN - ANTIOQUIA

FEBRERO DE 2011

Estimación de la Función de Supervivencia con Datos de Riesgos en Competencia Bivariados Dependientes

Un Estudio Comparativo Utilizando el Estimador Cópula Gráfico

Por:

Hugo Alberto Brango García

Presentado como requisito parcial para optar al título de
MAGISTER EN ESTADÍSTICA

Director:

Sergio Yáñez Canal

M.Sc Mathematics New York University

M.Sc Statistics Loughborough University

UNIVERSIDAD NACIONAL DE COLOMBIA

FACULTAD DE CIENCIAS

ESCUELA DE ESTADÍSTICA

MEDELLÍN - ANTIOQUIA

FEBRERO DE 2011

UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
ESCUELA DE ESTADÍSTICA

Los jurados abajo firmantes certifican que han leído y recomiendan a la **Facultad de Ciencias** aprobar el trabajo de grado titulado “**Estimación de la Función de Supervivencia con Datos de Riesgos en Competencia Bivariados Dependientes Un Estudio Comparativo Utilizando el Estimador Cópula Gráfico**” presentado por **Hugo Alberto Brango García** como requisito parcial para optar al título de **Magister en Estadística**.

Fecha: Febrero de 2011

Director:

Sergio Yáñez Canal
M.Sc Mathematics New York University
M.Sc Statistics Loughborough University

Jurados:

Carlos M. Lopera

Nelfi G. González



Universidad Nacional de Colombia
Sede Medellín
REGISTRO DE CALIFICACIONES
ESPECIALES

Código: M-FT-05 05.017.001
Versión: 0.0
Página: 1 de 1
Página: 1 de 1

01	2010
SEM	AÑO

FACULTAD: CIENCIAS			HABILITACIÓN	
ESCUELA - DEPARTAMENTO DE: ESTADÍSTICA				VALIDACIÓN
PROFESOR:	SERGIO YAÑEZ CANAL	CÓDIGO	GRUP	SUPLETORIOS
ASIGNATURA:	Tesis de maestría	3007980	0 1	

CÉDULA	APELLIDOS Y NOMBRES COMPLETOS DEL ESTUDIANTE	DEFINITIVA
10775518	BRANGO GARCIA HUGO ALBERTO	
	Título: "Estimación de la Función de Supervivencia con Datos de Riesgos en Competencia Bivariados Dependientes"	APROBADO

OBSERVACIONES:

La calificación corresponde a la sustentación pública de la Tesis para optar al título de Magister en Ciencias - Estadística.

Aprobado por Consejo de Facultad: Acta No: 05 Año: 2011 Mes: 03 Día: 28

Secretario Académico

Firma del Profesor

Vo.Bo. Director de Escuela - Departamento

NOTA: No se pueden mezclar calificaciones de conceptos diferentes. Si son calificaciones de habilitación, deben corresponder a estudiantes del mismo grupo y del mismo período académico. Si la calificación corresponde a un tipo no especificado arriba, debe hacerse constar en el espacio de observaciones

A mis padres
Milán Brango y Norma García

Índice general

1. Introducción	1
2. Riesgos competitivos	4
2.1. Identificabilidad	5
2.2. Estimación de la función de supervivencia bajo independencia	6
3. Estimador cópula gráfico	9
3.1. Cópulas	9
3.1.1. Medida de dependencia	11
3.1.1.1. Medidas de concordancia	11
3.2. El estimador Cópula-Gráfico	12
3.2.1. Construcción del estimador	13
4. Generación de tiempos de falla y esquema de simulación	18
4.1. Obtención de los tiempos de falla lognormales bivariados	18
4.2. Obtención de los tiempos de falla Weibull bivariados	19
4.3. Estudio de simulación	19
4.3.1. Esquema general de la metodología aplicada	19
5. Análisis de resultados	23
5.1. Intervalos de confianza empíricos para $S(t)$	23
5.1.1. Casos Weibull	23
5.1.2. Casos Lognormal	39

5.2. Comentarios sobre las marginales y la bondad de ajuste	51
5.2.1. Estimación marginal	51
5.2.2. Comentarios sobre la bondad de ajuste	52
5.2.2.1. Prueba para los tiempos Weibull y Lognormal con $n = 40$	53
5.2.2.2. Prueba para los tiempos Weibull y Lognormal con $n = 500$	54
6. Conclusiones y trabajo futuro	56
6.1. Trabajo futuro sobre predicción	56
A. Programa Cópula Gráfico	63

Índice de cuadros

5.1. Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0,5$; $\beta_2 = 2$ y $\theta_1 = \theta_2 = 1$	38
5.2. Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 2$; $\beta_2 = 1$ y $\theta_1 = \theta_2 = 1$	38
5.3. Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0,5$; $\beta_2 = 1$ y $\theta_1 = \theta_2 = 1$	39
5.4. Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0,5$; $\beta_2 = 0,5$ y $\theta_1 = \theta_2 = 1$	39
5.5. Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Lognormal con parámetros $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$	50
5.6. Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Lognormal con parámetros $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1, \sigma_2 = 2$	51
5.7. Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Lognormal con parámetros $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$	51
5.8. P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Weibull con tamaño de muestra $n = 40$	53
5.9. P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Lognormal con tamaño de muestra $n = 40$	54
5.10. P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $S^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Weibull con tamaño de muestra $n = 500$	54
5.11. P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Lognormal con tamaño de muestra $n = 500$	55

6.1.	Cambios porcentuales CP_p y $CP_p(\text{nuevo})$ para $S(t)$ y $S(t)_{\text{nuevo}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0.5$; $\beta_2 = 2$ y $\theta_1 = \theta_2 = 1$	58
6.2.	Cambios porcentuales CP_p y $CP_p(\text{nuevo})$ para $S(t)$ y $S(t)_{\text{nuevo}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0.5$; $\beta_2 = 1$ y $\theta_1 = \theta_2 = 1$	59
6.3.	Cambios porcentuales CP_p y $CP_p(\text{nuevo})$ para $S(t)$ y $S(t)_{\text{nuevo}}$ para tiempos de falla Weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\theta_1 = \theta_2 = 1$	60
6.4.	Cambios porcentuales CP_p y $CP_p(\text{nuevo})$ para $S(t)$ y $S(t)_{\text{nuevo}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 2$; $\beta_2 = 1$ y $\theta_1 = \theta_2 = 1$	61

Índice de figuras

5.1. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	24
5.2. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	25
5.3. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	25
5.4. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	26
5.5. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	26
5.6. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.9$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	27
5.7. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	28
5.8. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	28
5.9. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	29

5.10. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	29
5.11. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	30
5.12. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.9$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	30
5.13. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	31
5.14. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	32
5.15. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	32
5.16. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	33
5.17. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	33
5.18. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.9$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	34
5.19. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	35
5.20. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	35

5.21. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	36
5.22. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	36
5.23. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	37
5.24. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.9$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	37
5.25. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	40
5.26. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	41
5.27. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	41
5.28. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	42
5.29. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	42
5.30. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 1.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	43
5.31. Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	44

5.32. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	44
5.33. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	45
5.34. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	45
5.35. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	46
5.36. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 1.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	46
5.37. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	47
5.38. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	48
5.39. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	48
5.40. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	49
5.41. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	49
5.42. Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 1.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95	50

5.43. Intervalos de confianza empíricos aproximados del 95% para la distribución marginal $S_1(t_p)$ para el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95 52

Resumen

En esta investigación el objetivo central es estimar la función de sobrevivencia con datos de riesgos competitivos bivariados dependientes. Para estimar esta función para el tiempo mínimo de falla teniendo en cuenta la dependencia verdadera entre estos tiempos bivariados, se utilizó un modelo cópula mediante el estimador cópula gráfico, también se estimó la función de sobrevivencia como en la metodología tradicional asumiendo independencia entre los tiempos de falla. Estas estimaciones se compararon con la función de sobrevivencia verdadera, para ello se hacen simulaciones mediante las cuales se obtuvieron tiempos bivariados asociados a diferentes distribuciones Weibull y Lognormales, cada escenario de simulación contiene 6 situaciones diferentes, asociadas a 6 valores del parámetro de dependencia entre los tiempos de cada modo de falla. Se concluye acerca de la bondad del ajuste de los dos estimadores. También se estudian las marginales con los dos estimadores. En general se encontró, que el estimador cópula gráfico estima mejor la función de sobrevivencia en comparación al estimador tradicional que asume independencia.

Palabras clave: *Riesgos competitivos; Cópulas; Estimador cópula gráfico..*

Abstract

In this research the main objective is to estimate the function survival with data competing risks bivariate dependent. To estimate this function for the minimum time of failure taking into account the true dependence between these times was used copula model through of the copula graphic estimator, also estimated the survival function as in the traditional methodology that assumes independence between failure times. These estimates were compared with the true survival function, for this are make simulations which were obtained at different times associated to distributions Weibull and lognormal bivariate, each simulation scenario contains 6 different situations, associated with 6 values of the parameter of dependence between the times of each failure mode. We conclude about the goodness of fit of the two estimators. Also are studied the marginal with the two estimators. In general it was found that the copula graphic estimator estimated better the survival function compared to the traditional estimator than assumes independence.

Key words: *Competing risks; Copula; Copula Graphic estimator.*

Agradecimientos

El autor agradece a todos los profesores, compañeros de maestría y demás integrantes de la Escuela de Estadística. En especial agradece la asesoría del profesor Sergio Yáñez, así como los importantes aportes para el desarrollo de este trabajo de grado de los jurados y de los integrantes del grupo de investigación en Estadística Industrial de la Universidad Nacional, Sede Medellín.

Medellín, Antioquia
Febrero de 2011

Hugo Alberto Brango García

Capítulo 1

Introducción

Las diferentes maneras en que puede fallar un sistema o componente, reciben el nombre de modos de falla. Existen muchos sistemas, subsistemas y componentes que tienen más de un modo de falla; en ciertas aplicaciones y para algunos propósitos es importante distinguir entre esos diferentes modos de falla con el objeto de mejorar la confiabilidad (Meeker & Escobar, 1998). El tiempo de falla de un sistema con varios modos de falla puede ser modelado considerando un sistema en serie o un modelo de riesgos competitivos (“competing risks”), donde cada modo de falla es una componente del sistema en serie, y cada componente tiene un tiempo de vida aleatorio, por lo tanto el sistema fallará cuando el modo de falla con el tiempo de vida más corto falle.

Sean X y Y los tiempos de falla asociados a un problema de riesgos competitivos (“competing risks”) con dos modos de falla y supóngase que estamos interesados en estimar la función de supervivencia marginal de X , sin embargo, a menudo es imposible medir X debido a la ocurrencia de algún otro evento con tiempo Y . Si solamente estamos interesados en X entonces Y puede considerarse como censura del evento de interés. La información de la muestra observada es $T = \min(X, Y)$.

Un supuesto común para estimar la función de supervivencia de X es que el tiempo de censura Y es independiente de X . Bajo este supuesto se sabe que los datos observados proporcionan información suficiente para determinar de manera única la distribución marginal de X (Peterson, 1977). El supuesto de independencia entre X y Y no siempre es cierto, así la estimación de la distribución marginal de X se complica ya que la dependencia puede ser formulada de muchas maneras. En muchas situaciones prácticas la dependencia es una condición usual, lo que ha generado el creciente interés por su estudio.

Tsiatis (1975) mostró que las distribuciones marginales y la distribución conjunta de dos riesgos competitivos, son en general no identificables, esto es, hay muchas funciones de

distribución conjuntas diferentes que comparten las mismas funciones de sub-distribución (problema de identificabilidad). Si los riesgos son independientes, sus distribuciones marginales son identificables, estas marginales son las funciones de sub-distribución.

El problema clásico de riesgos competitivos es identificar las distribuciones marginales y las distribuciones conjuntas de los tiempos de falla asociados a los modos de falla a partir de datos de la forma (T, δ) , donde T es un tiempo de falla y δ es una variable discreta que representa el modo de falla, que toma valores de 0 cuando se presenta una censura a derecha, 1 cuando falla debido al primer modo, y 2 cuando falla debido al segundo modo de falla.

Yañez et al. (2006) y Manotas et al. (2008), estudiaron, para el caso Weibull y Lognormal, el efecto en la estimación de la sobrevivencia $S(t)$ (denotado $\hat{S}^*(t)$), cuando se asume el supuesto de independencia entre tiempos de falla en competencia que realmente son dependientes. Encontraron que el estimador $\hat{S}^*(t)$ bajo el supuesto de independencia, que es la metodología tradicional, no genera errores significativos en la estimación de la función de sobrevivencia del tiempo mínimo. Puesto que la literatura reciente reporta que el supuesto de independencia en riesgos competitivos generalmente da una visión pesimista del comportamiento del sistema, ellos comentan la importancia de estudiar en más detalle las propiedades del estimador $\hat{S}^*(t)$.

En este trabajo se estima $S(t)$ bajo dependencia mediante el estimador cópula gráfico propuesto en Zheng & Klein (1995). Se hace un estudio de simulación para comparar este estimador con el estimador de la confiabilidad $\hat{S}^*(t)$ propuesto en Yañez et al. (2006) y luego modificado por Manotas et al. (2008).

En el capítulo 2, se definen algunas nociones importantes de la teoría de riesgos en competitivos, el problema de identificabilidad y como se estima la confiabilidad $S(t)$ mediante la metodología tradicional asumiendo independencia entre dos modos de falla.

En el capítulo 3, se muestra la teoría de los modelos cópulas y cómo se emplea esta para modelar la estructura de dependencia para los tiempos de falla bivariados y solucionar así el problema de identificabilidad. También, se describe el algoritmo cópula gráfico con el que se estima la función de sobrevivencia para el tiempo mínimo de falla en un sistema con dos modos de falla.

En el capítulo 4 se describe la metodología utilizada en este estudio para la obtención de los resultados gráficos y analíticos, que permitieron desarrollar los objetivos propuestos.

En el capítulo 5 se presentan los resultados analíticos asociados a cada uno de los casos Weibull y Lognormal. Se describe el comportamiento de la estimación de la función de sobrevivencia mediante el modelo cópula para los distintos valores de dependencia en las

dos distribuciones bivariadas estudiadas. Se presentan intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ basados en la distribución empírica de $\widehat{S}^*(t_p)$ y $\widehat{S}(t_p)_{\text{Copula}}$ con 1000 réplicas de cada uno de estimadores, para $p = 0.05, 0.25, 0.50, 0.75, 0.95$. De esta manera se obtienen intervalos de confianza puntuales aproximados del 95 % las cuales se comparan con la sobrevivencia verdadera. De aquí se concluye sobre la bondad de ajuste de los dos estimadores. También se estudian las marginales con los dos estimadores.

El capítulo 6 presenta las conclusiones del estudio: En general se encontró que $\widehat{S}(t)_{\text{copula}}$ es mejor estimador para la función de sobrevivencia verdadera que $\widehat{S}^*(t)$ y un tópico que se propone como trabajo futuro.

Capítulo 2

Riesgos competitivos

El tiempo de falla de un sistema con dos modos de falla puede ser modelado como un sistema en serie o un modelo de riesgos competitivos. Cada unidad tiene un tiempo potencial de falla asociado a cada modo de falla. El tiempo de falla observado es el mínimo de esos tiempos potenciales individuales.

Por ejemplo, para un sistema con dos modos de falla, sean X y Y los respectivos tiempos potenciales, entonces el tiempo de falla observado es

$$T = \text{mín}(X, Y) \quad (2.1)$$

De esta manera en la práctica sólo se observa el mínimo entre X y Y pero se conoce cuál es el modo de falla. Así los datos son parejas de la forma (T, δ) , donde $T = \text{mín}(X, Y)$ y δ es una variable discreta que representa el modo de falla, que toma valores de 0 cuando se presenta una censura a derecha, 1 cuando falla debido al primer modo de falla, y 2 cuando falla debido al segundo modo de falla.

Para estudiar este modelo de riesgos competitivos, es necesario presentar algunas funciones básicas.

■ La CIF ó función de Subdistribución

Para un evento de tipo i ($i = 1, 2$) la CIF es definida como la probabilidad conjunta:

$$G_i(t) = P(T \leq t, \delta = i) \quad (2.2)$$

En otras palabras la CIF es la probabilidad que un evento de tipo i ocurra hasta el tiempo t . La función de distribución conjunta es la probabilidad que un evento de

cualquier tipo ocurra hasta el tiempo t , es decir es igual a la suma de las CIF's para todos los tipos de eventos, así:

$$F(t) = P(T \leq t) = \sum_{i=1}^2 P(T \leq t, \delta = i) = \sum_{i=1}^2 G_i(t) \quad (2.3)$$

Nótese que cuando hay riesgos competitivos la función de distribución total se encuentra en el intervalo $[0; 1]$. En contraste, en presencia de los riesgos competitivos la CIF sólo toma valores hasta $P(\delta = i)$ debido a que

$$\lim_{t \rightarrow \infty} G_i(t) = P(\delta = i) \quad (2.4)$$

■ Función de Subsobrevivencia

Es la probabilidad que un evento de tipo i no ocurra hasta el tiempo t es decir, es definida como la probabilidad de que un evento de tipo i no haya sido observado hasta el tiempo t :

$$\bar{G}_i(t) = P(T > t, \delta = i) \quad (2.5)$$

■ Función de Subdensidad

La función de subdensidad para el modo de falla i y T continuo, se define como

$$g_i(t) = \frac{dG_i(t)}{dt} = -\frac{d\bar{G}_i(t)}{dt} \quad (2.6)$$

2.1. Identificabilidad

La aproximación tradicional para especificar el modelo de riesgos competitivos es vía tiempos de falla latentes. En este contexto surge el denominado problema de identificabilidad. Para el caso bivariado tenemos dos tiempos de falla potenciales X, Y asociados a dos modos de falla, de manera que si ocurre el primero el segundo no puede ser observado, en este sentido se dice que los tiempos son latentes. Se supone que la pareja (X, Y) tiene una distribución multivariada de la forma $F(x, y) = P(X \leq x, Y \leq y)$. Así $T = \min(X, Y)$ determina el tiempo de falla del sistema completo. De manera que una vez que el sistema haya fallado, el otro tiempo de falla no puede ser observado. Las funciones de distribución marginales para los dos modos de falla son $F_1(t) = P(X \leq t)$ y $F_2(t) = P(Y \leq t)$, respectivamente.

El problema de identificabilidad (Tsiatis, 1975) establece que para el caso de modelos de tiempos de falla latentes, dada una función de distribución conjunta con dependencia arbitraria entre las componentes, existe siempre una función de distribución conjunta en la cual las variables son independientes, y la cual reproduce las subdensidades $g_i(t)$ de manera precisa. Así, uno no puede conocer a partir solamente de los datos observados (T, δ) , cual de los dos modelos es correcto, puesto que ambos se ajustan muy bien a los datos. Crowder (2001) da detalles técnicos sobre este problema que él denomina “El impase de Cox-Tsiatis”. En resumen podemos decir que en la práctica se pueden encontrar las subdistribuciones marginales, asociados con datos de riesgos competitivos, y ese conjunto es consistente con un número infinito de distribuciones conjuntas de tiempos de falla potenciales, esto quiere decir, que es imposible identificar las distribuciones marginales a partir solamente de los datos de riesgos competitivos.

Supongamos que se tienen datos de la forma (T, δ) y es necesario identificar el modelo conjunto $F(x, y)$. Algunas posibles soluciones son: i) Asuma independencia, lo cual es frecuentemente no realista ; ii) Asuma una estructura de dependencia conocida usando una cópula específica para la distribución conjunta de los modos de falla; Zheng & Klein (1995) proponen el llamado estimador cópula-gráfico, el cual es un estimador máximo verosímil y trabaja muy bien independiente de la cópula escogida (ver sección 2.4); iii) Especifique un modelo paramétrico. En este último caso a pesar del fuerte supuesto, los datos de sobrevivencia pueden contener poca información acerca de la asociación entre riesgos competitivos debido, a veces, a la alta censura y a la cantidad limitada de datos. Meeker & Escobar (2009) muestran un caso de este tipo donde el problema de identificabilidad prevalece.

En este trabajo se van a comparar las soluciones i) y ii). También como trabajo futuro se propone, incorporar los resultados de este trabajo a la solución iii).

2.2. Estimación de la función de sobrevivencia bajo independencia

Una estructura en serie con s componentes funciona si y solo si, todas las componentes funcionan. Para un sistema en serie con dos componentes independientes, la función de distribución para el tiempo mínimo T es

$$\begin{aligned}
F(t) &= (T \leq t) = 1 - P(T > t) \\
&= (T \leq t) = 1 - P(T > t) \\
&= 1 - P(X > t)P(Y > t) \\
&= 1 - [1 - F_1(t)][1 - F_2(t)] = F_1(t) + F_2(t) - F_1(t)F_2(t)
\end{aligned} \tag{2.7}$$

donde (X, Y) son los tiempos de vida asociados a las dos componentes y F_i , $i = 1, 2$ son respectivamente las funciones de distribución para los componentes y $T = \min(X, Y)$ es el tiempo de falla observado del sistema.

La función de sobrevivencia $S(t) = 1 - F(t)$, para el caso de independencia es

$$S(t) = P(T > t) = P(X > t)P(Y > t) = S_1(t)S_2(t) \tag{2.8}$$

donde $S_i(t)$, $i = 1, 2$ son las funciones de sobrevivencia para los dos componentes.

Para un sistema con dos componentes en serie y tiempos de falla dependientes

$$F(t) = P(T \leq t) = 1 - P(T > t) = 1 - P(X > t, Y > t) \tag{2.9}$$

En este caso, la evaluación debe hacerse con respecto a la distribución conjunta de X y Y , que incluye los parámetros de dependencia.

La función de sobrevivencia para este caso es

$$S(t) = P(T > t) = P(X > t, Y > t) \tag{2.10}$$

Para modos de falla en competencia con dependencia positiva se sabe que (Barlow y Proschan, 1975),

$$S(t) = P(X > t, Y > t) \geq P(X > t)P(Y > t) = S^*(t) = S_1(t)S_2(t) \tag{2.11}$$

donde $S^*(t)$ es la función de sobrevivencia asumiendo independencia. Por lo tanto $S^*(t)$ es pesimista en relación con $S(t)$, esto es para un mismo tiempo t , la probabilidad de sobrevivencia bajo el supuesto de independencia está siempre por debajo de la probabilidad conjunta que incluye la dependencia.

El estimador de $S^*(t)$ se denotará $\widehat{S}^*(t)$, que es un estimador de la sobrevivencia del tiempo mínimo de falla de un sistema con dos modos de falla que compiten. Este se obtiene como el producto de las funciones de sobrevivencia estimadas bajo una distribución específica,

donde para cada función marginal se estiman los parámetros, considerando los tiempos del otro modo de falla como tiempos de censura, así:

$$\widehat{S}^*(t) = \widehat{S}_1(t)\widehat{S}_2(t) \quad (2.12)$$

donde $\widehat{S}_1(t)$ y $\widehat{S}_2(t)$ se suponen son las funciones de supervivencia marginales estimadas para cada modo de falla.

Capítulo 3

Estimador cópula gráfico

3.1. Cópulas

Los modelos cópula son clases de distribuciones de sobrevivencia bivariada, especificada en términos de las funciones de sobrevivencia marginal y una función cópula, que es una función de distribución continua bivariada sobre el cuadrado unitario $[0,1]$ con marginales uniformes. Una característica de la clase cópula es que la determinación de las marginales a través de la cópula ayuda a modelar y entender eficazmente la estructura de dependencia, (Escarela et al. (2003)).

En el contexto de riesgos competitivos, es posible especificar la distribución de sobrevivencia conjunta en términos de dos distribuciones de sobrevivencia marginal y una cópula que permite relacionar la dependencia entre las variables aleatorias individuales que corresponden a cada riesgo de falla, (Escarela et al. (2003)). Así, sea la cópula de sobrevivencia C_α con $\alpha \in \mathbb{R}$ para dos variables aleatorias X y Y , donde α representa el parámetro de dependencia entre estas dos variables.

Si (X, Y) proviene de una cópula para algún α , entonces la función de sobrevivencia conjunta de (X, Y) está dada por

$$S(t) = C_\alpha [S_1(t), S_2(t)] \quad (3.1)$$

Teorema 1. (Teorema de Sklar). Sean X, Y variables aleatorias con función de distribución conjunta F , con marginales F_1 y F_2 respectivamente. Entonces existe una cópula C que satisface

$$F(x, y) = C[F_1(x), F_2(y)] \quad (3.2)$$

para toda $y_1, y_2 \in \mathbb{R}$. Si F_1, F_2 son continuas, entonces C es única. Inversamente, si C es

una cópula y F_1, F_2 son funciones de distribución, entonces F definida en la ecuación (3.2) es una función de distribución conjunta con marginales F_1 y F_2 .

Este teorema establece que, en el contexto de parejas aleatorias continuas, es posible construir una función de distribución bivariada en términos de dos funciones de distribución continuas univariadas y una cópula que permite relaciones de dependencia entre dos variables aleatorias individuales. Una demostración del Teorema de Sklar puede encontrarse en Schweizer & Sklar (1983).

Corolario 1. Dada una función de distribución conjunta F con marginales continuas F_1 y F_2 , como está indicado en el teorema de Sklar, es fácil construir la cópula correspondiente como se muestra a continuación:

$$C(v_1, v_2) = F(F_1^{-1}(v_1), F_2^{-1}(v_2))$$

donde $F_j^{(-1)}$ es la función inversa de F_j , dada por $F_j[F_j^{-1}(u)] = u$ si $u \in \text{rango } F_j$, o por $F_j^{-1}(u) = \sup\{z | F_j(z) \leq u\}$ si $u \notin \text{rango } F_j$, para $j = 1, 2$; aquí las inversas se usan para funciones de distribución no estrictamente crecientes. Nótese que si X y Y son variables aleatorias continuas con funciones de distribución F_1 y F_2 , respectivamente, entonces C es la función de distribución conjunta de $V_1 = F_1(X)$ y $V_2 = F_2(Y)$ ya que $F_1(X)$ y $F_2(Y)$ se distribuyen uniformemente en \mathbf{I} .

Nelsen (1999), la desigualdad de las cotas Fréchet-Hoeffding indica que si F es una función de distribución bivariada con marginales F_1 y F_2 , entonces

$$\text{máx}\{F_1(x) + F_2(y) - 1, 0\} \leq F(x, y) \leq \text{mín}\{F_1(x), F_2(y)\}$$

Este resultado es consecuencia del teorema de Sklar. En términos de cópulas, la desigualdad puede expresarse como (Fréchet 1951).

$$W(v_1, v_2) = \text{máx}\{v_1 + v_2 - 1, 0\} \leq C(y_1, y_2) \leq \text{mín}\{v_1, v_2\} = M(v_1, v_2)$$

la cual se conoce como desigualdad de las cotas de Fréchet e indica que cualquier cópula C representa un modelo de dependencia que se encuentra entre los extremos W y M . Las funciones W y M son conocidas como las cotas inferior y superior, respectivamente; de hecho, Fréchet (1951) demuestra que W y M son también cópulas.

3.1.1. Medida de dependencia

Cuando los modelos cópula son usados para construir una distribución conjunta de una pareja aleatoria continua, pueden verse como versiones de funciones de distribución conjuntas libres de marginales que tienen la poderosa habilidad de capturar propiedades de dependencia invariante al reescalamiento de las parejas aleatorias; aquí, invariante al reescalamiento significa que las propiedades y las medidas se quedan sin cambiar cuando se realizan transformaciones estrictamente crecientes a las variables aleatorias. Las medidas de asociación invariantes bajo reescalamiento más conocidas son τ de Kendall y ρ de Spearman, las cuales miden una forma de dependencia conocida como concordancia.

3.1.1.1. Medidas de concordancia

Cuando se considera una pareja de variables aleatorias, es útil saber qué tanto tienden a estar asociados valores grandes de una de las variables aleatorias con valores grandes de la otra, y que tanto están asociados valores pequeños de una con valores pequeños de la otra. Una formalización de la idea intuitiva de este grado de asociación fue propuesta por Yanagimoto & Okamoto (1969), quienes proponen el uso del orden de concordancia de distribuciones bivariadas con marginales univariadas dadas de acuerdo con la fuerza de su asociación positiva. Este orden estocástico se define a continuación.

Definición. Sean (X_1, Y_1) y (X_2, Y_2) vectores aleatorios independientes e idénticamente distribuidos, tales que $(X_i, Y_i) \sim F, i = 1, 2$. Entonces, la τ de Kendall se define como

$$\begin{aligned}\tau &= E[\text{signo}\{(X_1 - X_2)(Y_1 - Y_2)\}] \\ &= P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\} \\ &= P\{\text{concordancia}\} - P\{\text{disconcordancia}\} \\ &= 2P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - 1 \\ &= 2P\{\text{concordancia}\} - 1\end{aligned}$$

La versión de la τ de Kendall de las entradas de una pareja aleatoria continua (Y_1, Y_2) , dada en términos de la cópula C , puede expresarse como

$$\tau = 4 \int_0^1 \int_0^1 C(v_1, v_2) dC(v_1, v_2) - 1 \quad (3.3)$$

La ecuación (3.3) indica que la τ de Kendall está completamente determinada por la cópula y no está relacionada con las distribuciones marginales de (Y_1, Y_2) .

3.2. El estimador Cópula-Gráfico

Zheng & Klein (1995) proponen un estimador no paramétrico para las distribuciones marginales, al cual denominan estimador cópula gráfico como una solución al problema de identificabilidad, descrito en la sección 2.1. Para ello utilizan cópulas como una función no paramétrica que captura la dependencia entre dos variables aleatorias. El problema de identificabilidad es resuelto bajo el supuesto que la cópula entre X y Y y su parámetro de asociación son conocidos y con los datos de riesgos competitivos disponibles.

Suponga que se tiene, en el contexto de riesgos competitivos, una muestra de tamaño n de tiempos de vida $T_j = \min(X_j, Y_j)$. Si $F_1(t)$ y $F_2(t)$ son las distribuciones marginales de X y Y , entonces para cualquier cópula, tenemos

$$\mu_c(A_t) = P(X > t, Y > t) \quad (3.4)$$

donde

$$\mu_c(A_t) = \int_{I^2} \mathbf{1}_{E_t} \{F_1^{-1}(x), F_2^{-1}(y)\} dC(x, y)$$

con $E_t = \{(x, y) : x > t, y > t\}$ y $A_t = \{(x, y) | F_1(t) < x \leq 1, F_2(t) < y \leq 1\}$

$$\mu_c(B_t) = P(X \leq t, X < Y) \quad (3.5)$$

donde

$$\mu_c(B_t) = \int_{I^2} \mathbf{1}_{D_t} \{F_1^{-1}(x), F_2^{-1}(y)\} dC(x, y)$$

con $D_t = \{(x, y) : x < y, 0 < x < t\}$ y $B_t = \{(x, y) | 0 < x \leq F_1(t), F_2 F_1^{-1}(t) < y \leq 1\}$

Estas dos relaciones determinan de manera única a $F_1(t)$ y $F_2(t)$. Por tanto se pueden hallar estimadores $\hat{F}_1(t)$ y $\hat{F}_2(t)$ respectivamente que preservan estas relaciones sobre una rejilla seleccionada de m tiempos distintos de falla o censura $0 < t_1 < t_2 < \dots < t_m < \max\{T_j, j = 1, \dots, n\}$.

Para construir nuestro estimador, sean

$$\hat{A}_t = \{(x, y) | \hat{F}_1(t) < x \leq 1, \hat{F}_2(t) < y \leq 1\}, \hat{B}_t = \{(x, y) | 0 < x \leq \hat{F}_1(t), \hat{F}_2 \hat{F}_1^{-1}(t) < y \leq 1\}$$

Adicionalmente, sean

$\widehat{P}(X > t, Y > t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(T_i > t)$, $\widehat{P}(X \leq t, X < Y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(T_i \leq t, \delta_i = 1)$ las estimaciones empíricas de $P(X > t, Y > t)$ y $P(X \leq t, X < Y)$.

Finalmente se encuentra $\widehat{F}_1(t_i)$ como la raíz de (3.7) sujeto a la definición de $\widehat{F}_2(t_i)$ como una función $\widehat{F}_1(t_i)$ resolviendo la ecuación (3.6). Siendo \widehat{F}_1 y \widehat{F}_2 funciones de escalonadas en cada intervalo $(t_i, t_{i+1}]$

$$\mu_c(\widehat{A}_{t_i}) - \widehat{P}(X > t_i, Y > t_i) = 0 \quad (3.6)$$

$$\mu_c(\widehat{B}_{t_i}) - \widehat{P}(X \leq t_i, X < Y) = 0 \quad (3.7)$$

A continuación se describe el algoritmo cópula gráfico para la construcción de \widehat{F}_1 y \widehat{F}_2 , basado en un algoritmo de bisección para encontrar raíces.

3.2.1. Construcción del estimador

Por propiedades de función de supervivencia conjunta se sabe que:

$$P(X > t, Y > t) = 1 - P(X \leq t) - P(Y \leq t) + P(X \leq t, Y \leq t)$$

Donde $P(X > t, Y > t) = \mu_c(A_t) = S(t)$, las marginales $P(X \leq t) = F_1(t)$, $P(Y \leq t) = F_2(t)$ y la función de distribución conjunta $P(X \leq t, Y \leq t) = C(F_1(t), F_2(t))$ es la cópula. Así,

$$\mu_c(A_t) = 1 - F_1(t) - F_2(t) + C(F_1(t), F_2(t))$$

Un estimador sería

$$\mu_c(\widehat{A}_{t_i}) = 1 - \widehat{F}_1(t_i) - \widehat{F}_2(t_i) + C(\widehat{F}_1(t_i), \widehat{F}_2(t_i)) = \widehat{P}(X > t_i, Y > t_i)$$

Para construir el estimador cópula gráfico seleccionamos una rejilla tomando $t_1 < t_2 < \dots < t_m$ que sean tiempos de fallas o censuras. Aquí m es el número de tales tiempos distintos. Con esta rejilla y usando una función step en cada intervalo $(t_{i-1}, t_i]$ el estimador es más fácil de calcular. Para este estimador si $\delta_i = 1$ entonces $\widehat{F}_2(t_i) = \widehat{F}_2(t_{i-1})$ mientras si $\delta_i = 2$ entonces $\widehat{F}_1(t_i) = \widehat{F}_1(t_{i-1})$. Para el tiempo $t_0 = 0$ se tiene que $\widehat{F}(t_0) = \widehat{G}(t_0) = 0$

Tenemos que, para $\delta_i = 1$

$$\mu_c(\widehat{A}_{t_i}) = 1 - \widehat{F}_1(t_i) - \widehat{F}_2(t_{i-1}) + C\left(\widehat{F}_1(t_i), \widehat{F}_2(t_{i-1})\right) = \frac{1}{n} \sum_{i=1}^m I(T > t_i) \quad (3.8)$$

y para $\delta_i = 2$

$$\mu_c(\widehat{A}_{t_i}) = 1 - \widehat{F}_1(t_{i-1}) - \widehat{F}_2(t_i) + C\left(\widehat{F}_1(t_{i-1}), \widehat{F}_2(t_i)\right) = \frac{1}{n} \sum_{i=1}^m I(T > t_i) \quad (3.9)$$

Entonces $\widehat{F}_1(t_i)$ y $\widehat{F}_2(t_i)$ se encuentran resolviendo (3.8) y (3.9) iterativamente. Donde $C(\widehat{F}_1, \widehat{F}_2)$ es la cópula que relaciona las marginales estimadas \widehat{F}_1 y \widehat{F}_2 .

Para estimar F_1 y F_2 en este trabajo seleccionamos la cópula gamma

$$C(F_1, F_2) = F_1 + F_2 - 1 + \left\{ \left(\frac{1}{1-F_1} \right)^{\alpha-1} + \left(\frac{1}{1-F_2} \right)^{\alpha-1} - 1 \right\}^{-\frac{1}{\alpha-1}} \quad (\alpha > 1) \quad (3.10)$$

donde F_1 y F_2 son las funciones de distribución marginales para los tiempos de falla X y Y respectivamente, α es el parámetro de dependencia de la cópula, este parámetro relaciona los tiempos de falla X y Y mediante la correlación τ de Kendall's así: $\tau = \frac{1}{\alpha-1}$.

Se seleccionó la cópula gamma, puesto que permite variables con rango $(0, \infty)$ como es el caso de las distribuciones Weibull y Lognormal. Además, Zheng & Klein (1995) muestran que el estimador cópula gráfico es razonablemente robusto a la escogencia de la cópula, es decir que esta escogencia no afecta significativamente la estimación de la función de sobrevivencia

El algoritmo para construir \widehat{F}_1 y \widehat{F}_2 , sugerido por Zheng & Klein (1995) es como sigue:

Paso 1. Para $i = 1$, dado un punto inicial $\widehat{F}_1(t_1)$, digamos $\widehat{F}_1^{(1)}(t_1)$, encontrar $\widehat{F}_2^{(1)}(t_1)$ resolviendo (3.6).

Observación del paso 1

Para el caso de la cópula seleccionada, si $\delta_i = 2$ entonces $\widehat{F}_1(t_i) = \widehat{F}_1(t_{i-1})$ y el valor de $\widehat{F}_2(t_i)$ en función de $\widehat{F}_1(t_i)$ es:

$$\widehat{F}_2(t_i) = 1 - \left\{ 1 + \left[\frac{1}{n} \sum_{i=1}^m I(T > t_i) \right]^{1-\alpha} - \left(\frac{1}{1-\widehat{F}_1(t_{i-1})} \right)^{\alpha-1} \right\}^{\frac{1}{1-\alpha}} \quad (3.11)$$

La deducción de este despeje es como sigue, sabemos que

$$1 - \widehat{F}_1(t_{i-1}) - \widehat{F}_2(t_i) + C\left(\widehat{F}_1(t_{i-1}), \widehat{F}_2(t_i)\right) = \frac{1}{n} \sum_{i=1}^m I(T > t_i)$$

Reemplazando la cópula gamma tenemos

$$\begin{aligned}
1 - \widehat{F}_1(t_{i-1}) - \widehat{F}_2(t_i) + \widehat{F}_1(t_{i-1}) + \widehat{F}_2(t_i) - 1 + \left\{ \left(\frac{1}{1 - \widehat{F}_1(t_{i-1})} \right)^{\alpha-1} + \left(\frac{1}{1 - \widehat{F}_2(t_i)} \right)^{\alpha-1} - 1 \right\}^{-\frac{1}{\alpha-1}} &= \frac{1}{n} \sum_{i=1}^m I(T > t_i) \\
\left\{ \left(\frac{1}{1 - \widehat{F}_1(t_{i-1})} \right)^{\alpha-1} + \left(\frac{1}{1 - \widehat{F}_2(t_i)} \right)^{\alpha-1} - 1 \right\}^{-\frac{1}{\alpha-1}} &= \frac{1}{n} \sum_{i=1}^m I(T > t_i) \\
\left(\frac{1}{1 - \widehat{F}_1(t_{i-1})} \right)^{\alpha-1} + \left(\frac{1}{1 - \widehat{F}_2(t_i)} \right)^{\alpha-1} - 1 &= \left[\frac{1}{n} \sum_{i=1}^m I(T > t_i) \right]^{1-\alpha} \\
\left(\frac{1}{1 - \widehat{F}_2(t_i)} \right)^{\alpha-1} &= 1 + \left[\frac{1}{n} \sum_{i=1}^m I(T > t_i) \right]^{1-\alpha} - \left(\frac{1}{1 - \widehat{F}_1(t_{i-1})} \right)^{\alpha-1} \\
1 - \widehat{F}_2(t_i) &= \left\{ 1 + \left[\frac{1}{n} \sum_{i=1}^m I(T > t_i) \right]^{1-\alpha} - \left(\frac{1}{1 - \widehat{F}_1(t_{i-1})} \right)^{\alpha-1} \right\}^{\frac{1}{1-\alpha}} \\
\widehat{F}_2(t_i) &= 1 - \left\{ 1 + \left[\frac{1}{n} \sum_{i=1}^m I(T > t_i) \right]^{1-\alpha} - \left(\frac{1}{1 - \widehat{F}_1(t_{i-1})} \right)^{\alpha-1} \right\}^{\frac{1}{1-\alpha}}
\end{aligned}$$

De igual forma, si $\delta_i = 1$ entonces $\widehat{F}_2(t_i) = \widehat{F}_2(t_{i-1})$ donde el valor de $\widehat{F}_2(t_{i-1})$ en función de $\widehat{F}_1(t_i)$ debe satisfacer

$$\widehat{F}_2(t_{i-1}) = 1 - \left\{ 1 + \left[\frac{1}{n} \sum_{i=1}^m I(T > t_i) \right]^{1-\alpha} - \left(\frac{1}{1 - \widehat{F}_1(t_i)} \right)^{\alpha-1} \right\}^{\frac{1}{1-\alpha}}$$

Paso 2. Si el punto $(\widehat{F}_1^{(1)}(t_1), \widehat{F}_2^{(1)}(t_1))$ satisface (3.7), vaya al Paso 4. En otro caso use (3.7) para decidir si el siguiente valor, $\widehat{F}_1^{(2)}(t_1)$ es mayor o menor que $\widehat{F}_1^{(1)}(t_1)$. Use el punto medio de los intervalos $(\widehat{F}_1^{(1)}(t_1), 1)$ o $(0, \widehat{F}_1^{(1)}(t_1))$ respectivamente, como el valor de $\widehat{F}_1^{(2)}(t_1)$.

Observación del paso 2

En la ecuación (3.7) se tiene que $\mu_c(\widehat{B}_t) - \widehat{P}(X \leq t, X < Y) = 0$, donde:

$$\begin{aligned}
\mu_c(\widehat{B}_t) &= \widehat{F}_1(t) - \int_0^t dFC(\widehat{F}_1, \widehat{F}_2) \\
&= \widehat{F}_1(t) - \int_0^t C_{F_1}(\widehat{F}_1, \widehat{F}_2) d\widehat{F}_1 dt \\
&= \widehat{F}_1(t) - \int_0^t C_{F_1}(\widehat{F}_1, \widehat{F}_2) f_1(t) dt
\end{aligned}$$

donde $C_{F_1}(\widehat{F}_1, \widehat{F}_2)$ es la derivada parcial de la cópula respecto a F_1 , esto es:

$$\begin{aligned} C_{F_1}(\widehat{F}_1, \widehat{F}_2) &= \frac{\partial C(\widehat{F}_1, \widehat{F}_2)}{\partial \widehat{F}_1} \\ &= 1 + \left(\frac{1}{1-\alpha}\right) \left\{ \left(\frac{1}{1-\widehat{F}_1}\right)^{\alpha-1} + \left(\frac{1}{1-\widehat{F}_2}\right)^{\alpha-1} - 1 \right\}^{-\frac{1}{\alpha-1}-1} (1-\alpha) (1-\widehat{F}_1)^{-\alpha} (-1) \\ &= 1 - (1-\widehat{F}_1)^{-\alpha} \left\{ \left(\frac{1}{1-\widehat{F}_1}\right)^{\alpha-1} + \left(\frac{1}{1-\widehat{F}_2}\right)^{\alpha-1} - 1 \right\}^{\frac{\alpha}{1-\alpha}} \end{aligned}$$

$$\Rightarrow \mu(\widehat{B}_i) = \widehat{F}(t_i) - \int_0^{t_i} \left\{ 1 - (1-\widehat{F}_1(t_i))^{-\alpha} \left[\left(\frac{1}{1-\widehat{F}_1(t_i)}\right)^{\alpha-1} + \left(\frac{1}{1-\widehat{F}_2(t_i)}\right)^{\alpha-1} - 1 \right]^{\frac{\alpha}{1-\alpha}} \right\} f_1(t_i) dt_i$$

Como se trata de un estimador no paramétrico entonces (3.7) se puede escribir de la siguiente forma

$$\widehat{F}_1(t_i) - \sum_{i \geq 0} \left\{ 1 - (1-\widehat{F}_1(t_i))^{-\alpha} \left[\left(\frac{1}{1-\widehat{F}_1(t_i)}\right)^{\alpha-1} + \left(\frac{1}{1-\widehat{F}_2(t_i)}\right)^{\alpha-1} - 1 \right]^{\frac{\alpha}{1-\alpha}} \right\} [\widehat{F}(t_i) - \widehat{F}(t_{i-1})] - \widehat{P}(X \leq t, X < Y) = 0 \quad (3.12)$$

Paso 3. Repita los pasos 1 y 2 usando la estimación actual de $\widehat{F}_1(t_1)$ y $\widehat{F}_2(t_1)$. Después de i pasos el nuevo valor de $\widehat{F}_1(t_1)$ es el punto medio de los intervalos $(a_i, \widehat{F}_1^{(i)}(t_i))$ o $(\widehat{F}_2^{(i)}(t_i), b_i)$, donde

$$a_i = \max \left\{ \widehat{F}_1^{(1)}(t_1) \mid k < i, \widehat{F}_1^{(k)}(t_1) < \widehat{F}_1^{(i)}(t_i) \right\}, b_i = \min \left\{ \widehat{F}_1^{(k)}(t_1) \mid k < i, \widehat{F}_1^{(k)}(t_1) > \widehat{F}_1^{(i)}(t_i) \right\}.$$

Este proceso continúa hasta que encontremos $\widehat{F}_1(t_1)$ y $\widehat{F}_2(t_1)$ que satisfacen (3.6) y (3.7). Esto indica que la convergencia del algoritmo es garantizada.

En este trabajo la convergencia es garantizada si el valor absoluto de la ecuación 2.18 es menor que 1×10^{-6} .

Paso 4. Repita los Pasos (1-3) para $i = 2, \dots, m$. Para $t > \max \{T_i\}$, definimos $\widehat{F}_1(t)$ y $\widehat{F}_2(t)$ que serán $\widehat{F}_1(t_m)$ y $\widehat{F}_2(t_m)$ respectivamente.

Zheng & Klein (1995) sugieren que combinando los datos observados con una cópula asumida nos proporcionará una estimación significativamente igual a la función de supervivencia marginal y prueban que el estimador cópula gráfico es un estimador de máxima verosimilitud y bajo condiciones de independencia coincide con el estimador de Kaplan-Meier.

Con la aplicación del anterior algoritmo encontramos los estimadores de las marginales \widehat{F}_1 y \widehat{F}_2 , y con estas marginales se calcula el estimador de la sobrevivencia con la cópula gamma así

$$\widehat{S}(t)_{\text{copula}} = \left\{ \left(\frac{1}{1 - \widehat{F}_1} \right)^{\alpha-1} + \left(\frac{1}{1 - \widehat{F}_2} \right)^{\alpha-1} - 1 \right\}^{-1/\alpha-1} \quad (3.13)$$

En el Apéndice A se muestra el programa en R para la obtención de \widehat{F}_1 y \widehat{F}_2 mediante el algoritmo cópula gráfico.

Capítulo 4

Generación de tiempos de falla y esquema de simulación

Para realizar este estudio se simulan tiempos de falla Weibull y Lognormal, los cuales se pueden visualizar en varios contextos, tales como, los tiempos hasta la primera y segunda falla de un equipo reparable, los tiempos de falla “breakdown” de generadores duales en una planta de energía, o los tiempos de sobrevivencia en un sistema de dos órganos tal como los pulmones o riñones en el cuerpo humano (Lu & Bhattacharyya, 1990).

4.1. Obtención de los tiempos de falla lognormales bivariados

Para obtener tiempos de falla lognormales bivariados se aplica el algoritmo expuesto en Johnson (1987). Donde se explica que si se tienen variables aleatorias normales bivariadas, al aplicar función exponencial a esas variables, se obtienen variables aleatorias lognormales bivariadas. Esto es, si Z y W son variables aleatorias normales bivariadas, entonces $X = \exp(Z)$ y $Y = \exp(W)$, son variables aleatorias lognormales bivariadas.

Así, para simular los tiempos de falla Lognormal bivariados con coeficiente de correlación ρ_L , se generan normales bivariadas con coeficiente de correlación ρ_N , donde ρ_L está definido mediante la siguiente relación:

$$\rho_L = \frac{\exp(\rho_N \sigma_1 \sigma_2) - 1}{\{[\exp(\sigma_1^2) - 1][\exp(\sigma_2^2) - 1]\}^{\frac{1}{2}}} \quad (4.1)$$

y luego se exponencia cada uno de los tiempos normales bivariados.

4.2. Obtención de los tiempos de falla Weibull bivariados

Para simular los tiempos de falla Weibull bivariados se usó el algoritmo cópula compuesta de Gumbel-Hougaard (Frees y Valdes, 1997). La función de supervivencia conjunta de la Weibull bivariada utilizada es (Lu y Bhattacharyya, 1990)

$$S(x,y) = \exp \left\{ - \left[\left(\frac{x}{\theta_1} \right)^{\frac{\beta_1}{1-\lambda}} + \left(\frac{y}{\theta_2} \right)^{\frac{\beta_2}{1-\lambda}} \right]^{1-\lambda} \right\} \quad (4.2)$$

donde $T_1 > 0$ y $T_2 > 0$ son tiempos de falla Weibull $\beta_1 > 0$ y $\beta_2 > 0$, los parámetros de forma y $\theta_1 > 0$ y $\theta_2 > 0$, los parámetros de escala asociados a X y Y respectivamente, $0 \leq \lambda \leq 1$, es el parámetro de dependencia entre X y Y . Cuando el parámetro de dependencia λ entre tiempos Weibull es 0, entonces hay independencia entre X y Y . A medida que λ aumenta, la dependencia entre X y Y aumenta.

4.3. Estudio de simulación

Antes de presentar el esquema general de la metodología, se explica la notación utilizada para la función de supervivencia del tiempo mínimo de falla:

- Función de supervivencia verdadera del tiempo mínimo de falla: $S(t)$
- Función de supervivencia estimada del tiempo mínimo de falla, estimando los parámetros asociados a la distribución de los tiempos de cada modo de falla, y asumiendo independencia entre los modos de falla: $\widehat{S}^*(t)$.
- Función de supervivencia estimada del tiempo mínimo de falla, teniendo en cuenta la dependencia entre los modos de falla y asumiendo un modelo cópula: $\widehat{S}(t)_{\text{Cópula}}$.

4.3.1. Esquema general de la metodología aplicada

A continuación se presenta un esquema general que permite la obtención de las comparaciones de la función de supervivencia verdadera del tiempo mínimo de falla $S(t)$, con las funciones de supervivencia estimada $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Cópula}}$.

1. Se consideran 42 casos posibles de funciones de supervivencia verdaderas del tiempo mínimo de falla de un sistema con dos modos de falla ($S(t)$). Los modos de falla están asociados a tiempos de falla Lognormales y Weibull.

Las funciones de sobrevivencia verdaderas del tiempo mínimo de falla de un sistema corresponden a 18 escenarios diferentes en el caso de tiempos de falla lognormales; y 24 escenarios diferentes en el caso de tiempos de falla Weibull.

Los casos de tiempos lognormales a estudiar son:

- Tiempos de falla lognormales con parámetros iguales: $\mu_1 = \mu_2 = 0$ y $\sigma_1 = \sigma_2 = 1$.
- Tiempos de falla lognormales con parámetros μ_i 's iguales y σ_i 's diferentes: $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1$ y $\sigma_2 = 2$.
- Tiempos de falla lognormales con parámetros μ_i 's y σ_i 's diferentes: $\mu_1 = 0$ y $\mu_2 = 1$, $\sigma_1 = 2$ y $\sigma_2 = 1$.

Asociados a cada uno de estos casos hay 6 situaciones diferentes asociadas al parámetro de dependencia entre los tiempos de falla. Para tiempos de falla lognormales X y Y , el parámetro de dependencia es el coeficiente de correlación ρ entre $\log(X)$ y $\log(Y)$, y este toma los siguientes valores: $\rho = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$. Es decir para cada uno de los 4 casos de tiempos lognormales hay 6 situaciones distintas que van desde independencia hasta dependencia extrema.

Los casos de tiempos de falla Weibull a estudiar, se diferencian entre sí, de acuerdo a la forma de la función hazard de los tiempos de falla. Por lo tanto para los 4 casos a estudiar, los valores de los parámetros de escala θ_i serán de 1, ya que la forma de la función hazard de tiempos de falla Weibull, depende del parámetro de forma β_i . Los casos de estudio son:

- Tiempos de falla con parámetros de forma β_i diferentes. Modo de falla con tasa de falla decreciente vs. Modo de falla con tasa de falla creciente: $\beta_1 = 0,5$ y $\beta_2 = 2$.
- Tiempos de falla con parámetros de forma β_i diferentes. Modo de falla con tasa de falla decreciente vs. Modo de falla con tasa de falla constante: $\beta_1 = 0,5$ y $\beta_2 = 1$.
- Tiempos de falla con parámetros de forma β_i iguales. Modo de falla con tasa de falla decreciente vs. Modo de falla con tasa de falla decreciente: $\beta_1 = 0,5$ y $\beta_2 = 0,5$.
- Tiempos de falla con parámetros de forma β_i diferentes. Modo de falla con tasa de falla creciente vs. Modo de falla con tasa de falla constante: $\beta_1 = 2$ y $\beta_2 = 1$.

Asociados a cada uno de los 4 casos de tiempos de falla Weibull expuestos anteriormente, hay también 6 situaciones diferentes de acuerdo al parámetro de dependencia

λ entre tiempos de falla Weibull, y este toma los valores $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8, 0.9$. Así como en tiempos de falla lognormales, para cada uno de los casos de tiempos de falla Weibull, hay situaciones distintas que van desde independencia entre los tiempos de falla hasta dependencia extrema entre éstos.

2. Hay 42 funciones de sobrevivencia verdaderas $S(t)$, las cuales estimaremos asumiendo independencia entre los tiempos de falla con el estimador $\widehat{S}^*(t)$ y sin ignorar la dependencia asumiendo un modelo cópula con el estimador $\widehat{S}(t)_{\text{cópula}}$. Para llevar a cabo estas estimaciones se realiza lo siguiente:

- Se generan muestras de tamaño $n = 50$ de tiempos bivariados, tanto Lognormales como Weibull.
- En cada una de las muestras de tiempos bivariados obtenidas, se elige el tiempo más pequeño y se etiqueta cada uno de estos tiempos con 1 ó 2, para indicar si ese tiempo proviene de los tiempos del modo de falla 1, o si ese tiempo proviene de los tiempos del modo de falla 2.
- Se estiman los parámetros asociados a las distribuciones de los tiempos de falla de cada modo de falla, con estos se estiman las funciones de sobrevivencia del tiempo mínimo de falla para cada uno de los casos de interés, mediante $\widehat{S}^*(t)$.
- Se estiman las distribuciones marginales asociadas a cada modo de falla mediante la aplicación del algoritmo cópula gráfico descrito en la sección 3.2 y con estas funciones de distribución marginales se halla $\widehat{S}(t)_{\text{cópula}}$, utilizando la ecuación 3.13.

3. Se construyen intervalos de confianza empíricos basados en simulación montecarlo para $S(t)$, usando los estimadores $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Cópula}}$.

- Para cada uno de los escenarios descritos se obtienen 1000 estimaciones de $S(t_p)$ usando ambos estimadores $\widehat{S}^*(t_p)$ y $\widehat{S}(t_p)_{\text{Cópula}}$ en cada percentil t_p con $p = 0.05, 0.25, 0.50, 0.75, 0.95$. De esta forma se obtiene la distribución empírica de $\widehat{S}(t_p)$ en cada uno de los percentiles considerados.
- Finalmente, de estas distribuciones se obtienen los límites puntuales de confianza empíricos aproximados del 95%. Estos límites puntuales sirven como región de aceptación para realizar pruebas de hipótesis empíricas en cada percentil. Se tienen entonces, dos hipótesis nulas en cada punto, en las que se prueban la igualdad estadística de las funciones $\widehat{S}^*(t_p)$ vs. $S(t_p)$ y $\widehat{S}(t_p)_{\text{Cópula}}$ vs. $S(t_p)$.

4. Adicionalmente, para comparar puntualmente los dos estimadores se calcula la eficiencia relativa (ER_p), de $\widehat{S}^*(t_p)$ con relación a $\widehat{S}(t_p)_{\text{Copula}}$ en los percentiles $p = 0.05, 0.25, 0.50, 0.75, 0.95$ así:

$$ER_p = \frac{\text{ECM}(\widehat{S}(t_p)_{\text{Copula}})}{\text{ECM}(\widehat{S}^*(t_p))} = \frac{\text{E} \left[\left(\widehat{S}(t_p)_{\text{Copula}} - S(t_p) \right)^2 \right]}{\text{E} \left[\left(\widehat{S}^*(t_p) - S(t_p) \right)^2 \right]}$$

donde $\text{ECM}(\widehat{S}(t_p)_{\text{Copula}})$ y $\text{ECM}(\widehat{S}^*(t_p))$ son los errores cuadrático medio que estiman la dispersión de las diferencias entre la función de supervivencia verdadera y los estimadores $\widehat{S}(t_p)_{\text{Copula}}$ y $\widehat{S}^*(t_p)$ respectivamente.

Si $ER_p < 1$ entonces $\text{ECM}(\widehat{S}^*(t_p)) > \text{ECM}(\widehat{S}(t_p)_{\text{Copula}})$ lo cual indicará que $\widehat{S}^*(t_p)$ es menos eficiente frente a $\widehat{S}(t_p)_{\text{Copula}}$ para la estimación de la función de supervivencia verdadera en t_p .

Capítulo 5

Análisis de resultados

En este capítulo se presentan los resultados del estudio, en el que se realizan comparaciones entre las funciones de sobrevivencia verdaderas del tiempo mínimo de falla $S(t)$, con la función de sobrevivencia estimada donde los parámetros asociados a los tiempos de falla son estimados asumiendo independencia entre los tiempos de falla, $\hat{S}^*(t)$ y la función de confiabilidad estimada sin ignorar la dependencia entre los tiempos de falla asumiendo un modelo cópula, $\hat{S}(t)_{\text{Cópula}}$. Estas comparaciones se realizan para tiempos de falla Weibull y Lognormal.

5.1. Intervalos de confianza empíricos para $S(t)$

En aplicaciones de sobrevivencia generalmente se tienen pocas fallas y muchos datos censurados. Bajos estas características, las pruebas formales de bondad de ajuste no son útiles (Meeker & Escobar, 2009). En lugar de ello se realizaron pruebas de hipótesis empíricas en cada percentil, utilizando como región de aceptación los intervalos de confianza empíricos descritos en el capítulo anterior.

5.1.1. Casos Weibull

En las figuras 5.1 a 5.6 se muestran los límites de confianza para los casos Weibull en que los parámetros de forma son $\beta_1 = 0.5$ y $\beta_2 = 2.0$, y parámetros de dependencia $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8$ y 0.9 respectivamente. En estos gráficos, para cada uno de los percentiles t_p se tiene el valor verdadero de $S(t_p)$, el límite inferior de $S^*(t_p)$ (denotado como $\text{LI}.S^*(t_p)$), el límite superior de $S^*(t_p)$ (denotado como $\text{LS}.S^*(t_p)$) y los límites inferior y superior de

$S(t_p)_{\text{Copula}}$ (denotados $LI.S(t)_{\text{Copula}}$ y $LS.S(t)_{\text{Copula}}$ respectivamente). Se observa en general que los intervalos para $S(t_p)$ mediante la distribución empírica de $\widehat{S}^*(t)$ son más amplios que los intervalos de confianza para $S(t_p)$ mediante la distribución empírica de $\widehat{S}(t_p)_{\text{Copula}}$, lo cual indica que el estimador $\widehat{S}(t_p)_{\text{Copula}}$ tiene mayor precisión a la hora de estimar $S(t_p)$. La amplitud de estos intervalos se hace más evidente a medida que aumenta la dependencia λ . Al ser tan amplios los límites de confianza, la estimación de $S(t)$ con $\widehat{S}^*(t)$ puede tomar diferentes comportamientos, sobreestimando o subestimando la sobrevivencia verdadera.

Así por ejemplo, en la figura 5.4, para $t_{0.25} = 0.0830$, $S(0.0830) = 0.75$, $LI.S(t)_{\text{Copula}} = 0.6880$ y $LS.S(t)_{\text{Copula}} = 0.8026$; $LI.S^*(t) = 0.6389$ y $LS.S^*(t) = 0,8439$. Observe que el valor verdadero de $S(t_p)$ cae dentro de los intervalos de confianza empíricos, pero los límites obtenidos con $S(t_p)_{\text{Copula}}$ son más pequeños que los que se obtienen con $S^*(t_p)$.

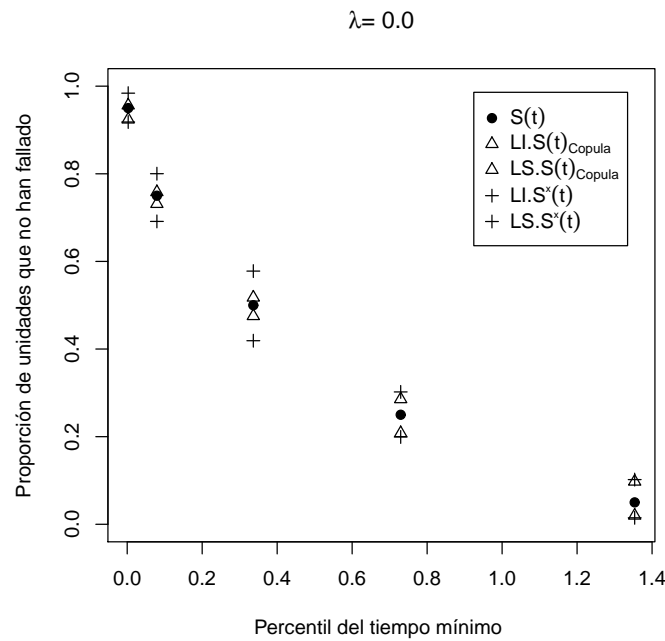


Figura 5.1: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

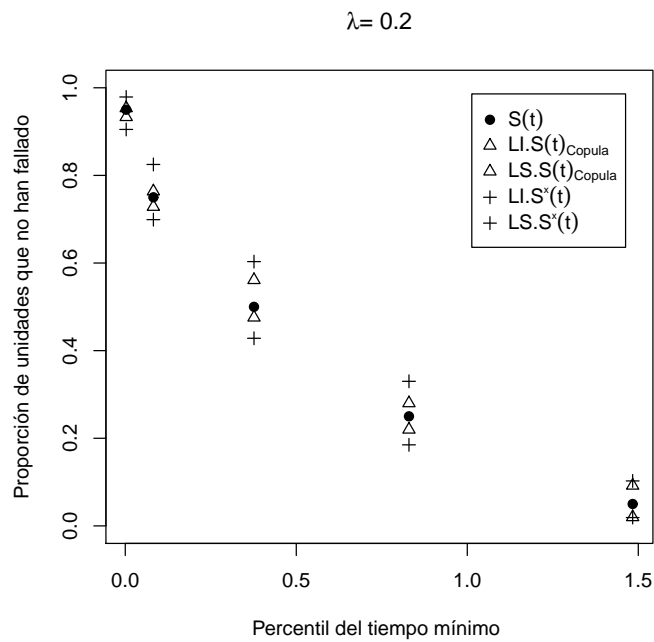


Figura 5.2: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

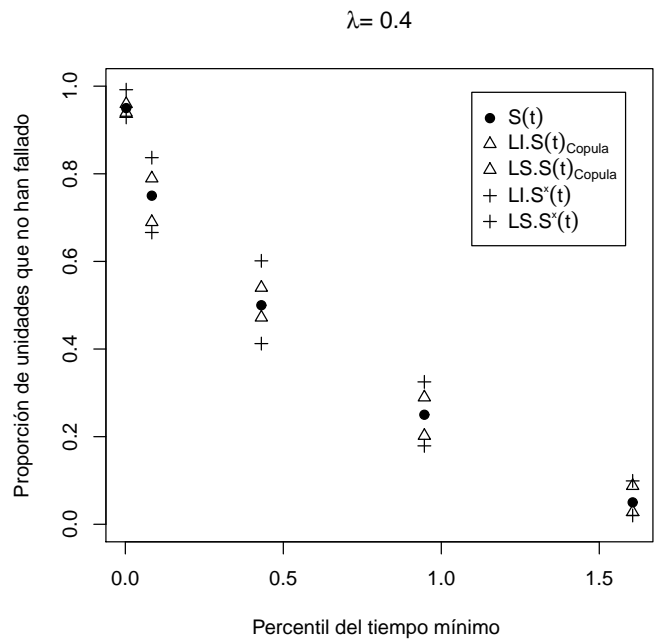


Figura 5.3: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

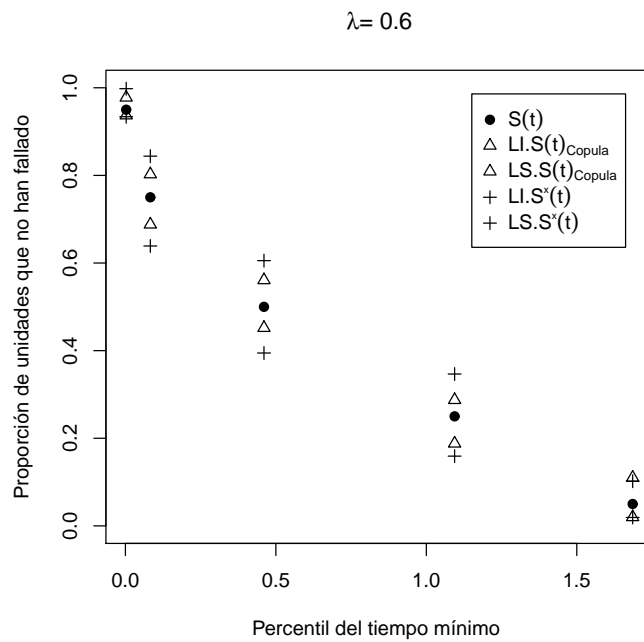


Figura 5.4: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

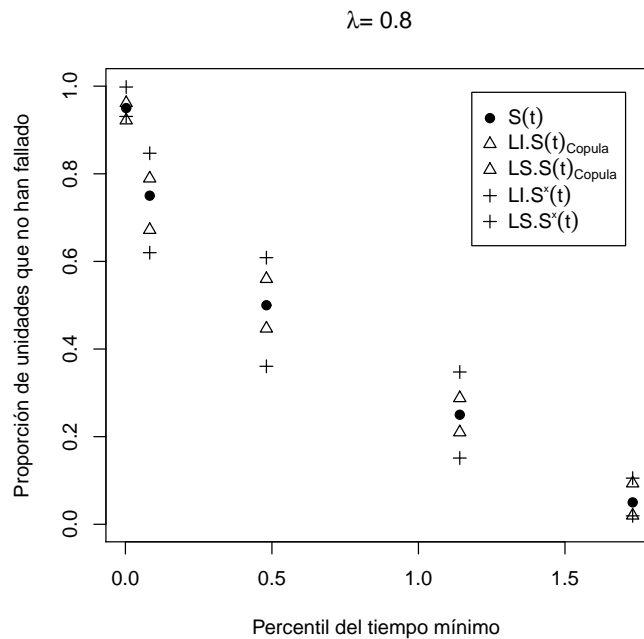


Figura 5.5: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

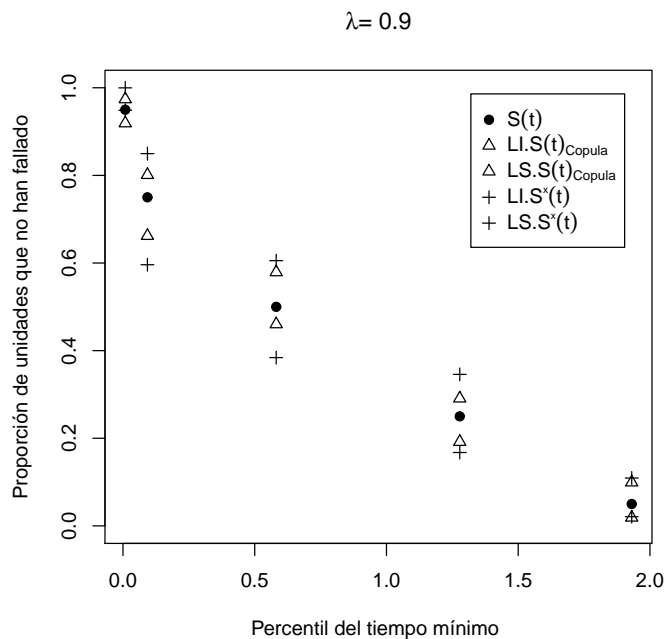


Figura 5.6: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.9$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

En las figuras 5.7 a 5.12 se muestran los límites de confianza para los casos Weibull en que los parámetros de forma son $\beta_1 = 0.5$ y $\beta_2 = 1.0$, y parámetros de dependencia $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8$ y 0.9 respectivamente.

El comportamiento para este caso es muy parecido al anterior, el valor verdadero de $S(t_p)$ cae dentro de los intervalos de confianza empíricos, pero los intervalos obtenidos con $\hat{S}(t_p)_{Copula}$ son más estrechos que los obtenidos con $\hat{S}^*(t_p)$. En estos casos al ser tan amplios los límites de confianza, la estimación de $S(t_p)$ con $\hat{S}^*(t_p)$ puede tomar diferentes comportamientos, sobreestimando o subestimando la sobrevivencia verdadera.

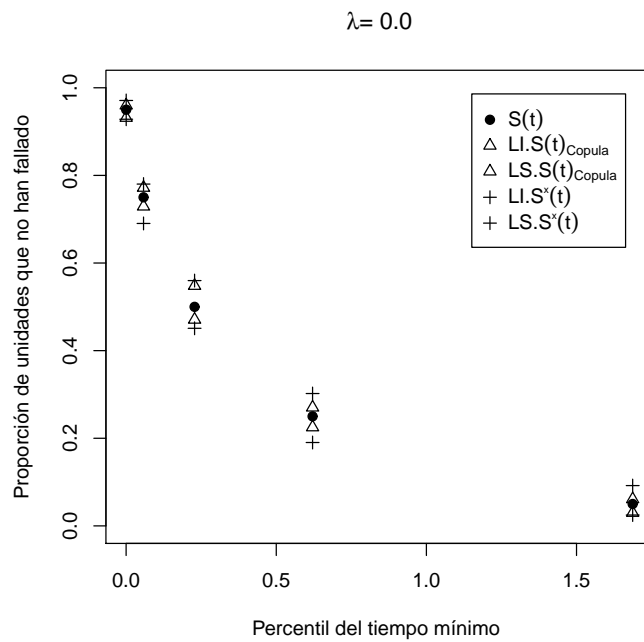


Figura 5.7: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

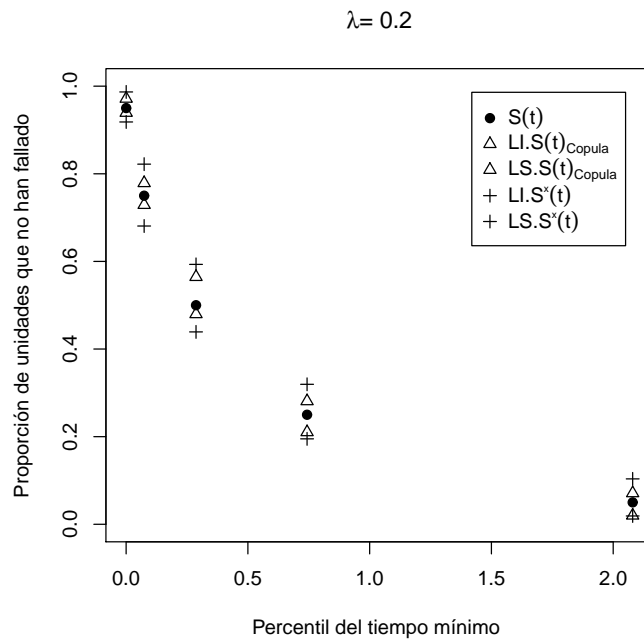


Figura 5.8: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

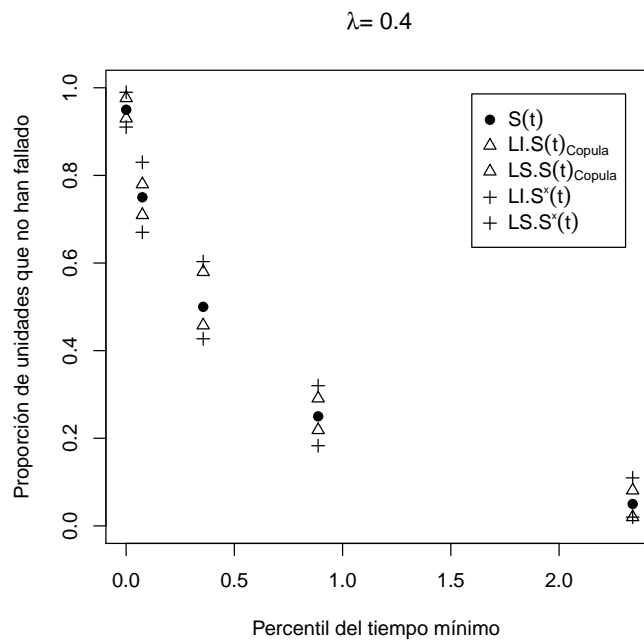


Figura 5.9: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

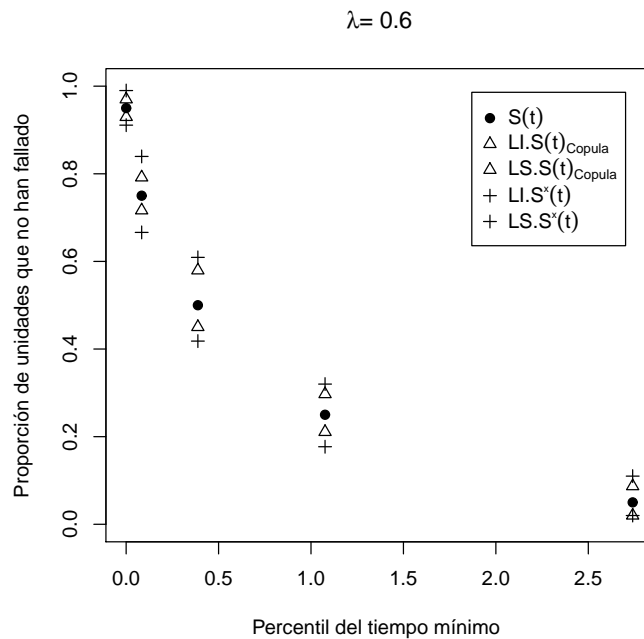


Figura 5.10: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

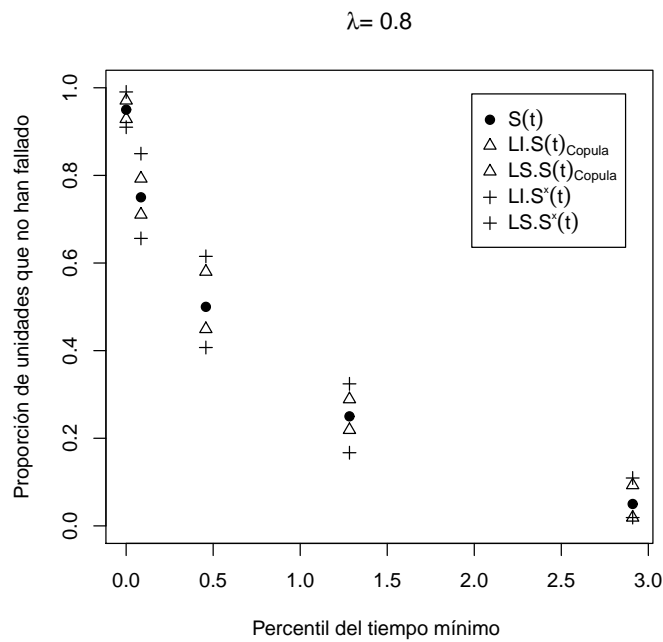


Figura 5.11: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

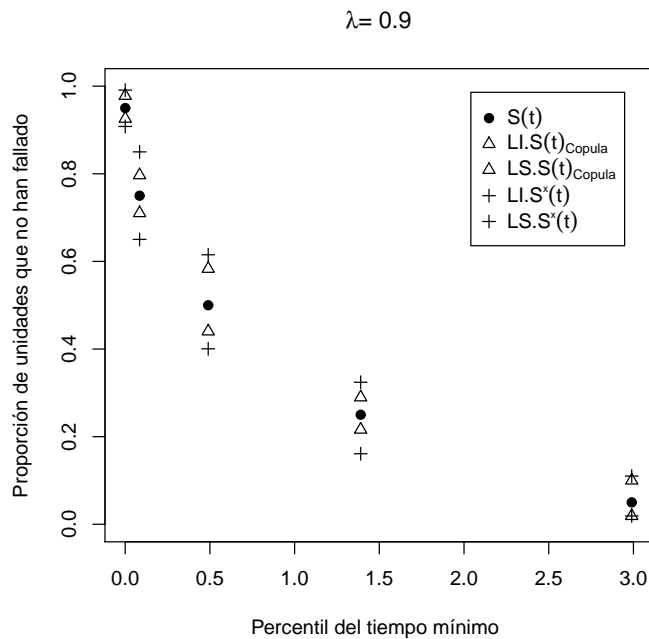


Figura 5.12: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\lambda = 0.9$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

En las figuras 5.13 a 5.18 se muestran los límites de confianza para los casos Weibull en que los parámetros de forma son $\beta_1 = 2.0$ y $\beta_2 = 1.0$, y parámetros de dependencia $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8$ y 0.9 respectivamente. El comportamiento para este caso es muy parecido a los casos Weibull con parámetros de forma $\beta_1 = 0.5 ; \beta_2 = 2.0$ y $\beta_1 = 0.5 ; \beta_2 = 1.0$.

El valor verdadero de $S(t_p)$ cae dentro de los intervalos de confianza empíricos, pero los intervalos obtenidos con $\hat{S}(t_p)_{\text{Copula}}$ son más estrechos que los obtenidos con $\hat{S}^*(t_p)$. En estos casos al ser tan amplios los límites de confianza, la estimación de $S(t_p)$ con $\hat{S}^*(t_p)$ puede tomar diferentes comportamientos, sobreestimando o subestimando la sobrevivencia verdadera.

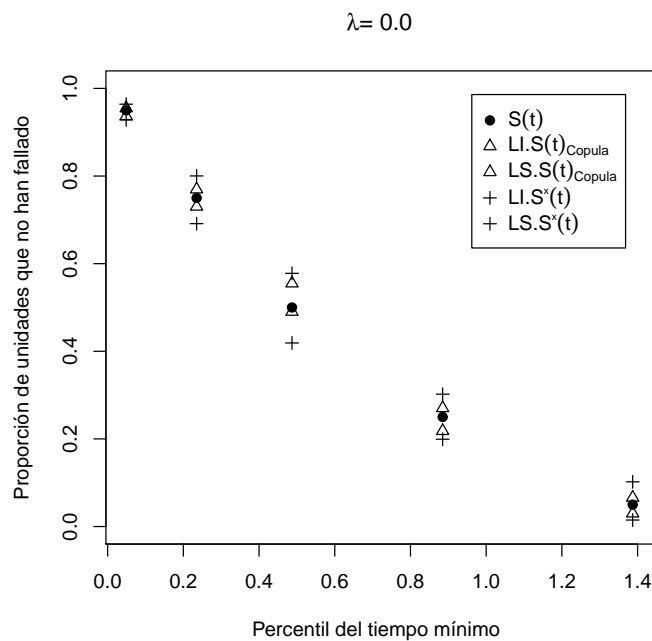


Figura 5.13: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0 ; \beta_2 = 1.0$ y $\lambda = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

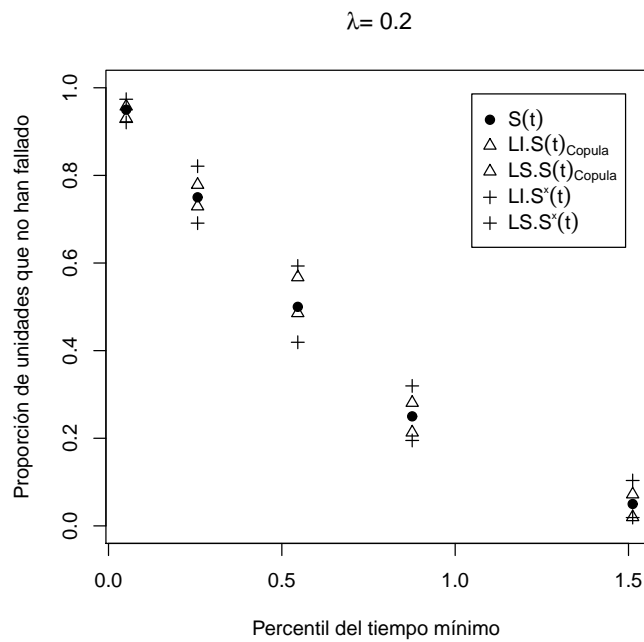


Figura 5.14: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

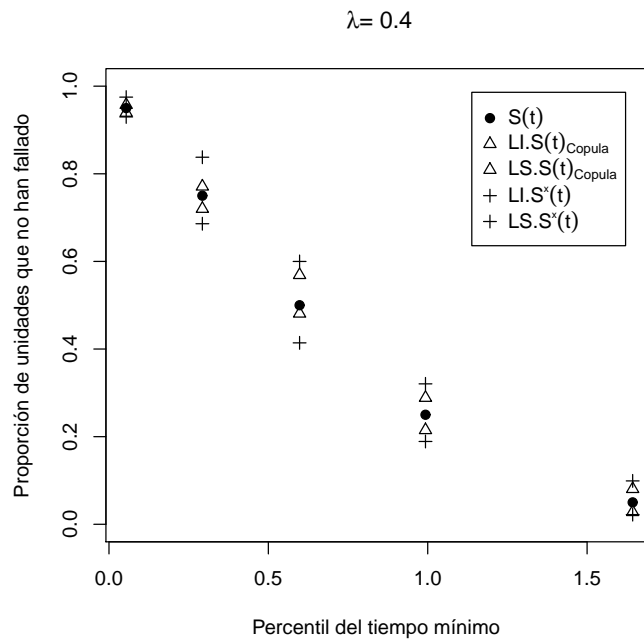


Figura 5.15: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

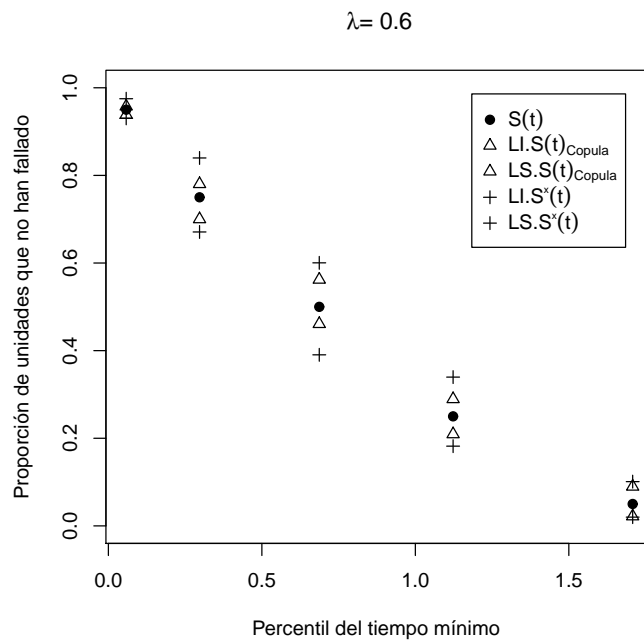


Figura 5.16: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

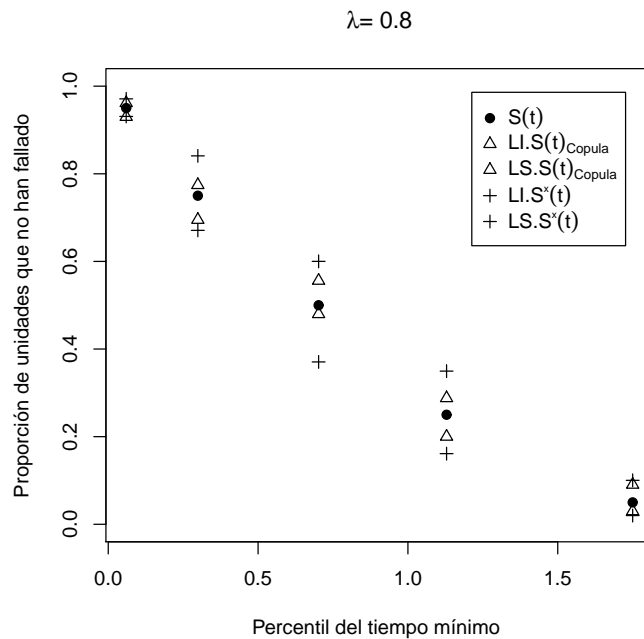


Figura 5.17: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

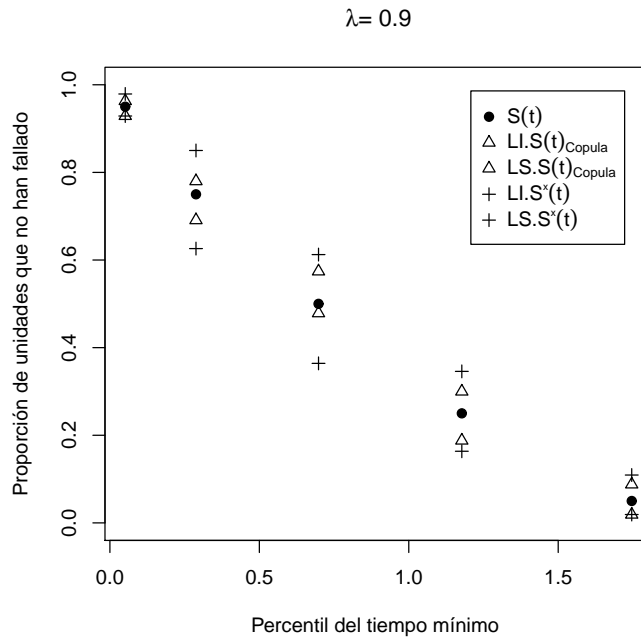


Figura 5.18: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 2.0$; $\beta_2 = 1.0$ y $\lambda = 0.9$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

En las figuras 5.19 a 5.24 se muestran los límites de confianza para los casos Weibull en que los parámetros de forma son $\beta_1 = 0.5$ y $\beta_2 = 0.5$, y parámetros de dependencia $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8$ y 0.9 respectivamente. El comportamiento para este caso es muy parecido a los casos Weibull con parámetros de forma $\beta_1 = 0.5$; $\beta_2 = 2.0$ con $\beta_1 = 0.5$; $\beta_2 = 1.0$ y $\beta_1 = 2.0$; $\beta_2 = 1.0$

El valor verdadero de $S(t_p)$ cae dentro de los intervalos de confianza empíricos, pero los intervalos obtenidos con $\hat{S}(t_p)_{\text{Copula}}$ son más estrechos que los obtenidos con $\hat{S}^*(t_p)$. En estos casos al ser tan amplios los límites de confianza, la estimación de $S(t_p)$ con $\hat{S}^*(t_p)$ puede tomar diferentes comportamientos, sobreestimando o subestimando la sobrevivencia verdadera.

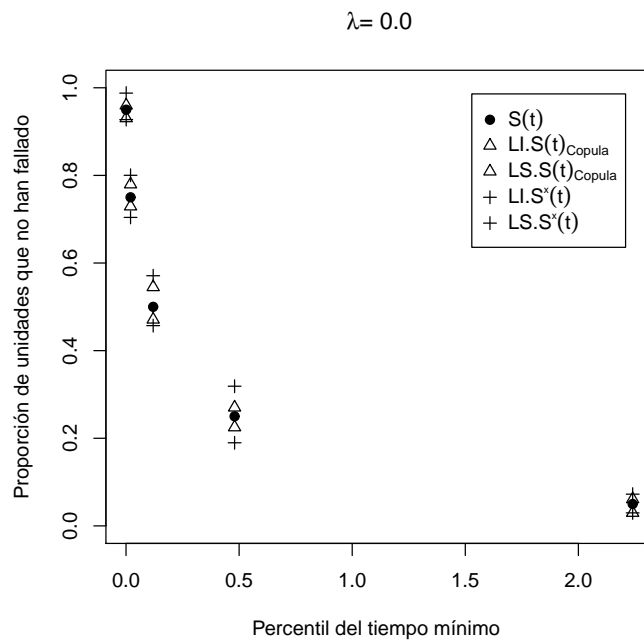


Figura 5.19: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

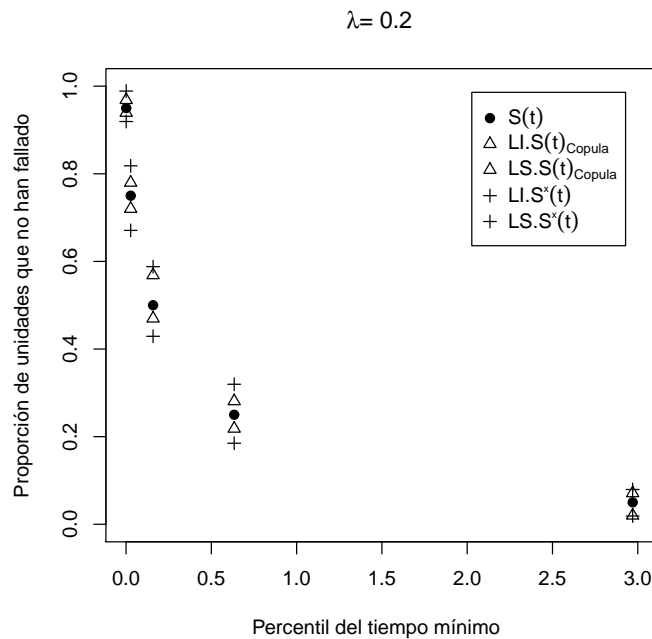


Figura 5.20: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

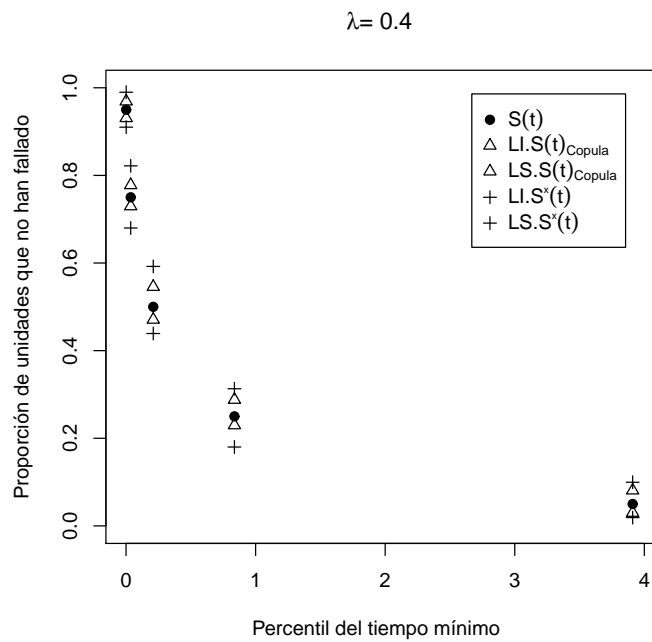


Figura 5.21: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

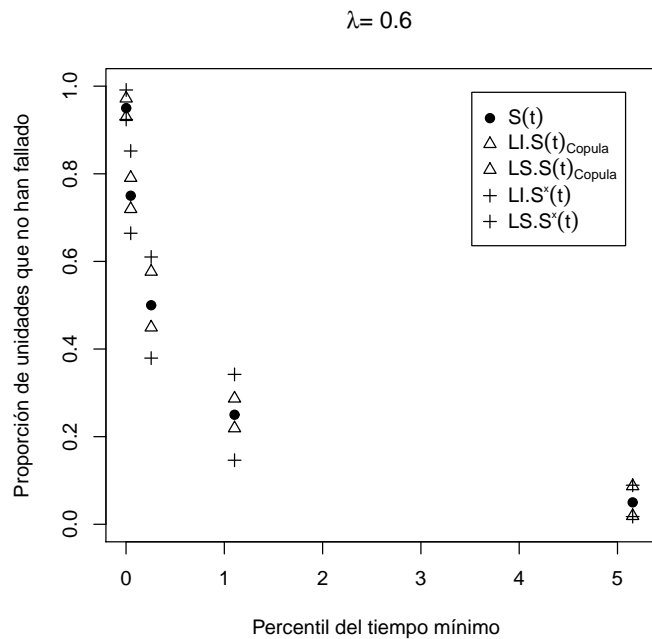


Figura 5.22: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

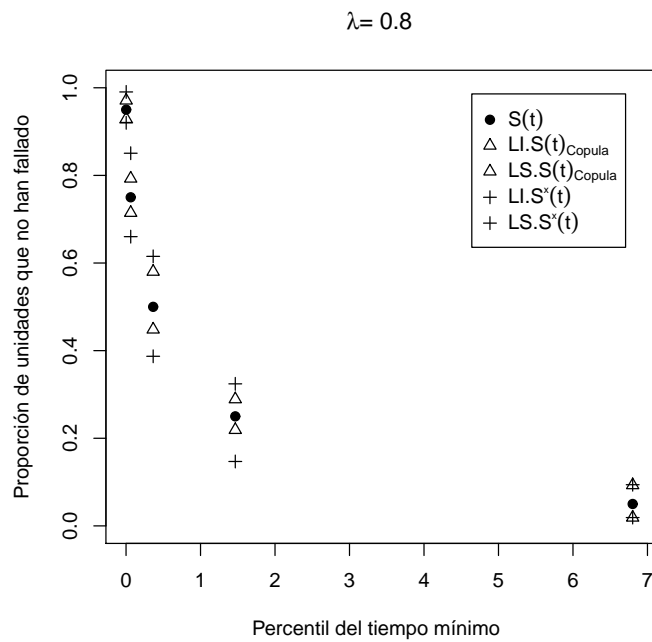


Figura 5.23: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

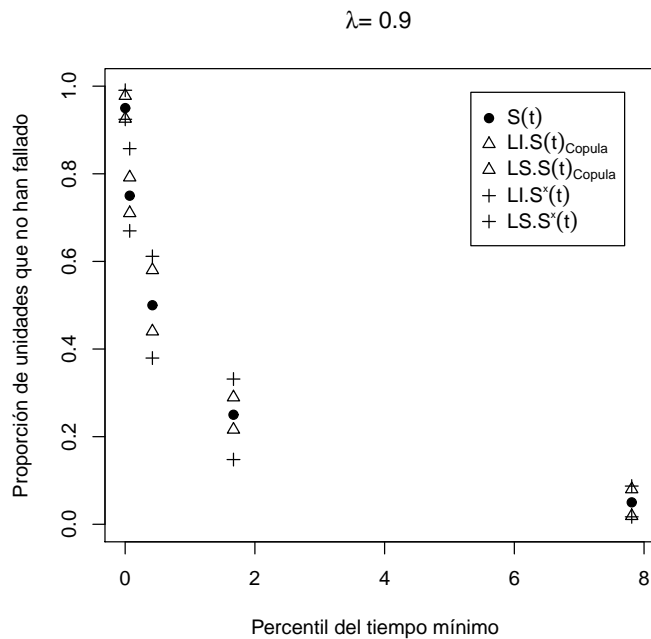


Figura 5.24: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 0.5$ y $\lambda = 0.9$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

En todos los casos Weibull mostrados se observa como los intervalos de confianza empíricos para $S(t_p)$ basados en $\widehat{S}^*(t_p)$ son más amplios que aquellos basados en $\widehat{S}(t_p)_{\text{Copula}}$, indicando que el estimador $\widehat{S}(t_p)_{\text{Copula}}$ tiene mayor precisión a la hora de estimar $S(t)$, esta precisión se hace más evidente a medida que aumenta la dependencia λ .

En la tablas 5.1 a 5.4 se muestra la eficiencia relativa (ER_p) del estimador $\widehat{S}^*(t_p)$ con relación a $\widehat{S}(t_p)_{\text{Copula}}$ en los percentiles $p = 0.25, 0.50, 0.75$. En la subsección 4.3.1 se muestra la forma como se estima ER_p . En estas tablas se observa como a medida que aumenta la dependencia λ disminuye la eficiencia relativa, es decir la eficiencia del estimador $\widehat{S}^*(t_p)$ con relación al estimador $\widehat{S}(t_p)_{\text{Copula}}$ es menor al aumentar la dependencia.

λ	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.2991	0.1357	0.1581
0.2	0.2763	0.1286	0.1374
0.4	0.2747	0.1093	0.1321
0.6	0.2367	0.0964	0.1102
0.8	0.1951	0.0682	0.1013
0.9	0.1940	0.0671	0.1005

Cuadro 5.1: Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0,5$; $\beta_2 = 2$ y $\theta_1 = \theta_2 = 1$

λ	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.3487	0.3104	0.3248
0.2	0.3397	0.3010	0.3132
0.4	0.2928	0.2271	0.2804
0.6	0.2670	0.1943	0.2184
0.8	0.2269	0.1821	0.1908
0.9	0.1512	0.1110	0.1353

Cuadro 5.2: Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 2$; $\beta_2 = 1$ y $\theta_1 = \theta_2 = 1$

λ	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.3341	0.3189	0.3760
0.2	0.2978	0.2890	0.3329
0.4	0.2370	0.2176	0.2536
0.6	0.2153	0.2012	0.2250
0.8	0.1815	0.1730	0.1989
0.9	0.1427	0.1210	0.1532

Cuadro 5.3: Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0,5$; $\beta_2 = 1$ y $\theta_1 = \theta_2 = 1$

λ	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.4358	0.4296	0.4434
0.2	0.4275	0.4022	0.4106
0.4	0.3950	0.3782	0.4050
0.6	0.3501	0.3316	0.3793
0.8	0.3366	0.2854	0.3423
0.9	0.2988	0.2606	0.2921

Cuadro 5.4: Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0,5$; $\beta_2 = 0,5$ y $\theta_1 = \theta_2 = 1$

De todo lo anterior, se puede concluir que el estimador $\widehat{S}(t_p)_{\text{Copula}}$ es mejor que $\widehat{S}^*(t_p)$ para estimar $S(t_p)$ en todos los casos Weibull estudiados, ya los intervalos obtenidos con $\widehat{S}(t_p)_{\text{Copula}}$ son más estrechos que los que se obtienen con $\widehat{S}^*(t_p)$, y además el ER_p es siempre menor que uno, lo que nos asegura que la dispersión del estimador $\widehat{S}(t_p)_{\text{Copula}}$ frente a $S(t_p)$ es siempre menor que la dispersión del estimador $\widehat{S}^*(t_p)$ frente a $S(t_p)$.

5.1.2. Casos Lognormal

En las figuras 5.25 a 5.30 se muestran los límites de confianza para los casos Lognormal con $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y parámetros de dependencia $\rho = 0.0, 0.2, 0.4, 0.6, 0.8$ y 0.9 respectivamente. En estos gráficos, para cada uno de los percentiles t_p se tiene el valor verdadero de $S(t_p)$, el límite inferior de $S^*(t_p)$ (denotado como $LI.S^*(t_p)$), el límite superior de $S^*(t_p)$ (denotado como $LS.S^*(t_p)$) y los límites inferior y superior de $S(t)_{\text{Copula}}$ (denotados $LI.S(t)_{\text{Copula}}$ y $LS.S(t)_{\text{Copula}}$ respectivamente).

Al igual que para los casos Weibull, se observa como los intervalos de confianza para $S(t_p)$ estimados mediante la distribución empírica de $\widehat{S}^*(t_p)$ son más amplios que los estimados mediante la distribución empírica de $\widehat{S}(t_p)_{\text{Copula}}$, indicando esto que el estimador $\widehat{S}(t_p)_{\text{Copula}}$ tiene mayor precisión a la hora de estimar $S(t_p)$. La amplitud de estos intervalos se hace más evidente a medida que aumenta la dependencia λ . Al ser tan amplios los límites de confianza, la estimación de $S(t)$ con $\widehat{S}^*(t)$ puede tomar diferentes comportamientos, sobreestimando o subestimando la sobrevivencia verdadera.

Por ejemplo, en la figura 5.27, para $t_{0.5} = 0.5330$, $S(0.5330) = 0.5$, $\text{LI}.S(t)_{\text{Copula}} = 0.4647$ y $\text{LS}.S(t)_{\text{Copula}} = 0.5396$; $\text{LI}.S^*(t) = 0.4101$ y $\text{LS}.S^*(t) = 0.5919$. Observe que el valor verdadero de $S(t_p)$ cae dentro de las bandas de confianza, pero los límites obtenidos con $S(t_p)_{\text{Copula}}$ son más pequeños que los que se obtienen con $S^*(t_p)$.

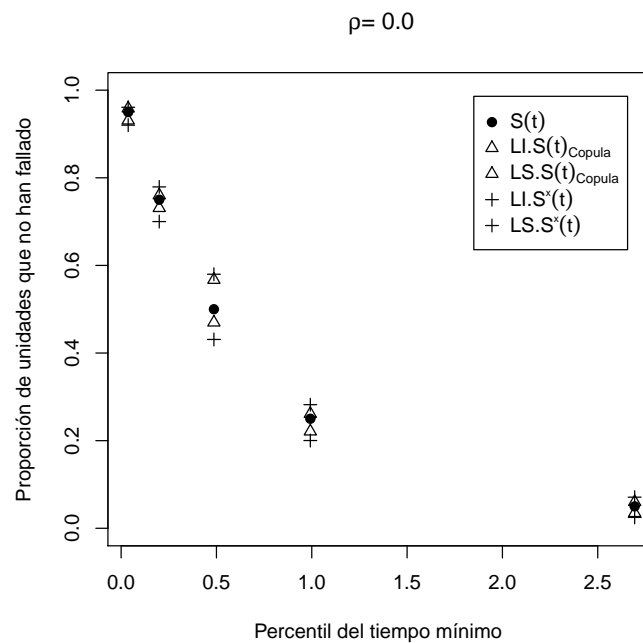


Figura 5.25: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

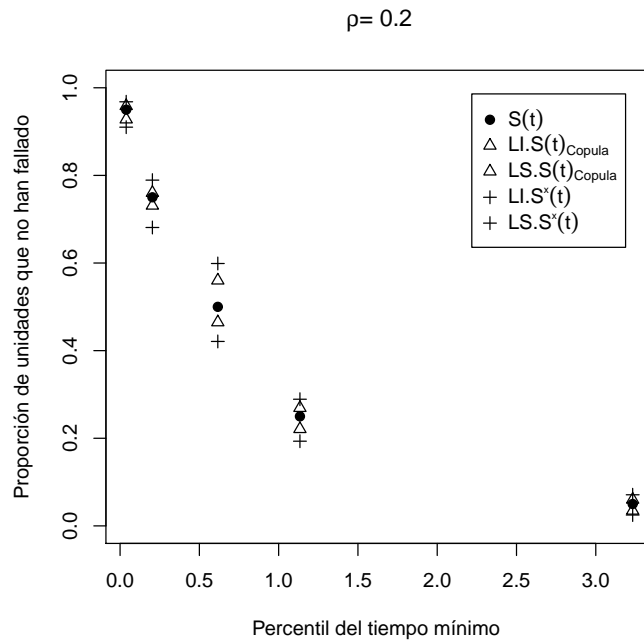


Figura 5.26: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

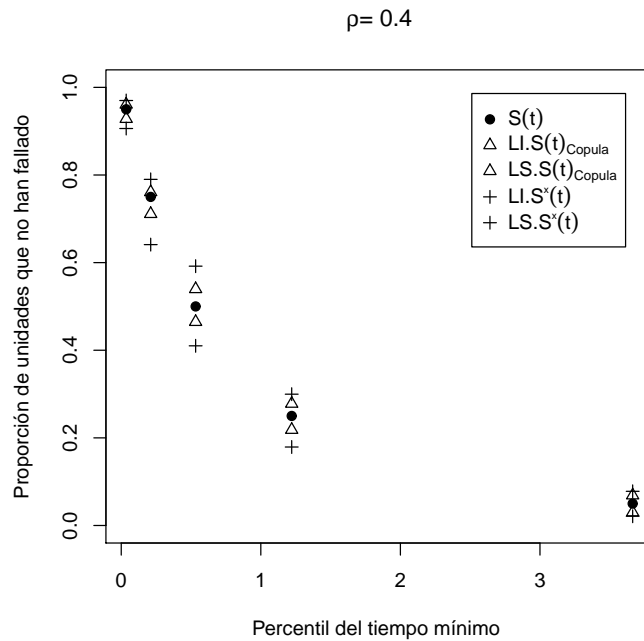


Figura 5.27: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

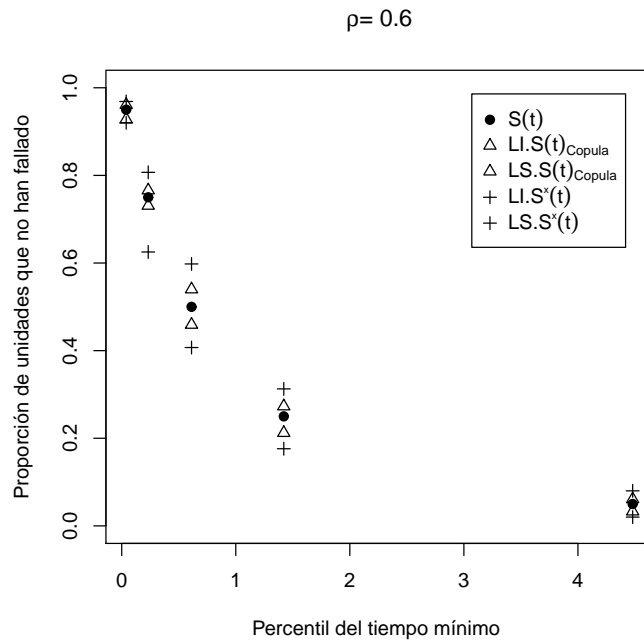


Figura 5.28: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

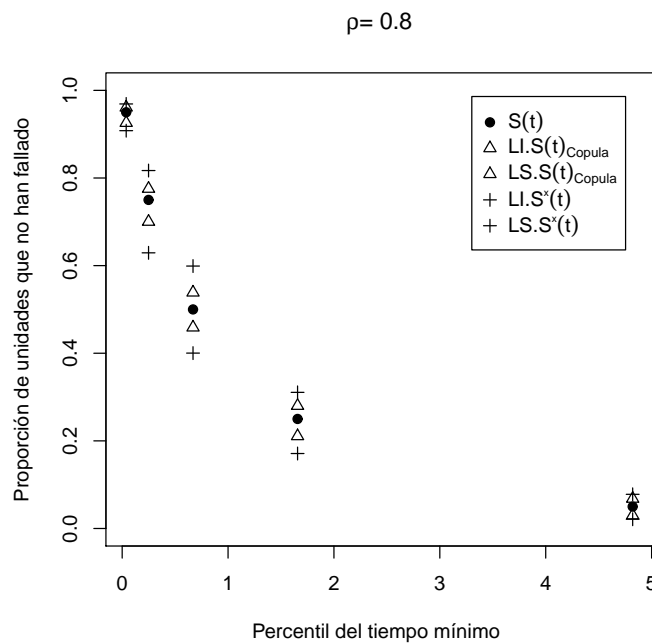


Figura 5.29: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

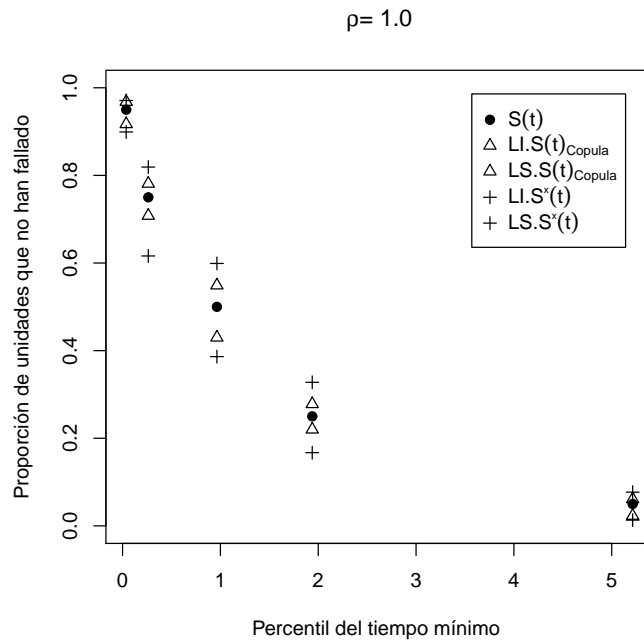


Figura 5.30: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$ y $\rho = 1.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

En las figuras 5.31 a 5.36 se muestran los límites de confianza para los casos Lognormal con $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y parámetros de dependencia $\rho = 0.0, 0.2, 0.4, 0.6, 0.8$ y 0.9 respectivamente. En estos gráficos, para cada uno de los percentiles t_p se tiene el valor verdadero de $S(t_p)$, el límite inferior de $S^*(t_p)$ (denotado como $LI.S^*(t_p)$), el límite superior de $S^*(t_p)$ (denotado como $LS.S^*(t_p)$) y los límites inferior y superior de $S(t)_{Copula}$ (denotados $LI.S(t)_{Copula}$ y $LS.S(t)_{Copula}$ respectivamente).

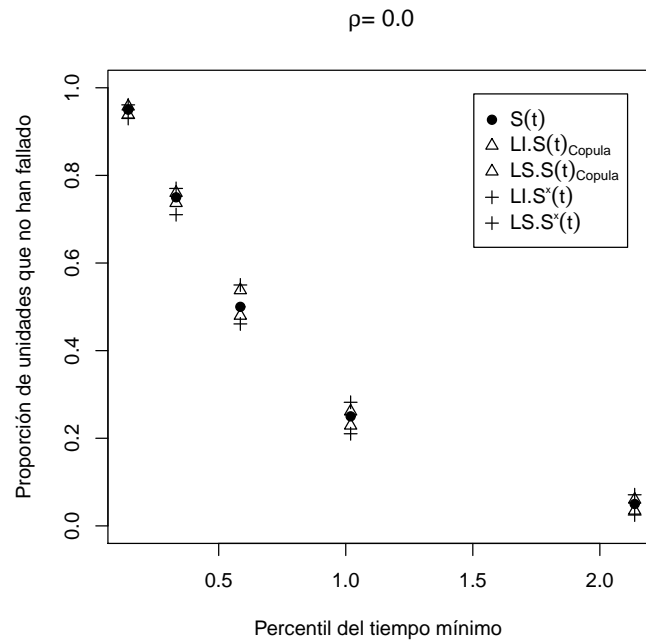


Figura 5.31: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

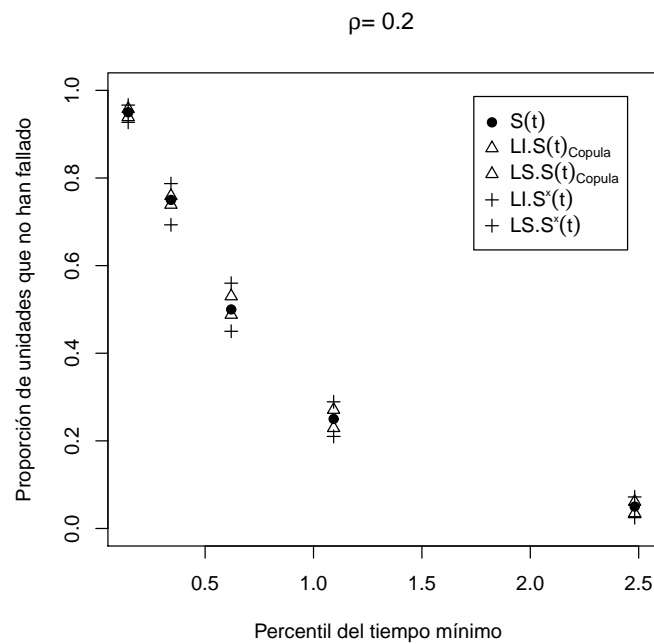


Figura 5.32: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

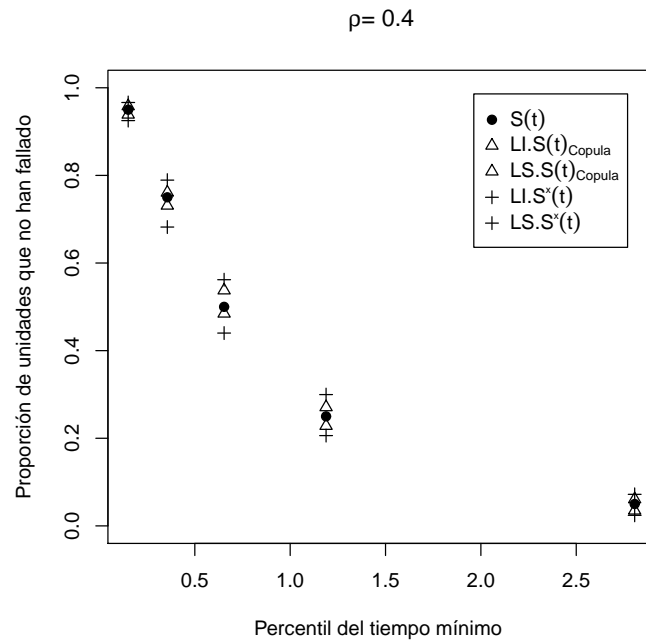


Figura 5.33: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

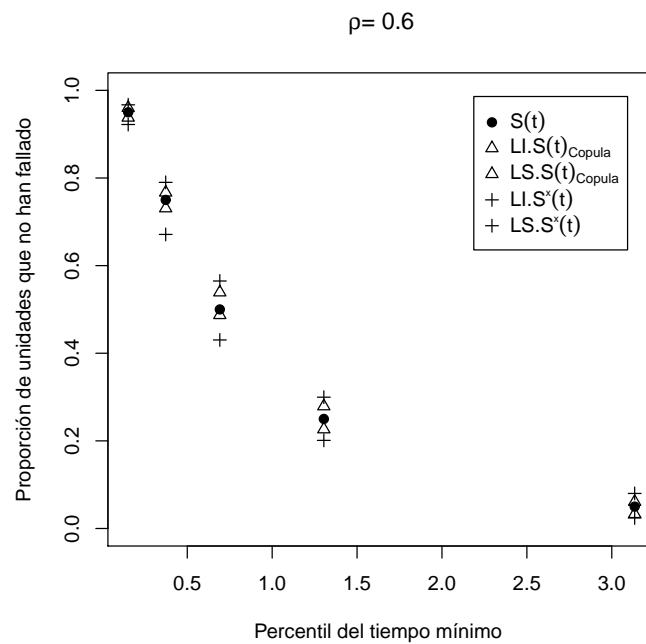


Figura 5.34: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

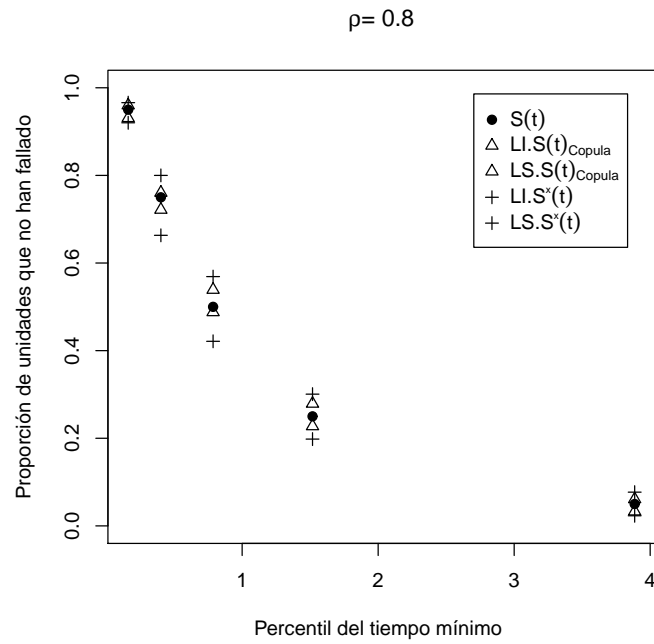


Figura 5.35: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

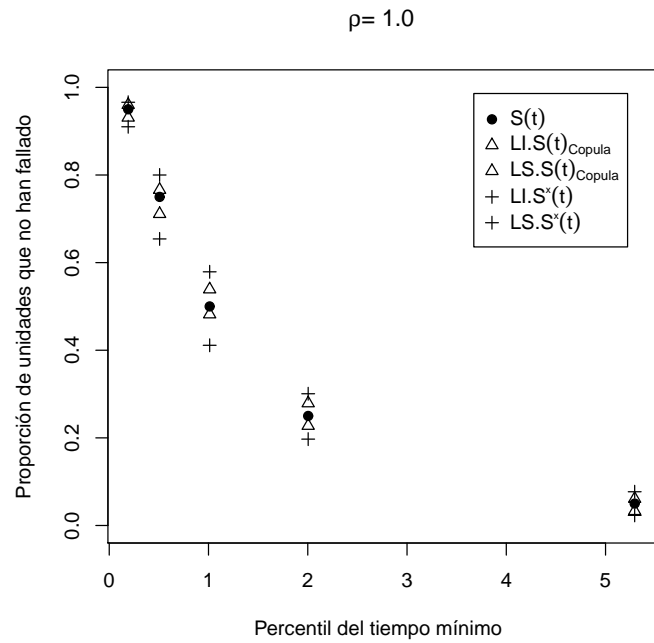


Figura 5.36: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$ y $\rho = 1.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

En las figuras 5.37 a 5.42 se muestran los límites de confianza para los casos Lognormal con $\mu_1 = 0$, $\mu_2 = 1$; $\sigma_1 = 2$, $\sigma_2 = 1$ y parámetros de dependencia $\rho = 0.0, 0.2, 0.4, 0.6, 0.8$ y 0.9 respectivamente. En estos gráficos, para cada uno de los percentiles t_p se tiene el valor verdadero de $S(t_p)$, el límite inferior de $S^*(t_p)$ (denotado como $LI.S^*(t_p)$), el límite superior de $S^*(t_p)$ (denotado como $LS.S^*(t_p)$) y los límites inferior y superior de $S(t)_{\text{Copula}}$ (denotados $LI.S(t)_{\text{Copula}}$ y $LS.S(t)_{\text{Copula}}$ respectivamente).

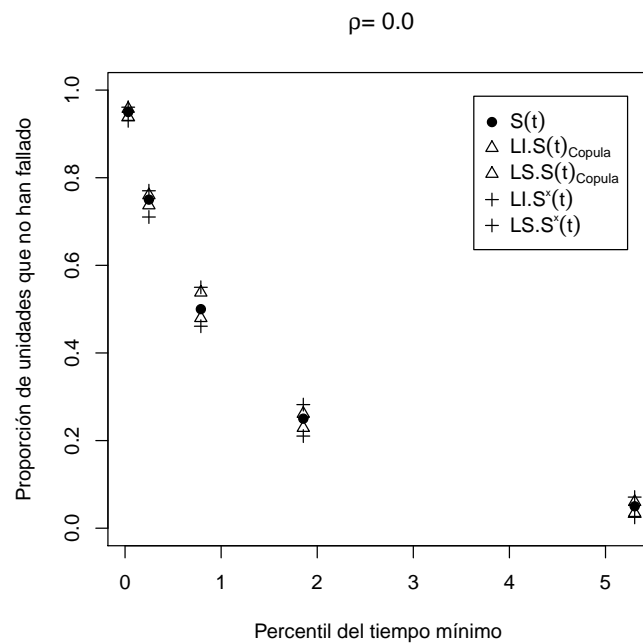


Figura 5.37: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = 0$, $\mu_2 = 1$; $\sigma_1 = 2$, $\sigma_2 = 1$ y $\rho = 0.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

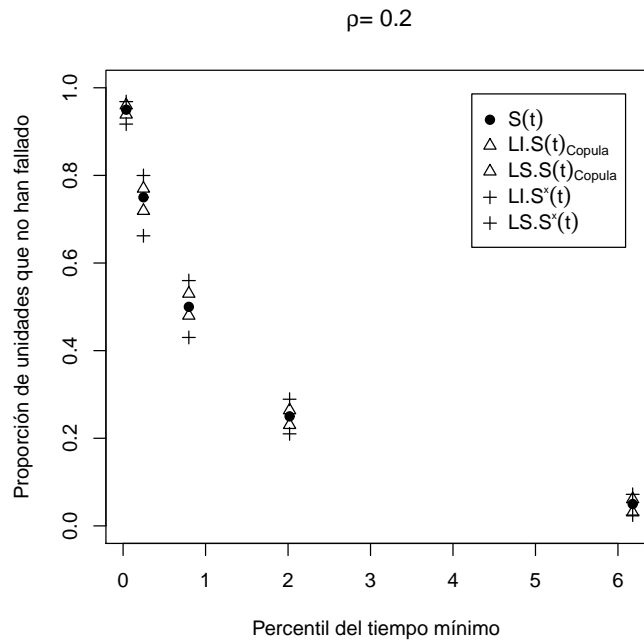


Figura 5.38: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 0.2$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

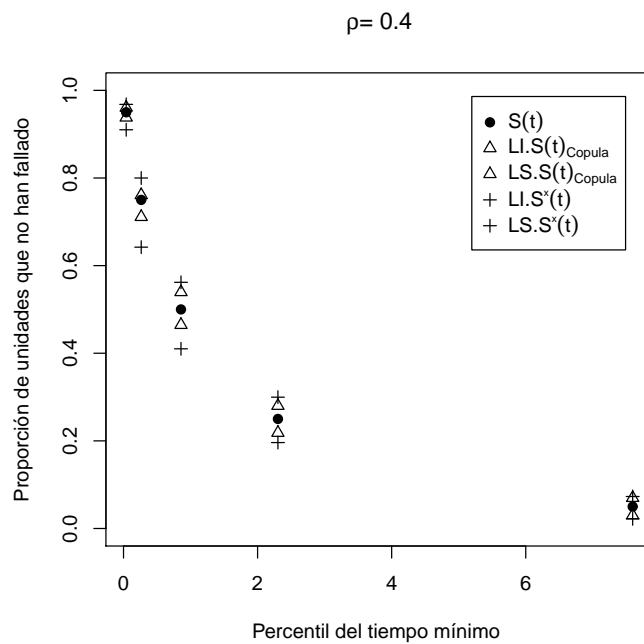


Figura 5.39: Intervalos de confianza empíricos aproximados del 95% para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 0.4$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

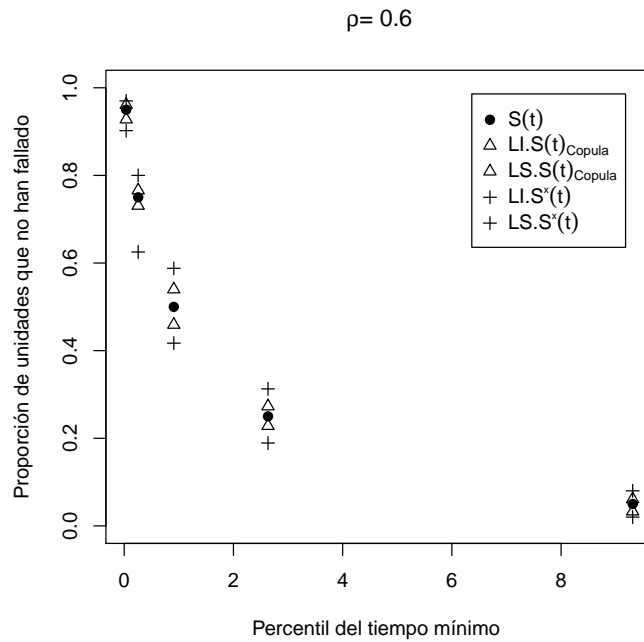


Figura 5.40: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

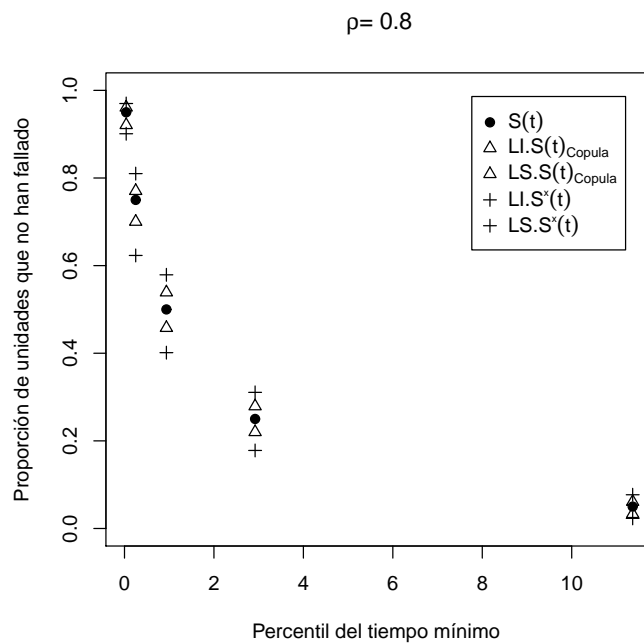


Figura 5.41: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 0.8$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

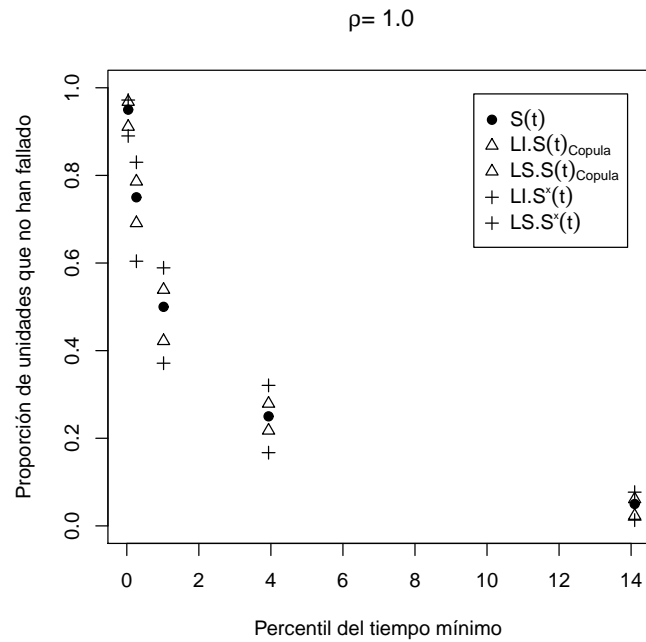


Figura 5.42: Intervalos de confianza empíricos aproximados del 95 % para $S(t_p)$ en el caso Lognormal $\mu_1 = 0, \mu_2 = 1$; $\sigma_1 = 2, \sigma_2 = 1$ y $\rho = 1.0$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

En la tablas 5.5 a 5.7 se muestra la eficiencia relativa (ER_p) del estimador $\widehat{S}^*(t_p)$ con relación a $\widehat{S}(t_p)_{Copula}$ en los percentiles $p = 0.25, 0.50, 0.75$ para los casos Lognormal. Al igual que en los casos Weibull en estas tablas se observa como a medida que aumenta la dependencia ρ disminuye la eficiencia relativa, es decir la eficiencia del estimador $\widehat{S}^*(t_p)$ con relación al estimador $\widehat{S}(t_p)_{Copula}$ es menor al aumentar la dependencia. De todo esto se puede concluir que el estimador $\widehat{S}^*(t_p)$ es menos eficiente que $\widehat{S}(t_p)_{Copula}$ para estimar $S(t_p)$.

ρ	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.5334	0.5546	0.3839
0.2	0.4565	0.4143	0.3169
0.4	0.4009	0.3234	0.2730
0.6	0.3261	0.3148	0.2636
0.8	0.3051	0.2859	0.2582
1.0	0.2984	0.2810	0.2304

Cuadro 5.5: Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{Copula}$ para tiempos de falla Log-normal con parámetros $\mu_1 = \mu_2 = 0$; $\sigma_1 = \sigma_2 = 1$

ρ	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.3821	0.4203	0.3666
0.2	0.30315	0.3353	0.3247
0.4	0.2841	0.2599	0.2601
0.6	0.2341	0.2080	0.2193
0.8	0.2050	0.1816	0.1821
1.0	0.1617	0.1294	0.1423

Cuadro 5.6: Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Log-normal con parámetros $\mu_1 = \mu_2 = 0$; $\sigma_1 = 1$, $\sigma_2 = 2$

ρ	$p = 0.25$	$p = 0.50$	$p = 0.75$
0.0	0.3856	0.3951	0.3224
0.2	0.2911	0.3156	0.3091
0.4	0.2644	0.2659	0.2503
0.6	0.2116	0.2003	0.1893
0.8	0.2051	0.1916	0.1581
1.0	0.1417	0.1193	0.1023

Cuadro 5.7: Eficiencia relativa de $\widehat{S}^*(t_p)$ respecto a $\widehat{S}(t_p)_{\text{Copula}}$ para tiempos de falla Log-normal con parámetros $\mu_1 = 0$, $\mu_2 = 1$; $\sigma_1 = 2$, $\sigma_2 = 1$

Al igual que en los casos Weibull, en los casos Lognormal estudiados se puede concluir que el estimador $\widehat{S}(t_p)_{\text{Copula}}$ es mejor que $\widehat{S}^*(t_p)$ para estimar $S(t_p)$ ya que los intervalos de confianza puntuales empíricos con $\widehat{S}(t_p)_{\text{Copula}}$ son más estrechos que los obtenidos con $\widehat{S}^*(t_p)$, además el ER_p es siempre menor que uno, lo que nos asegura que la dispersión del estimador $\widehat{S}(t_p)_{\text{Copula}}$ frente a $S(t_p)$ es siempre menor que la dispersión del estimador $\widehat{S}^*(t_p)$ frente a $S(t_p)$.

5.2. Comentarios sobre las marginales y la bondad de ajuste

5.2.1. Estimación marginal

Se compara la función de distribución marginal para el tiempo X , con las funciones de distribución marginales estimadas $\widehat{S}_1(t)_{\text{Copula}}$ y $\widehat{S}_1^*(t)$.

La figura 5.43 muestra los intervalos de confianza puntuales empíricos de la distribución marginal $S_1(t)$, basados en el método de simulación montecarlo, para tiempos de falla bivariados Weibull con parámetros de forma $\beta_1 = 0.5$; $\beta_2 = 2$ y dependencia $\lambda = 0.6$ estimados mediante el algoritmo cópula gráfico (gráfico de la izquierda) y mediante la distribución “marginal” estimada $S_1(t)$, que asume independencia entre los modos de falla X y Y (gráfico de la derecha).

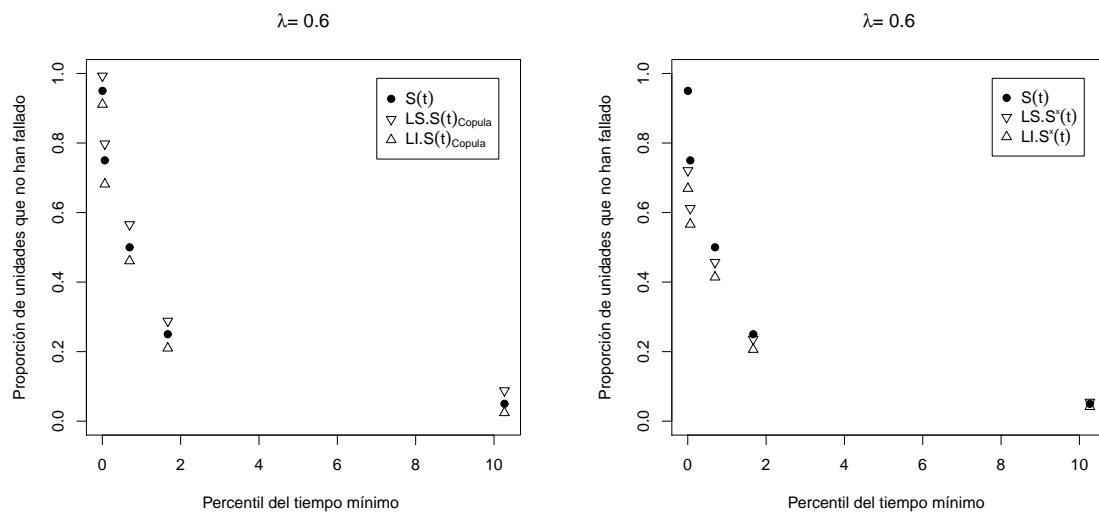


Figura 5.43: Intervalos de confianza empíricos aproximados del 95 % para la distribución marginal $S_1(t_p)$ para el caso Weibull $\beta_1 = 0.5$; $\beta_2 = 2$ y $\lambda = 0.6$ en los percentiles t_p con $p = 0.05, 0.25, 0.50, 0.75$ y 0.95

Se observa como el estimador que emplea el algoritmo cópula gráfico estima la marginal ya que $S_1(t_p)$ cae dentro de los límites del intervalo empírico, mientras que el otro estimador realmente no estima marginales puesto que los límites del intervalo construido no contiene la marginal $S_1(t)$ como era de esperarse, ya que el algoritmo cópula gráfico resuelve el problema de identificabilidad al asumir una cópula y la dependencia.

5.2.2. Comentarios sobre la bondad de ajuste

Las pruebas de bondad de ajuste en general carecen de poder en la práctica en aplicaciones de sobrevivencia (y otras) con cantidades pequeñas o moderadas de datos y fuerte censura. En estas situaciones, la bondad de las pruebas formales sólo sirven para detectar grandes desviaciones de un modelo. Con muestras de gran tamaño, sin embargo, uno puede fácilmente rechazar un modelo que puede ser apropiado para el propósito que nos ocupa. Debido a estas características, las pruebas formales de bondad de ajuste no se utilizan en

sobrevivencia (Meeker and Escobar, 2009).

Siguiendo el argumento presentado por Meeker and Escobar (2009), en lugar de realizar pruebas formales como una medida de ajuste en este trabajo, realizamos pruebas de hipótesis para cada percentil utilizando como región de aceptación el intervalo de confianza

A manera de chequeo se hizo la prueba de Cramer Von Mises para tamaños de muestra de $n = 40$ (tablas 4.8 y 4.9) y $n = 500$ (tablas 4.10 y 4.11). En el primer caso se aceptó la igualdad de distribución para la sobrevivencia verdadera estimada con los dos métodos estudiados y para el segundo caso ambas rechazan.

5.2.2.1. Prueba para los tiempos Weibull y Lognormal con $n = 40$

Las tablas 5.8 y 4.9 muestran lo p-valores para la prueba de Cramer Von Mises que compara igualdad de distribución para la sobrevivencia verdadera $S(t)$ contra la confiabilidad estimada mediante $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ con $n = 40$, para los casos de estudio Weibull y Lognormal. En todos los casos se acepta la hipótesis de igualdad de distribución de la confiabilidad verdadera con la sobrevivencia estimada con los estimadores $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$.

λ	$\beta_1 = 0,5$ y $\beta_2 = 2$		$\beta_1 = 0,5$ y $\beta_2 = 1$		$\beta_1 = 2$ y $\beta_2 = 1$		$\beta_1 = \beta_2 = 0,5$	
	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$
0.0	0.7920	0.2990	0.6732	0.6039	0.1782	0.3891	0.8613	0.9306
0.2	0.8415	0.1881	0.2079	0.5841	0.9900	0.7722	0.8019	0.5841
0.4	0.9702	0.2970	0.2673	0.9702	0.2475	0.4554	0.8019	0.5742
0.6	0.1089	0.8514	0.1584	0.9306	0.3267	0.6930	0.9900	0.9702
0.8	0.2178	0.1782	0.6831	0.2871	0.2871	0.9504	0.2574	0.0990
1.0	0.3069	0.4356	0.1584	0.4653	0.8009	0.9009	0.9900	0.8415

Cuadro 5.8: P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Weibull con tamaño de muestra $n = 40$

	$\mu_1 = \mu_2 = 0$ y $\sigma_1 = 1; \sigma_2 = 2$		$\mu_1 = 0; \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 1$		$\mu_1 = 0, \mu_2 = 1$ y $\sigma_1 = 2; \sigma_2 = 1$	
ρ	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$
0.0	0.6932	0.6239	0.4752	0.3891	0.6554	0.4306
0.2	0.5792	0.6141	0.7900	0.5722	0.6089	0.4843
0.4	0.3673	0.6702	0.4451	0.4154	0.4619	0.5182
0.6	0.3484	0.5306	0.4267	0.4930	0.3900	0.4707
0.8	0.2331	0.2911	0.2271	0.4513	0.2377	0.2912
1.0	0.1784	0.3653	0.3019	0.3021	0.2912	0.2425

Cuadro 5.9: P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Lognormal con tamaño de muestra $n = 40$

5.2.2.2. Prueba para los tiempos Weibull y Lognormal con $n = 500$

Las tablas 5.10 y 5.11 muestran lo p-valores para la prueba de Cramer Vonm Mises que compara igualdad de distribución para la sobrevivencia verdadera $S(t)$ contra la sobrevivencia estimada mediante $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ con $n = 500$, para los casos de estudio Weibull y Lognormal. En todos los casos se rechaza la hipótesis de igualdad de distribución de la sobrevivencia verdadera con la sobrevivencia estimada con los estimadores $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$

	$\beta_1 = 0,5$ y $\beta_2 = 2$		$\beta_1 = 0,5$ y $\beta_2 = 1$		$\beta_1 = 2$ y $\beta_2 = 1$		$\beta_1 = \beta_2 = 0,5$	
λ	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$
0,0	0.0387	0.0123	0.0392	0.0212	0.0234	0.0321	0.0321	0.0356
0,2	0.0032	0.0254	0.0176	0.0123	0.0191	0.0123	0.0245	0.0321
0,4	0.0030	0.0123	0.0037	0.0025	0.0012	0.0145	0.0212	0.0359
0,6	0.0007	0.0043	0.0032	0.0012	0.0023	0.0213	0.0129	0.0213
0,8	0.0087	0.0054	0.0016	0.0014	0.0015	0.0216	0.0256	0.0298
0,9	0.0065	0.0001	0.0018	0.0012	0.0045	0.0121	0.0189	0.0201

Cuadro 5.10: P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Weibull con tamaño de muestra $n = 500$

	$\mu_1 = \mu_2 = 0$ y $\sigma_1 = 1; \sigma_2 = 2$		$\mu_1 = 0; \mu_2 = 1$ y $\sigma_1 = \sigma_2 = 1$		$\mu_1 = 0, \mu_2 = 1$ y $\sigma_1 = 2; \sigma_2 = 1$	
ρ	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$	$\widehat{S}^*(t)$	$\widehat{S}(t)_{\text{Copula}}$
0,0	0.0392	0.0202	0.0286	0.0319	0.0271	0.0326
0,2	0.0176	0.0227	0.0175	0.0173	0.0155	0.0221
0,4	0.0037	0.0125	0.0012	0.0142	0.0118	0.0159
0,6	0.0032	0.0182	0.013	0.0273	0.0021	0.0123
0,8	0.0016	0.0145	0.0125	0.0215	0.0016	0.0098
1,0	0.0018	0.0101	0.0115	0.0141	0.0019	0.0100

Cuadro 5.11: P_{valor} para las pruebas analíticas de bondad de ajuste de las distribuciones de $\widehat{S}^*(t)$ y $\widehat{S}(t)_{\text{Copula}}$ para tiempos asociados a dos modos de falla Lognormal con tamaño de muestra $n = 500$

Capítulo 6

Conclusiones y trabajo futuro

Del estudio realizado se obtienen las siguientes conclusiones:

1. En todos los casos estudiados el estimador $\widehat{S}^*(t_p)$ es ineficiente frente al estimador $\widehat{S}(t_p)_{\text{Copula}}$.
2. En todos los casos estudiados, para estimar $S(t_p)$, el estimador $\widehat{S}(t_p)_{\text{Copula}}$ es mejor estimador que $\widehat{S}^*(t_p)$. Por ello recomendamos para trabajos de sobrevivencia que estén dentro del marco de los escenarios estudiados, utilizar el estimador cópula gráfico, cuando no se tenga certeza sobre el supuesto de independencia.
3. Manotas et. al (2008) encontraron que el supuesto de independencia en el contexto de riesgos competitivos no genera errores significativos en la estimación de la sobrevivencia. Pero para los casos estudiados en este trabajo, la estimación de la sobrevivencia bajo este supuesto de independencia puede subestimar o sobreestimar la sobrevivencia verdadera ya que los límites puntuales de los intervalos de confianza empíricos son más amplios en comparación a los encontrados mediante el estimador cópula gráfico, lo que puede generar sobrecostos y mala planeación en diseño, por ejemplo, en mantenimiento de equipos si se ignora la dependencia.

6.1. Trabajo futuro sobre predicción

En problemas de ingeniería donde se estudia la confiabilidad de algún sistema, es de interés realizar cambios en los parámetros de las marginales para mejorar y predecir la confiabilidad conjunta de dicho sistema. En esta sección se realizan cambios en uno de los parámetros de forma de las dos distribuciones y el mismo cambio se le hace al estimador

de ese parámetro en $\widehat{S}^*(t)$, y se estima el cambio porcentual con relación a la confiabilidad verdadera en los percentiles $p = 0.05, 0.25, 0.50, 0.75, 0.95$.

Un argumento a favor de este tipo de análisis es presentado en (Meeker & Escobar, 2009). "Los ingenieros de diseño tienen herramientas que les permiten obtener predicciones razonables acerca de los efectos que un simple cambio geométrico o cambios de tamaño de una pieza de un producto mecánico tendrá sobre la confiabilidad", y esto se puede traducir en cambios en los parámetros de las marginales.

Con este objetivo, se realizó un estudio de simulación para mostrar el efecto de un cambio en el parámetro β_2 de la distribución Weibull para predecir la nueva confiabilidad de acuerdo a dicho cambio. El esquema de simulación se diseñó siguiendo lo hecho por Meeker & Escobar (2009). La función de confiabilidad conjunta de la Weibull $S(t)$ depende de los parámetros $(\beta_1, \beta_2, \lambda)$ (con $\theta_1 = \theta_2 = 1$, de acuerdo a lo fijado en la sección 4.3) y se realiza un cambio en β_2 multiplicándolo por 0.3, obteniendo una función de confiabilidad nueva que denotamos por $S(t)_{\text{nuevo}} = S(t; \beta_1, 0.3\beta_2, \lambda)$. Ahora bien, el estimador bajo independencia nuevo (i.e. con el cambio en $0.3\beta_2$) se obtiene $\widehat{S}^*(t)_{\text{nuevo}} = \widehat{S}_1(t)\widehat{S}_2(t)_{\text{nuevo}}(t; \widehat{\theta}_2, 0.3\widehat{\beta}_2)$, donde $\widehat{S}_2(t)_{\text{nuevo}}$ se obtiene al reemplazar el estimador $\widehat{\beta}_2$ por $0.3\widehat{\beta}_2$, donde $\widehat{\beta}_2$ es obtenido de los datos de la distribución marginal 2 y el estimador $\widehat{\theta}_2$ se deja igual; como se describió en la sección 2.2. Para comparar se obtuvieron los cambios porcentuales (CP_p), así

$$CP_p = \frac{|S(t_p) - \widehat{S}^*(t_p)|}{S(t_p)} \times 100.$$

que corresponde al cambio relativo entre el valor verdadero $S(t_p)$ y el valor estimado $\widehat{S}^*(t_p)$ antes de hacer cambios en el parámetro β_2 donde $p = 0.05, 0.25, 0.50, 0.75, 0.95$. El cambio porcentual nuevo $CP_p(\text{nuevo})$ se obtiene, así

$$CP_p(\text{nuevo}) = \frac{|S(t_p)_{\text{nuevo}} - \widehat{S}^*(t_p)_{\text{nuevo}}|}{S(t_p)_{\text{nuevo}}} \times 100.$$

En las tablas 6.1 a 6.4 se muestran los cambios porcentuales CP_p y $CP_p(\text{nuevo})$.

λ	$p = 0.05$		$p = 0.25$		$p = 0.50$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p
0.0	1.14279	1.2891	1.12186	1.1097	1.0822	1.0714
0.2	71.8997	49.3511	16.3558	15.2779	7.4906	3.7173
0.4	147.1947	79.6577	34.1554	31.4991	15.9938	5.8585
0.6	226.0480	97.6448	51.7345	49.1091	20.1146	7.5282
0.8	348.2726	99.9365	78.7846	74.7928	22.5856	11.5464
0.9	379.5557	99.9736	83.7510	81.8147	31.7949	12.1519
0.98	389.0409	99.9731	93.4373	82.6682	35.2286	15.7612

λ	$p = 0.75$		$p = 0.95$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p
0.0	1.0234	0.9419	0.4668	0.3255
0.2	2.3615	2.1102	0.9193	0.8953
0.4	5.3259	2.8302	1.3844	1.2676
0.6	3.6267	3.5698	1.8549	1.5949
0.8	6.5551	3.7108	1.9075	1.8547
0.9	8.3761	4.0974	3.0218	2.9878
0.98	11.8975	6.0977	3.8366	3.7381

Cuadro 6.1: Cambios porcentuales CP_p y $CP_p(\text{nuevo})$ para $S(t)$ y $S(t)_{\text{nuevo}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0.5$; $\beta_2 = 2$ y $\theta_1 = \theta_2 = 1$

λ	$p = 0.05$		$p = 0.25$		$p = 0.50$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p
0.0	3.932	0.711	1.161	0.253	1.111	0.594
0.2	58.571	43.195	19.991	16.945	10.970	9.564
0.4	107.588	75.660	35.948	32.507	17.049	11.797
0.6	214.397	94.884	58.399	51.368	28.690	20.310
0.8	252.888	99.821	75.933	63.525	32.522	22.283
0.9	305.851	99.986	86.827	69.369	37.279	22.370
0.98	363.355	99.997	90.752	69.662	39.789	23.979

λ	$p = 0.75$		$p = 0.95$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p
0.0	0.917	0.194	0.829	0.166
0.2	4.956	2.385	1.045	0.856
0.4	5.943	3.560	1.133	1.119
0.6	6.020	4.997	1.831	1.840
0.8	6.430	5.800	5.506	2.302
0.9	8.557	6.364	5.884	2.421
0.98	12.852	7.406	8.162	3.101

Cuadro 6.2: Cambios porcentuales CP_p y $CP_p(\text{nuevo})$ para $S(t)$ y $S(t)_{\text{nuevo}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 0.5$; $\beta_2 = 1$ y $\theta_1 = \theta_2 = 1$

λ	$p = 0.05$		$p = 0.25$		$p = 0.50$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p
0.0	0.320	0.460	0.211	0.226	0.129	0.181
0.2	53.181	38.460	18.775	17.600	9.908	8.182
0.4	80.371	59.279	34.664	33.698	19.026	18.435
0.6	80.783	76.157	52.154	51.307	30.424	27.723
0.8	102.197	88.443	64.439	63.088	38.123	35.652
0.9	105.833	90.799	74.269	67.416	42.105	38.702
0.98	109.546	95.693	74.504	68.295	50.362	42.158

λ	$p = 0.75$		$p = 0.95$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p
0.0	0.125	0.138	0.064	0.056
0.2	4.213	3.776	1.003	0.426
0.4	8.100	8.048	2.344	1.512
0.6	14.645	12.371	2.934	1.797
0.8	16.036	13.915	3.592	3.355
0.9	21.262	19.480	4.343	3.598
0.98	25.654	23.632	4.992	4.331

Cuadro 6.3: Cambios porcentuales CP_p y $CP_p(\text{nuevo})$ para $S(t)$ y $S(t)_{\text{nuevo}}$ para tiempos de falla Weibull con parámetros $\beta_1 = \beta_2 = 0.5$ y $\theta_1 = \theta_2 = 1$

λ	$p = 0.05$		$p = 0.25$		$p = 0.50$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p	$CP_{p(0.3)}$	CP_p
0.0	6.557	6.291	2.322	2.247	2.212	2.086
0.2	47.728	40.560	15.924	14.806	6.299	4.078
0.4	87.381	73.002	35.515	34.022	14.082	13.712
0.6	108.797	94.014	52.541	52.072	22.641	18.171
0.8	110.576	99.610	70.628	69.324	26.667	23.315
0.9	111.101	99.997	84.700	84.507	20.273	14.017
0.98	111.996	99.997	85.186	84.904	24.480	19.198

λ	$p = 0.75$		$p = 0.95$	
	$CP_p(\text{nuevo})$	CP_p	$CP_p(\text{nuevo})$	CP_p
0.0	1.588	1.586	0.626	0.381
0.2	2.582	1.878	0.793	0.583
0.4	3.210	3.143	1.121	0.747
0.6	3.225	3.218	2.961	2.271
0.8	6.231	6.076	4.764	3.455
0.9	6.383	6.288	4.473	3.550
0.98	7.186	7.173	5.146	3.631

Cuadro 6.4: Cambios porcentuales CP_p y $CP_p(\text{nuevo})$ para $S(t)$ y $S(t)_{\text{nuevo}}$ para tiempos de falla Weibull con parámetros $\beta_1 = 2$; $\beta_2 = 1$ y $\theta_1 = \theta_2 = 1$

Analizando por columna para cada percentil, se observa que a medida que aumenta la dependencia λ entre los tiempos de falla Weibull, el cambio porcentual CP_p es siempre menor que el cambio porcentual $CP_p(\text{nuevo})$. En la tabla 6.1 para el percentil $t_{0.05}$ la diferencia entre los dos llega a ser hasta de casi cuatro veces. De lo cual se puede decir que en la práctica el estimador $\widehat{S}^*(t)_{\text{nuevo}}$, que es el estimador de la confiabilidad bajo independencia, no logra captar correctamente el cambio en el parámetro β_2 , luego para efectos de predicción en este contexto es muy mal estimador. En esta situación lo que ocurre es que $\widehat{S}_2(t)_{\text{nuevo}}$ se calcula cambiando $\widehat{\beta}_2$ obtenido de los datos simulados para la marginal del modo de falla 2, por $0.3\widehat{\beta}_2$, pues al no tener datos nuevos, se tienen una estimación inapropiada de $S_2(t)_{\text{nuevo}}$. Por ello afirmamos que $\widehat{S}^*(t)_{\text{nuevo}}$ no es útil para predecir cuando hay cambios en los parámetros.

Obsérvese además, que el estimador $\widehat{S}(t)_{\text{cópula}}$ no sirve para fines de predicción, ya que este estimador es no paramétrico y al hacer cambios en los parámetros de las marginales no hay forma de incluir dicho cambio en el estimador. Pero se puede pensar que con las estimaciones de las distribuciones marginales obtenidas con el estimador cópula-gráfico se puede hacer un modelamiento paramétrico, como paso previo al proceso de predicción de

la confiabilidad bajo cambios en los parámetros, lo cual mejoraría el análisis exploratorio, ya que se utilizan estimaciones adecuadas de las distribuciones marginales en lugar de las estimaciones basadas en asumir independencia. Estos ajustes paramétricos permitirían hacer predicciones en experimentos acelerados como los descritos en Meeker and Escobar (2009).

Así, para efectos de predicción en problemas de Ingeniería que implican cambios en los parámetros de las marginales para mejorar la confiabilidad conjunta, ninguno de los dos estimadores funciona. Este sería un tópico de importancia para trabajo futuro tratando de adecuar el estimador cópula gráfico para tales efectos, y comparar con resultados como los reportados por Meeker & Escobar (2009). Se piensa que con las distribuciones marginales estimadas se pueden hacer ajustes paramétricos que permitan incluir la dependencia en la estimación de la predicción de la vida útil de un equipo sujeto a cambios en su diseño original, en el contexto de experimentos acelerados.

Apéndice A

Programa Cópula Gráfico

```
#####
```

Este programa estima las marginales F y G, también la función de confiabilidad con la cópula

```
#####
```

```
datos=read.table("datos.txt")
tao=0.4 # Dependencia de los tiempos de falla
alpha= 1+1/tao # Dependencia de la copula
t1=datos$V1
t2=datos$V2
delta = (t1<t2)*1
T = apply(cbind(t1,t2),1,min)
ti = sort(T)
n = length(ti)
desp = function(i,n,ant) (1-((1-i/n)^(1-alpha)+1-(1/(1-ant))^(alpha-1))^(1/(1-alpha)))
Ti = Ft = Gt = error = 0
for(i in 1:(n+1)){
  if(i == 1) Gt[1] = Ft[1] = Ti[i] = 0
  if(i>1){
    Ti[i] = ti[i-1]
    if(di[i-1]==1) { Gt[i] = Gt[i-1] ; Ft[i] = ifelse(i==2,..05,Ft[i-1]) }
    if(di[i-1]==0) { Ft[i] = Ft[i-1]; Gt[i] = desp((i-1),n,Ft[i-1]) }
    error=Ft[i]-sum((1-(1-Ft[2:i])^(-alpha))*((1-Ft[2:i])^(1-alpha)+(1-Gt[2:i])
    ^(-alpha)-1)^(alpha/(1-alpha)))*(Ft[2:i]-Ft[2:(i-1)]))-1/n*sum(di[1:(i-1)]==1) Fopt = Ft[i] ; j=1
    if(di[i-1]==1){ while(abs(error) > 0.000001){
      j = j+1
      if(j==2){ b = 1 ; Fopt[j] = (Fopt[j-1]+b)/2 }
      if(j>2){ if(error>0) { a = max(Fopt[Fopt<Fopt[j-1]]) ; Fopt[j] = (a+Fopt[j-1])/2 }
      if(error<0) { b = min(c(Fopt[Fopt>Fopt[j-1]],1)) ; Fopt[j] = (Fopt[j-1]+b)/2 }
      }
      Ft[i]=Fopt[j]
      error=Ft[i]-sum((1-(1-Ft[2:i])^(-alpha))*((1-Ft[2:i])^(1-alpha)+(1-Gt[2:i])
      ^(-alpha)-1)^(alpha/(1-alpha)))*(Ft[2:i]-Ft[2:(i-1)]))-1/n*sum(di[1:(i-1)]==1) }
    }
  }
}
```



```
}  
Ft; Gt  
S_Copula= ((1-Ft)^(1-alpha)+(1-Gt)^(1-alpha)-1)^(1/(1-alpha))
```

Bibliografía

- [1] Barlow, R.E, y Proschan, F. (1975), *Statistical Theory of Reliability and Life Testing*. New York: Holt, Richard Winston.
- [2] Berford, T., (2005). *Competing Risk Modeling in Reliability*. Chapter 1. *Modern Statistical and Mathematical Methods in Reliability*. Singapore: World Scientific Publishing Co.
- [3] Berman, S. M. (1963). Notes on extreme values, competing risks, and semi-Markov process. *Ann. Math. Statist.* **34**, 1104-1106.
- [4] Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., Nesbitt, C.J., (1997). *Actuarial Mathematics*, vol. III. Society of Actuaries. Itasca.
- [5] Bryant, T., (2004). Approaches and experiences in projecting mortality patterns for oldest-old. *North American Actuarial Journal* **6** (3), 14-29.
- [6] Carriere, J., (1994). Dependent decrement theory. *Transactions, Society of Actuaries* XLVI, 45-65.
- [7] Crowder, M. (2001), *Classical Competing Risks*. Chapman Hall/CRC.
- [8] David, H.A and Moeschberger, M.L., (1978), *Theory of Competing Risks*. Griffin.
- [9] Denuit, M. Dhaene, J. Goovaerts, M and Kaas, R., (2005), *Actuarial Theory for Dependent Risks. Measures, Orders and Models*. Great Britain: John Wiley & Sons.
- [10] Elandt, R. Johnson, N., (1980), *Survival Models and Data Analysis*. John Wiley & Sons.
- [11] Escarela, G., Carriere, J., (2003) Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research* **12** (4), 333-349.
- [12] Fréchet, M. (1951), Sur les Tableaux de Corr-elation dont les Marges sont Donnés. *Annales de l'Université de Lyon* (**14**), 53-77.

- [13] Genest, C. & Rivest, L. (1993), Statistical Inference Procedures for Bivariate Archimedian Copula. *Journal of the American Statistical Association* **88**, 1034-1043.
- [14] Johnson, M. (1987). Multivariate Statistical Simulation. *John Wiley Sons*.
- [15] Lu, J. & Bhattacharryya, G. K. (1990), Some New Constructions of Bivariate Weibull Models. *Annals of the Institute of Statistical Mathematics* **42**(3), 543-559.
- [16] Manotas, E., Yáñez, S., Lopera C.M., Jaramillo, M., (2008) Estudio del Efecto de la Dependencia en la Estimación de la Confiabilidad de un Sistema con dos Modos de Falla Concurrentes. *DYNA*, Medellín, **v. 75, n. 154**, 5-21.
- [17] Meeker, W. y Escobar, L.A., (1998), Statistical Methods for Reliability Data. John Wiley & Sons.
- [18] Miller, D. R. (1977). A note on independence on multivariate lifetimes in competing risks models. *Ann. Statist.* **5**, 576-579.
- [19] Nelsen, R.B., (1999) An Introduction to Copulas. *Springer*.
- [20] Peterson, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. Am. Statist. Assoc.* **73**, 11-13.
- [21] Rausand, M. and Hoyland, A., (2004), System Reliability Theory. *John Wiley & Sons*.
- [22] Salinas-Torres, V.H., de Branca Pereira, C.A., Tiwardi, R.C., (2002). Bayesian non-parametric estimation in a series or a competing risk model. *Journal on Nonparametric Statistics* **14**, 449-458.
- [23] Schweizer, B. & Sklar, A. (1983), Probabilistic Metric Spaces, *Dover Publications*, New York, United States.
- [24] Tsiatis, A., (1975), A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci. USA.* **72**, 20-22.
- [25] Yanagimoto, T. & Okamoto, M. (1969), Partial Orderings of Permutations and Monotonicity of rank a Rank Correlation Statistic. *Annals of the Institute of Statiscal Mathematics* **21**, 489-506.
- [26] Yáñez, S., Manotas, E., Lopera, C., (2006). Estudio de la Dependencia en la Estimación de la Confiabilidad de un Sistema con dos Modos de Falla, *Memorias del XV Simposio de Estadística, Paipa*.

- [27] Zheng, M., Klein, J.P., (1994). A self-consistent estimator of marginal survival functions based on dependent competing risk data and assumed copula. *Communications in Statistics- Theory and Methods* **23 (8)**, 2299-2311.
- [28] Zheng, M., Klein, J.P., (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82 (1)**, 12-38