



Centro de Investigación en Matemáticas, A.C.

CIMAT

**ESTUDIO DEL AIC Y BIC EN LA
SELECCIÓN DE MODELOS DE VIDA
CON DATOS CENSURADOS**

que para obtener el grado de

Maestro en Ciencias con Especialidad en
Probabilidad y Estadística

P r e s e n t a

Abelardo Montesinos López

Directores de Tesina

Dr. Enrique Raúl Villa Diharce

Dr. Luis A. Escobar R.

Guanajuato, Gto. Agosto de 2011

Índice general

Índice general	I
Agradecimientos	III
1 Introducción	1
2 Criterio de información AIC y BIC	3
2.1 Introducción	3
2.2 Distancia de Kullback-Leibler	3
2.2.1 Estimación por máxima verosimilitud y su relación con la información de K-L	5
2.2.2 Derivación alternativa del índice AIC, Konishi y Kitagawa (2008)	6
2.2.3 Akaike (1974)	9
2.3 Criterio de información Bayesiana (BIC)	10
3 Estudio de simulación	13
3.1 Introducción	13
3.2 Implementación del mecanismo de censura tipo I	14
3.2.1 Distribuciones de uno y dos parámetros	14
3.2.2 Distribuciones de dos parámetros	25
3.2.3 Distribuciones de uno, dos y tres parámetros	41
4 Conclusiones	49
A Algunas distribuciones de tiempo de vida	52
A.1 Distribución Weibull	52
A.2 Distribución Lognormal	53
A.3 Distribución Loglogística	54

A.4	Distribución Inversa Gausiana	54
A.5	Distribución Gama Generalizada	55
B	Algunos resultados	56
C	Anexos	58
	Bibliografía	60

Agradecimientos

A mi familia, amigos y a todas aquellas personas que directa e indirectamente contribuyeron en la realización de esta tesina.

En particular a los Dres. Enrique Villa y Luis Escobar, por sus consejos y paciencia.

Al departamento de Probabilidad y Estadística del CIMAT por la oportunidad que me brindó.

Al CONACYT por la beca que recibí durante mis estudios de maestría.

Al Dr. Rogelio Ramos Quiroga y al Dr. Jorge Domínguez y Domínguez por sus valiosos comentarios a este trabajo.

A la Dra. Angélica Hernández Quintero por su apoyo incondicional.

CAPÍTULO 1

Introducción

En el trabajo de modelación estadística, es de primordial importancia la selección del modelo, es decir, elegir dentro de un conjunto de modelos alternativos el modelo más apropiado para el conjunto disponible de datos. Por ejemplo, en teoría de valores extremos algunas veces se desea elegir entre la distribución generalizada de valores extremos con un parámetro de forma muy pequeño o una distribución gumbel, donde ésta última se toma como un caso límite de la primera cuando el parámetro de forma tiende a cero. En tal caso es deseable un estadístico que permita seleccionar entre un modelo u otro. Los índices AIC y BIC (Criterio de información de Akaike y criterio de información bayesiano, respectivamente) son dos criterios de uso frecuente para la selección de modelos. El AIC fue propuesto por Akaike (1974) como un estimador insesgado asintótico de la información de Kullback-Leibler esperada, entre un modelo candidato ajustado y el verdadero modelo. El BIC fue derivado por Schwarz en 1978 como una aproximación a una transformación de la probabilidad posterior de un modelo candidato.

A través del tiempo el uso de ambos criterios para la selección de modelos ha crecido significativamente. Entre algunas de las primeras aplicaciones del AIC sugeridas por el autor de éste índice, Akaike, se encuentran el análisis factorial, análisis de componentes principales, regresión múltiple y series de tiempo. Otras aplicaciones recientes de ambos criterios

también se tienen en ecología (Anderson *et al.*, 1994; Johnson y Omland, 2004; Dennis *et al.*, 2006; Ponciano *et al.*, 2009) y bioinformática (Edwards *et al.*, 2010; Abreu *et al.*, 2010), por mencionar algunas. Así mismo, en los últimos años se han introducido en el área de confiabilidad, donde con frecuencia se usan con datos censurados, a pesar que ambos criterios no fueron construidos para modelos con datos censurados. Por esta razón, en el presente trabajo hacemos un estudio del desempeño de los criterios de información antes mencionados, considerando datos con censura por tiempo (Tipo I). En este estudio, consideramos los modelos de probabilidad más comunes en el análisis de datos de vida, como son: Weibull, Lognormal, Loglogística, Gaussiana inversa, Rayleigh, Exponencial y Gama Generalizada. Utilizamos un diseño experimental, donde los tamaños muestrales son 20, 30, 50 y 100, y las proporciones de censura son 0, 0.05, 0.10, 0.15, 0.20, 0.30., 0.40 y 0.50. A través de un estudio de simulación evaluamos la proporción de veces que el procedimiento (AIC y BIC) selecciona a cada uno de los modelos que compiten, considerando conjuntos de datos generados de uno de ellos.

El trabajo está organizado de la siguiente manera: En el Capítulo 2 se presenta una breve introducción de la información de Kullback-Leibler y su conexión con el AIC. Luego se presenta un bosquejo detallado de las ideas principales de una derivación alternativa del índice AIC por parte de Konishi y Kitagawa (2008), así como también las ideas del autor original, Akaike (1974). Por último, se definen los elementos necesarios que permiten dar una idea clara de la derivación del índice BIC. En el Capítulo 3 se presenta el estudio de simulación, donde comparamos el desempeño de los índices. Se describen los pasos generales que se siguen para su implementación. Primero se presentan y discuten los resultados que se obtienen al considerar el problema de discriminar entre los modelos Weibull, Rayleigh y Exponencial, con un número diferente de parámetros. Luego se discuten los resultados de la discriminación para cuatro modelos con dos parámetros: Weibull, Lognormal, Loglogística y Gaussiana inversa. También se estudia el desempeño de ambos criterios cuando las distribuciones candidatas son la Weibull, Rayleigh, Exponencial, Lognormal y Gama Generalizada. Finalmente, en el Capítulo 4 se presentan las conclusiones obtenidas en el estudio.

Criterio de información AIC y BIC

2.1. Introducción

En este Capítulo se discute la información de Kullback-Leibler (K-L) como un criterio para la evaluación estadística de modelos que aproximan a la verdadera distribución de probabilidad que genera los datos, y se dan algunas de sus propiedades teóricas más importantes que se relacionan al AIC. Luego, por un lado se describen las ideas principales de Konishi y Kitagawa (2008) de como el criterio de información estadística (K-L) lleva al concepto de información AIC. Por otro, también se describen las ideas del autor original Akaike (1974), para la derivación del AIC. Además, se define el criterio de información Bayesiana (BIC) con todos sus elementos, y las ideas principales de como llegar a este.

2.2. Distancia de Kullback-Leibler

La estimación de la información de Kullback-Leibler es el punto clave en la derivación del criterio de información AIC, el cual ha sido ampliamente usado para la selección de modelos estadísticos. La información de Kullback-Leibler definida en (2.1) es considerada como una medida de bondad de ajuste del modelo propuesto $f(x)$ hacia el modelo verdadero $g(x)$, ver Shibata (1995).

$$I(g, f) = \int \log \left[\frac{g(x)}{f(x)} \right] g(x) dx = E_X \left[\log \left(\frac{g(X)}{f(X)} \right) \right] \quad (2.1)$$

donde E_X denota que la esperanza es tomada con respecto a la variable aleatoria X .

Algunas de las propiedades de la información de K-L son: (i) $I(g, f) \geq 0$, (ii) $I(g, f) = 0 \Leftrightarrow g(x) = f(x)$, es decir, la información de K-L siempre es positiva, excepto cuando las dos distribuciones son iguales (Burnham y Anderson, 2002, página 430). De aquí, esta información puede interpretarse directamente como una “distancia” entre dos modelos, en este caso $f(x)$ y $g(x)$, aunque estrictamente no lo sea, ya que la medida de f a g no necesariamente es la misma que de g a f .

Aunque la información K-L es bastante razonable para evaluar qué tan adecuado es un modelo dado, en la práctica es bastante limitada ya que casi siempre se desconoce la verdadera distribución que genera los datos, lo cual impide calcular (2.1).

La ecuación (2.1) se puede re-expresar como

$$I(g, f) = E_X[\log(g(X))] - E_X[\log(f(X))] \quad (2.2)$$

de donde se tiene que, para la comparación de diferentes modelos es suficiente considerar $E_X[\log(f(X))]$, ya que $E_X[\log(g(X))]$ es un término común que puede ser ignorado. El segundo término de (2.2) se conoce como log-verosimilitud esperada por unidad para el modelo f . Así, de un conjunto de modelos candidatos el modelo que tenga mayor log-verosimilitud esperada es el que corresponde al que tiene menor información de K-L, y en consecuencia es el mejor modelo.

Si el modelo $f(x)$ está completamente especificado, entonces obsérvese que un estimador natural para $E_X[\log(f(X))]$ es

$$\frac{1}{n} \sum_{i=1}^n \log [f(X_i)] \quad (2.3)$$

donde $\mathbf{X}_n = (X_1, \dots, X_n)'$ es una muestra aleatoria de la verdadera distribución $g(x)$. Notése que (2.3) es un estimador insesgado para $E_X[\log(f(X))]$, y cuando n tiende a infinito converge a $E_X[\log(f(X))]$ con probabilidad 1, asumiendo que $|E_X[\log(f(X))]| < \infty$.

2.2.1. Estimación por máxima verosimilitud y su relación con la información de K-L

En situaciones prácticas trabajar con modelos completamente especificados no es común. Lo habitual es asumir un modelo paramétrico $\{f(x|\theta); \theta \in \Theta \subset R^p\}$ y luego estimar los parámetros θ por el “método de máxima verosimilitud”. Aún más, debido a que muchas veces no se tiene bien identificado un modelo, lo usual es proponer varios modelos paramétricos para el mismo problema, lo que posteriormente se hace en el estudio del AIC y BIC.

Se sabe que bajo ciertas condiciones de regularidad el estimador de máxima verosimilitud (EMV) de θ , $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X}_n)$, converge en probabilidad a $\theta_0 = \arg \min_{\theta \in \Theta} I[g(\cdot), f(\cdot|\theta)]$, es decir,

$$\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{p} \theta_0. \quad (2.4)$$

$f(x|\theta_0)$ es llamada la mejor aproximación a $g(x)$ (Claeskens y Hjort, 2008). Así, como en la práctica θ_0 es imposible de calcular ya que no se conoce $g(x)$, el emv de θ proporciona la mejor aproximación paramétrica a la verdadera distribución g dentro de la clase paramétrica $f(x|\theta)$. Para cuando θ es un escalar y $g(x) = f(x|\theta^*)$, para algún $\theta^* \in \Theta$, una prueba de (2.4) puede verse en Wasserman (2004). Una forma heurística de ver (2.4) es la siguiente:

Maximizar la log-verosimilitud $\ell_n(\theta) =: \sum_{i=1}^n \log(f(X_i|\theta))$ es equivalente a maximizar

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log \left(\frac{f(X_i|\theta)}{g(X_i)} \right)$$

y además por ley de los grandes números

$$M_n(\theta) \xrightarrow[n \rightarrow \infty]{p} E_g \left[\log \left(\frac{f(X|\theta)}{g(X)} \right) \right] = -E_g \left[\log \left(\frac{g(X)}{f(X|\theta)} \right) \right] = -I[g(\cdot), f(\cdot|\theta)].$$

De aquí, $M_n(\theta) \approx -I[g(\cdot), f(\cdot|\theta)]$. Pero como se mencionó anteriormente, $I[g(\cdot), f(\cdot|\theta)]$ se minimiza en θ_0 , así que $-I[g(\cdot), f(\cdot|\theta)]$ es maximizada en θ_0 . De esta forma, se espera que el θ que maximiza $M_n(\theta)$, $\hat{\theta}_n$, tienda a θ_0 .

2.2.2. Derivación alternativa del índice AIC, Konishi y Kitagawa (2008)

Como ya se dijo, una forma de evaluar la bondad de ajuste de $f(x|\hat{\theta}_n)$ al modelo verdadero $g(x)$, es por medio de la información K-L. Así, por (2.2), la información de K-L entre el modelo $g(x)$ y $f(x|\hat{\theta}_n)$ está dada por

$$I[g(\cdot), f(\cdot|\hat{\theta}_n)] = E_X[\log(g(X))] - E_X[\log(f(X|\hat{\theta}_n))] \quad (2.5)$$

Entonces por lo comentado en (2.2) solamente es de importancia obtener un buen estimador del segundo término del lado derecho de (2.5), $E_X[\log(f(X|\hat{\theta}_n))]$. Una posibilidad es usar la distribución empírica de los datos en vez de $g(x)$, resultando como estimador

$$\hat{E}_X[\log(f(X|\hat{\theta}_n))] = \frac{1}{n} \sum_{i=1}^n \log [f(X_i|\hat{\theta}_n)] = \frac{1}{n} \ell_n(\hat{\theta}_n). \quad (2.6)$$

Este estimador no es necesariamente insesgado para $E_X[\log(f(X|\hat{\theta}_n))]$, debido a que aquí se hace uso de la muestra aleatoria $\mathbf{X}_n = (X_1, \dots, X_n)'$ que inicialmente también se emplea para obtener $\hat{\theta}_n =: \hat{\theta}_n(\mathbf{X}_n)$, el EMV de θ . La forma tradicional de arreglar este problema es obtener el sesgo que se comete por usar (2.6) como estimador de $E_X[\log(f(X|\hat{\theta}_n))]$, y luego restar este a (2.6). Para ello, primero nótese que el sesgo de este estimador está dado por

$$\begin{aligned} \text{Sesgo} &= E_{\mathbf{X}_n} \left[\frac{1}{n} \ell_n(\hat{\theta}_n) - E_X[\log(f(X|\hat{\theta}_n))] \right] \\ &= \frac{1}{n} E_{\mathbf{X}_n} \left[\ell_n(\hat{\theta}_n) - n E_X[\log(f(X|\hat{\theta}_n))] \right] \end{aligned} \quad (2.7)$$

donde $E_{\mathbf{X}_n}$ significa que la esperanza es tomada respecto a la distribución conjunta del vector \mathbf{X}_n . Luego, obsérvese que el último término de la ecuación (2.7) se puede descomponer en tres partes, lo que será importante para estimar el sesgo:

$$\begin{aligned} E_{\mathbf{X}_n} \left[\ell_n(\hat{\theta}_n) - n E_X[\log(f(X|\hat{\theta}_n))] \right] &= E_{\mathbf{X}_n} [\ell_n(\theta_0) - n E_X[\log(f(X|\theta_0))]] \\ &\quad + E_{\mathbf{X}_n} [\ell_n(\hat{\theta}_n) - \ell_n(\theta_0)] \\ &\quad + E_{\mathbf{X}_n} \left[n E_X[\log(f(X|\theta_0))] - n E_X[\log(f(X|\hat{\theta}_n))] \right] \end{aligned} \quad (2.8)$$

A continuación se describe cómo se obtienen cada uno de estos términos, además en ese proceso se emplean resultados que son explicados en el apéndice B.

1. Directamente se obtiene que

$$E_{\mathbf{X}_n} [\ell_n(\theta_0) - nE_X[\log(f(X|\theta_0))]] = 0. \quad (2.9)$$

2. Expandiendo $\ell_n(\theta_0)$ alrededor de $\hat{\theta}_n$ se obtiene que

$$\ell_n(\theta_0) = \ell_n(\hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)' \frac{\partial \ell_n(\hat{\theta}_n)}{\partial \theta} + \frac{1}{2} (\theta_0 - \hat{\theta}_n)' \frac{\partial^2 \ell_n(\hat{\theta}_n)}{\partial \theta \partial \theta'} (\theta_0 - \hat{\theta}_n) + \dots \quad (2.10)$$

Pero $\frac{\partial \ell_n(\hat{\theta}_n)}{\partial \theta} =: \left. \frac{\partial \ell_n(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = 0$, por ser $\hat{\theta}_n$ el EMV de θ . Así, al usar la aproximación $\frac{1}{n} \frac{\partial^2 \ell_n(\hat{\theta}_n)}{\partial \theta \partial \theta'} =: \left. \frac{1}{n} \frac{\partial^2 \ell_n(\theta)}{\partial \theta \partial \theta'} \right|_{\theta=\hat{\theta}_n} \approx -J(\theta_0)$, donde $J(\theta_0)$ está definida en (B.2), entonces de (2.10)

$$\begin{aligned} E_{\mathbf{X}_n} [\ell_n(\hat{\theta}_n) - \ell_n(\theta_0)] &\approx E_{\mathbf{X}_n} \left[\frac{1}{2} (\theta_0 - \hat{\theta}_n)' n J(\theta_0) (\theta_0 - \hat{\theta}_n) \right] \quad (2.11) \\ &= \frac{1}{2} E_{\mathbf{X}_n} \left[\text{Tr} \left\{ \sqrt{n} (\theta_0 - \hat{\theta}_n)' J(\theta_0) \sqrt{n} (\theta_0 - \hat{\theta}_n) \right\} \right] \\ &= \frac{1}{2} E_{\mathbf{X}_n} \left[\text{Tr} \left\{ J(\theta_0) \sqrt{n} (\theta_0 - \hat{\theta}_n) \sqrt{n} (\theta_0 - \hat{\theta}_n)' \right\} \right] \\ &= \frac{1}{2} \text{Tr} \left\{ J(\theta_0) E_{\mathbf{X}_n} \left[\sqrt{n} (\hat{\theta}_n - \theta_0) \sqrt{n} (\hat{\theta}_n - \theta_0)' \right] \right\} \end{aligned}$$

pero como de (B.2) $\sqrt{n}(\hat{\theta}_n - \theta_0) \sim N_p [0, J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}]$,

$E_{\mathbf{X}_n} \left[\sqrt{n}(\hat{\theta}_n - \theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0)' \right] \approx J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}$, y en consecuencia

$$\begin{aligned} E_{\mathbf{X}_n} [\ell_n(\hat{\theta}_n) - \ell_n(\theta_0)] &\approx \frac{1}{2} \text{Tr} \left\{ J(\theta_0) E_{\mathbf{X}_n} \left[\sqrt{n}(\hat{\theta}_n - \theta_0) \sqrt{n}(\hat{\theta}_n - \theta_0)' \right] \right\} \\ &\approx \frac{1}{2} \text{Tr} \left\{ J(\theta_0) J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1} \right\} \\ &= \frac{1}{2} \text{Tr} \left\{ I(\theta_0) J(\theta_0)^{-1} \right\} \quad (2.12) \end{aligned}$$

donde $I(\theta_0)$ está definida en (B.1).

3. Expandiendo $E_X[\log(f(X|\hat{\theta}_n))]$ alrededor de θ_0 se obtiene

$$\begin{aligned} E_X[\log(f(X|\hat{\theta}_n))] &= E_g[\log(f(X|\theta_0))] + (\hat{\theta}_n - \theta_0)' \frac{\partial E_X[\log(f(X|\theta_0))]}{\partial \theta} \\ &\quad + \frac{1}{2} (\hat{\theta}_n - \theta_0)' \frac{\partial^2 E_X[\log(f(X|\theta_0))]}{\partial \theta \partial \theta'} (\hat{\theta}_n - \theta_0) + \dots \quad (2.13) \end{aligned}$$

Pero como

$$\frac{\partial E_X[\log(f(X|\theta_0))]}{\partial \theta} = 0$$

por ser θ_0 el que minimiza $I[g(\cdot), f(\cdot|\theta)]$, y

$$\frac{\partial^2 E_X[\log(f(X|\theta_0))]}{\partial \theta \partial \theta'} = E_X \left[\frac{\partial^2 \log(f(X|\theta_0))}{\partial \theta \partial \theta'} \right] = -J(\theta_0)$$

por (B.2), entonces sustituyendo estos valores en (2.13) se obtiene

$$\begin{aligned} E_{\mathbf{X}_n} \left[nE_X[\log(f(X|\theta_0))] - nE_X[\log(f(X|\hat{\theta}_n))] \right] \\ \approx \frac{n}{2} E_{\mathbf{X}_n} \left[(\hat{\theta}_n - \theta_0)' J(\theta_0) (\hat{\theta}_n - \theta_0) \right] \\ = \frac{1}{2} E_{\mathbf{X}_n} \left[(\hat{\theta}_n - \theta_0)' nJ(\theta_0) (\hat{\theta}_n - \theta_0) \right] \\ \approx \frac{1}{2} \text{Tr} \{ I(\theta_0) J(\theta_0)^{-1} \} \end{aligned} \quad (2.14)$$

donde la última aproximación es similar a la hecha en (2.12).

Por lo tanto, sustituyendo (2.9), (2.12) y (2.14) en (2.8) finalmente se llega a la siguiente aproximación:

$$\begin{aligned} E_{\mathbf{X}_n} \left[\ell_n(\hat{\theta}_n) - nE_X[\log(f(X|\hat{\theta}_n))] \right] &\approx 0 + \frac{1}{2} \text{Tr} \{ I(\theta_0) J(\theta_0)^{-1} \} \\ &+ \frac{1}{2} \text{Tr} \{ I(\theta_0) J(\theta_0)^{-1} \} \\ &= \text{Tr} \{ I(\theta_0) J(\theta_0)^{-1} \} \end{aligned}$$

y por (2.7), el sesgo aproximado es

$$\text{Sesgo} \approx \frac{1}{n} \text{Tr} \{ I(\theta_0) J(\theta_0)^{-1} \} \quad (2.15)$$

De esta manera restando (2.15) de (2.6), el estimador corregido aproximadamente insesgado para $E_X[\log(f(X|\hat{\theta}_n))]$ está dado por

$$\hat{E}_X^*[\log(f(X|\hat{\theta}_n))] = \frac{1}{n} \ell_n(\hat{\theta}_n) - \frac{1}{n} \text{Tr} \{ I(\theta_0) J(\theta_0)^{-1} \}.$$

Así, si se tiene un conjunto de modelos candidatos para un fenómeno de estudio, el mejor candidato será el que tenga mayor $\hat{E}_X^*[\log(f(X|\hat{\theta}_n))]$, lo que equivale a decir, el que tenga menor $-2n\hat{E}_X^*[\log(f(X|\hat{\theta}_n))]$. Notése que si el modelo verdadero $g(x)$ se encuentra anidado en $f(x|\theta)$, por el Resultado B.1 $I(\theta_0) = J(\theta_0)$, y de este modo $-2n\hat{E}_X^*[\log(f(X|\hat{\theta}_n))]$ se reduce al conocido criterio de información de Akaike:

$$\text{AIC} = -2\ell_n(\hat{\theta}_n) + 2p \quad (2.16)$$

2.2.3. Akaike (1974)

En la derivación del AIC por Akaike (1974), de entrada se considera la situación donde $g(x) = f(x|\theta_0)$, es decir, la densidad de probabilidad $g(x)$ verdadera se encuentra incluida en la familia dada, $\{f(x|\theta) : \theta \in \Theta \subset \mathbf{R}^p\}$. Si $K(\theta_0, \theta)$ denota $I(g, f(\cdot|\theta))$ y además si θ está suficientemente cercano a θ_0 , $K(\theta_0, \theta)$ se puede aproximar por (Kullback, 1977, página 28)

$$K(\theta_0, \theta_0 + \Delta\theta) \approx \frac{1}{2} \Delta\theta' I(\theta_0) \Delta\theta \quad (2.17)$$

donde

$$I(\theta_0) = \int g(x) \frac{\partial \log[f(x|\theta_0)]}{\partial \theta} \frac{\partial \log[f(x|\theta_0)]}{\partial \theta'} dx$$

y

$$\frac{\partial \log[f(x|\theta_0)]}{\partial \theta} := \left. \frac{\partial \log[f(x|\theta)]}{\partial \theta} \right|_{\theta=\theta_0}.$$

Una forma alternativa de obtener la aproximación en (2.17) es primero expandir por serie de Taylor de segundo grado la función $\log[f(x|\theta_0 + \Delta\theta)]$ alrededor de θ_0 , restar $\log[f(x|\theta_0)]$ y sacar esperanzas, donde para este último paso es necesario recordar que

$$E \left[\left. \frac{\partial \log[f(x|\theta)]}{\partial \theta} \right|_{\theta=\theta_0} \right] = 0.$$

Por (2.17), obsérvese que cuando el EMV $\hat{\theta}_n$ de θ se encuentra cerca a θ_0 , la desviación de la distribución ajustada $f(x|\hat{\theta}_n)$ a la verdadera $f(x|\theta_0)$ puede medirse por $(\hat{\theta}_n - \theta_0)' I(\theta_0) (\hat{\theta}_n - \theta_0) / 2$.

Ahora, considérese el caso en que θ es restringido a un sub-espacio de menor dimensión $\Theta_k \subset \Theta_p$ que no incluye a θ_0 . Si $\hat{\theta}_{n,k}$ denota el EMV de θ en Θ_k , y $\theta_k = \arg \min_{\theta \in \Theta_k} I[g(\cdot), f(\cdot|\theta)]$ está suficientemente cerca a θ_0 , entonces bajo condiciones de regularidad la distribución asintótica de $\sqrt{n}(\hat{\theta}_{n,k} - \theta_k)$ es aproximadamente MVN $[0, I(\theta_0)^{-1}]$, y de aquí la distribución de $n(\hat{\theta}_{n,k} - \theta_0)' I(\theta_0) (\hat{\theta}_{n,k} - \theta_0)$ es aproximadamente Ji-cuadrada no central $\chi_{k,\lambda}^2$, con parámetro de no centralidad $\lambda = n(\theta_k - \theta_0)' I(\theta_0) (\theta_k - \theta_0)$. Así de esto y (2.17)

$$E \left[2nK(\theta_0, \hat{\theta}_{n,k}) \right] \approx n(\theta_k - \theta_0)' I(\theta_0) (\theta_k - \theta_0) + k \quad (2.18)$$

donde k es la dimensión de Θ_k o el número de parámetros libres. De esto, si se cuenta con una estimación de $n(\theta_k - \theta_0)' I(\theta_0) (\theta_k - \theta_0)$ y se tienen varios modelos, es natural adoptar

como mejor el que tenga un valor $E[2nK(\theta_0, \hat{\theta}_{n,k})]$ más pequeño.

Para obtener una estimación de $n(\theta_k - \theta_0)'I(\theta_0)(\theta_k - \theta_0)$ nótese que como asintóticamente $2[\ell_n(\hat{\theta}_n) - \ell_n(\hat{\theta}_{n,k})] \sim \chi_{p-k,\lambda}^2$ (donde $\lambda = n(\theta_k - \theta_0)'I(\theta_0)(\theta_k - \theta_0)$), y $2[\ell_n(\hat{\theta}_n) - \ell_n(\theta_0)] \sim \chi_p^2$, entonces

$$\begin{aligned} E \left[2[\ell_n(\theta_0) - \ell_n(\hat{\theta}_{n,k})] \right] &= E \left[2[\ell_n(\hat{\theta}_n) - \ell_n(\hat{\theta}_{n,k})] - 2[\ell_n(\hat{\theta}_n) - \ell_n(\theta_0)] \right] \\ &\approx E \left[2[\ell_n(\hat{\theta}_n) - \ell_n(\hat{\theta}_{n,k})] \right] - E \left[2[\ell_n(\hat{\theta}_n) - \ell_n(\theta_0)] \right] \\ &= (p - k + \lambda) - p \\ &= n(\theta_k - \theta_0)'I(\theta_0)(\theta_k - \theta_0) - k. \end{aligned} \quad (2.19)$$

De aquí, un estimador aproximadamente insesgado para $n(\theta_k - \theta_0)'I(\theta_0)(\theta_k - \theta_0)$ está dado por

$$2[\ell_n(\theta_0) - \ell_n(\hat{\theta}_{n,k})] + k$$

y en consecuencia, por (2.18), un estimador aproximadamente insesgado para $E \left[2nK(\theta_0, \hat{\theta}_k) \right]$ es

$$\begin{aligned} \hat{E} \left[2nK(\theta_0, \hat{\theta}_k) \right] &=: 2[\ell_n(\theta_0) - \ell_n(\hat{\theta}_{n,k})] + k + k \\ &= 2[\ell_n(\theta_0) - \ell_n(\hat{\theta}_{n,k})] + 2k \end{aligned} \quad (2.20)$$

Así, por lo mencionado anteriormente, la comparación de varios modelos candidatos se puede llevar a cabo con (2.20), pero como $\ell_n(\theta_0)$ es un término común puede ignorarse, y tal comparación resulta equivalente a realizarla con

$$\text{AIC} = -2\ell(\hat{\theta}_{n,k}) + 2k \quad (2.21)$$

el llamado criterio de información de Akaike.

2.3. Criterio de información Bayesiana (BIC)

El criterio de información Bayesiana (BIC) propuesto por Schwarz en (1978), ha sido uno de los métodos más populares usado para la selección de modelos. Este es un criterio de evaluación de modelos en términos de sus probabilidades posteriores. Una motivación detrás

del BIC junto con un bosquejo de la derivación de este se presenta en seguida.

Se tiene el problema de seleccionar, dentro de un conjunto de r modelos no necesariamente anidados, el que mejor describa a un conjunto de datos $\mathbf{x}_n = (x_1, \dots, x_n)'$, donde la densidad condicional de estos dado por el i -ésimo modelo candidato (M_i) y su correspondiente vector de parámetros (θ^i), está dada por $f_i(\mathbf{x}_n|\theta^i) =: f(\mathbf{x}_n|M_i, \theta^i)$ ($\theta^i \in \Theta_i \subset \mathbf{R}^{k_i}$). Sea $\pi_i(\theta)$ la densidad a priori para el vector θ^i dado el modelo M_i , y $p(M_i)$ una densidad de probabilidad discreta a priori que asigna probabilidad positiva a cada uno de los modelos M_1, \dots, M_r . Dado estos supuestos, por el teorema de Bayes de probabilidad total, la probabilidad a posteriori del i -ésimo modelo está dada por

$$\begin{aligned}
 P(M_i|\mathbf{x}_n) &= \frac{P(M_i)f(\mathbf{x}_n|M_i)}{f(\mathbf{x}_n)} \\
 &= \frac{P(M_i) \int_{\Theta_i} f(\mathbf{x}_n|M_i, \theta)\pi_i(\theta)d\theta}{f(\mathbf{x}_n)} \\
 &= \frac{P(M_i) \int_{\Theta_i} f_i(\mathbf{x}_n|\theta)\pi_i(\theta)d\theta}{f(\mathbf{x}_n)} \\
 &= \frac{P(M_i)f_i(\mathbf{x}_n)}{f(\mathbf{x}_n)} \tag{2.22}
 \end{aligned}$$

donde

$$f_i(\mathbf{x}_n) = \int_{\Theta_i} f_i(\mathbf{x}_n|\theta)\pi_i(\theta)d\theta. \tag{2.23}$$

La probabilidad condicional $P(M|\mathbf{x}_n)$ dada en (2.22) se interpreta como la probabilidad de que los datos sean generados por el modelo M_i dado que se ha observado \mathbf{x}_n , entonces desde un punto de vista Bayesiano es natural adoptar como mejor modelo el que tenga mayor probabilidad a posteriori.

Para comparar diferentes modelos a través de sus probabilidades a posteriori, $f(\mathbf{x}_n)$ no es importante ya que es un término común para todos los modelos, y al ignorarse, tal comparación resulta equivalente a realizarla con solamente el numerador de (2.22), es decir, usando $P(M_i)f_i(\mathbf{x}_n)$. Además, si asume que las probabilidades a priori $P(M_i)$ son iguales para todos los modelos, entonces el modelo que maximice (2.23) es el que debe seleccionarse como

el mejor.

En la práctica los valores de (2.23) son difíciles de calcular, además de que para ello se requiere de la especificación de las densidades a priori $\pi_i(\theta)$. Una aproximación del logaritmo de (2.23) está dada por

$$\begin{aligned}\log[f_i(\mathbf{x}_n)] &\approx \log[f_i(\mathbf{x}_n|\hat{\theta}_n^i)] - \frac{k_i}{2}\log(n) \\ &= \ell_{n,i}(\hat{\theta}_n^i) - \frac{k_i}{2}\log(n)\end{aligned}\tag{2.24}$$

donde $\hat{\theta}_n^i$ es el EMV de θ^i y $\ell_{n,i}(\theta^i) = \log[f_i(\mathbf{x}_n|\theta^i)]$ es la log-verosimilitud correspondiente al modelo M_i (Konishi y Kitagawa, 2008; Claeskens y Hjort, 2008; Burnham y Anderson, 2002). Así, ya que la función logaritmo es monótona creciente, seleccionar como mejor modelo el que maximice (2.23) equivale aproximadamente a seleccionar al que maximice (2.24), lo que a su vez equivale a seleccionar el modelo que minimice

$$\text{BIC} =: -2\ell_{n,i}(\hat{\theta}_n^i) + k_i\log(n)\tag{2.25}$$

el llamado criterio de información Bayesiana.

3.1. Introducción

El objetivo de este trabajo de simulación es estudiar el desempeño de los métodos de selección AIC y BIC en la identificación del verdadero modelo cuando se trabaja con datos de tiempo de vida. Para el estudio se consideran las distribuciones de tiempo de vida más comunes y censura por tiempo, para diferentes fracciones de censura y diferentes tamaños de muestra.

Las distribuciones consideradas son de uno y dos parámetros, y cuando solamente se consideran distribuciones con dos parámetros, la elección de estos se hace de tal manera que todas las densidades tengan la misma media $E(T)$ y el mismo coeficiente de variación CV. Además, el estudio se hace para fracciones de censura $\gamma = 0, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50$ y tamaños de muestra $n = 20, 30, 50, 100$.

Para cada simulación, se generan 100,000 muestras de tamaño n de una de las distribuciones, la cual es tomada como la distribución verdadera. Para cada una de estas muestras por máxima verosimilitud se ajustan cada una de las distribuciones candidatas, y posteriormente se calcula el AIC y BIC. Finalmente, para cada distribución se calcula la proporción de veces

que es favorecida como mejor modelo por el AIC y BIC.

3.2. Implementación del mecanismo de censura tipo I

Para el estudio del desempeño del AIC y BIC considerando censura por tiempo con diferentes fracciones de censura (γ), el tiempo límite de observación $t_{1-\gamma}$ se determinó como el cuantil $100(1 - \gamma)\%$ de la distribución que generan los datos (distribución verdadera). Por ejemplo, si los datos son generados con la distribución Weibull, $t_{1-\gamma}$ es el valor tal que $\mathbb{P}(T \leq t_{1-\gamma}) = 1 - \gamma$, es decir, $t_{1-\gamma} = \eta [-\log(\gamma)]^{1/\beta}$. Esta forma de obtener el tiempo límite de observación no garantiza que cada muestra X_1, \dots, X_n extraída contenga exactamente $100(1 - \gamma)\%$ observaciones menores que $t_{1-\gamma}$, con lo cual solamente se tiene una aproximación del porcentaje de censura deseado γ .

3.2.1. Distribuciones de uno y dos parámetros

Las distribuciones que aquí se consideran son las distribución Weibull (WEI), Rayleigh (RAY) y Exponencial (EXP). Tres primeros escenarios a estudiar es cuando los datos son generados de la distribución $\text{WEI}(\eta = 1.108, \beta = 1.5)$, $\text{EXP}(\eta = 1) \equiv \text{WEI}(\eta = 1, \beta = 1)$ ó $\text{RAY}(\sigma = 0.7979) \equiv \text{WEI}(\eta = 1.128379, \beta = 2)$. Estas tres distribuciones elegidas comparten las característica de tener media y varianza igual a 1. La gráfica de probabilidad de tales distribuciones se muestran en la Figura 3.1.

Debido a que las tres distribuciones que en este apartado se estudian tienen un parámetro de escala, entonces del Apéndice C se sigue que tanto para el AIC como para el BIC, la elección del parámetro de escala para la distribución que genera los datos no altera los resultados de las proporciones de elección correspondientes a cada uno de los tres modelos candidatos que aquí se consideran. De este modo, solamente se pueden obtener escenarios diferentes con la distribución Weibull variando su parámetro de forma correspondiente.

Los resultados obtenidos para el primer escenario cuando los datos son generados de la distribución $\text{WEI}(\eta = 1.108, \beta = 1.5)$, para los tamaños de muestra considerados, $n = 20, 30, 50$ y 100 , respectivamente se encuentran en los Cuadros 3.1–3.4. De estos cuadros se observa que para cada tamaño de muestra (n), cuando la fracción de censura (γ) aumenta,

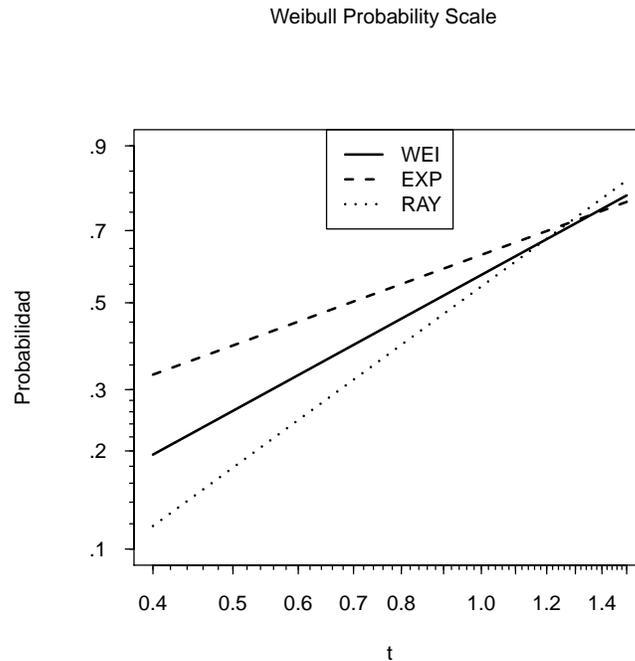


Figura 3.1. Gráficas de probabilidad de las distribuciones WEI(1.108, 1.5), EXP(1) y RAY(0.7979).

disminuye la proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es la Weibull, siendo que de esta se generaron los datos. Para todos los niveles de censura el criterio AIC tiene una probabilidad más alta de seleccionar el modelo correcto. El criterio BIC muestra un mayor grado de deterioro de la capacidad para seleccionar el modelo correcto.

Parte de la información contenida en los Cuadros 3.1–3.4 se presenta en la Figura 3.2, donde se puede apreciar de mejor manera lo dicho anteriormente. De aquí, aunque el decrecimiento de la proporción de veces, que tanto el AIC y BIC eligen al modelo correcto (en este caso el Weibull), comienza de valores cada vez más grandes a medida que el tamaño de muestra es mayor, estas se aproximan a cero conforme la fracción de censura γ se acerca a 0.5; para los tamaños de muestra 20, 30 y 50 en el BIC, y solamente para $n = 20$ y 30 en el AIC. Con $n = 50$ en el AIC y $n = 100$ en el BIC, las trayectorias de las proporciones de selección del modelo correcto aproximadamente coinciden en 0.33 al acercarse la fracción de censura a $\gamma = 0.5$, no obstante el decrecimiento de estas arranca de valores diferentes. De esta figura también se observa que con el AIC, para $n = 100$, caída de la proporción de selec-

ción del modelo correcto es mucho más lenta que cuando se consideran valores de n menores con el mismo AIC, y para todos los n con el BIC. Así, en este escenario el desempeño del AIC es superior al del BIC.

$n = 20$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.3740	0.3061	0.2508	0.1947	0.1428	0.0668	0.0296	0.0218
	Rayleigh	0.4845	0.4917	0.5109	0.5319	0.5529	0.5798	0.5900	0.5855
	Exponencial	0.1415	0.2023	0.2383	0.2735	0.3043	0.3534	0.3804	0.3927
BIC	Weibull	0.1611	0.0840	0.0398	0.0169	0.0083	0.0050	0.0068	0.0086
	Rayleigh	0.5964	0.6047	0.6131	0.6174	0.6193	0.6113	0.6044	0.5967
	Exponencial	0.2425	0.3113	0.3471	0.3658	0.3724	0.3837	0.3888	0.3947

Cuadro 3.1. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución WEI($\eta = 1.108$, $\beta = 1.5$) considerando $n = 20$.

$n = 30$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.6112	0.5548	0.4996	0.4458	0.3873	0.2715	0.1578	0.0646
	Rayleigh	0.3352	0.3504	0.3751	0.4009	0.4300	0.4833	0.5371	0.5722
	Exponencial	0.0536	0.0949	0.1253	0.1533	0.1827	0.2452	0.3051	0.3633
BIC	Weibull	0.3817	0.2939	0.2198	0.1493	0.0909	0.0206	0.0040	0.0033
	Rayleigh	0.4795	0.5017	0.5299	0.5582	0.5841	0.6070	0.6125	0.6043
	Exponencial	0.1389	0.2043	0.2502	0.2925	0.3249	0.3724	0.3836	0.3924

Cuadro 3.2. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución WEI($\eta = 1.108$, $\beta = 1.5$) considerando $n = 30$.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 50$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.8392	0.8096	0.7725	0.7321	0.6867	0.5820	0.4693	0.3298
	Rayleigh	0.1539	0.1708	0.1964	0.2250	0.2543	0.3173	0.3787	0.4498
	Exponencial	0.0070	0.0196	0.0311	0.0429	0.0590	0.1008	0.1520	0.2204
BIC	Weibull	0.6658	0.5959	0.5264	0.4560	0.3844	0.2304	0.0907	0.0143
	Rayleigh	0.2970	0.3287	0.3655	0.4056	0.4429	0.5167	0.5777	0.6052
	Exponencial	0.0372	0.0755	0.1082	0.1385	0.1727	0.2528	0.3316	0.3806

Cuadro 3.3. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución WEI($\eta = 1.108$, $\beta = 1.5$) considerando un tamaño de muestra $n = 50$.

$n = 100$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.9808	0.9756	0.9647	0.9514	0.9354	0.8908	0.8254	0.7321
	Rayleigh	0.0192	0.0241	0.0347	0.0471	0.0620	0.1011	0.1540	0.2225
	Exponencial	0.0000	0.0003	0.0006	0.0014	0.0026	0.0081	0.0206	0.0454
BIC	Weibull	0.9244	0.9017	0.8699	0.8292	0.7843	0.6727	0.5304	0.3454
	Rayleigh	0.0750	0.0947	0.1222	0.1568	0.1917	0.2731	0.3607	0.4623
	Exponencial	0.0006	0.0036	0.0080	0.0140	0.0240	0.0542	0.1089	0.1923

Cuadro 3.4. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución WEI($\eta = 1.108$, $\beta = 1.5$) considerando un tamaño de muestra $n = 100$.

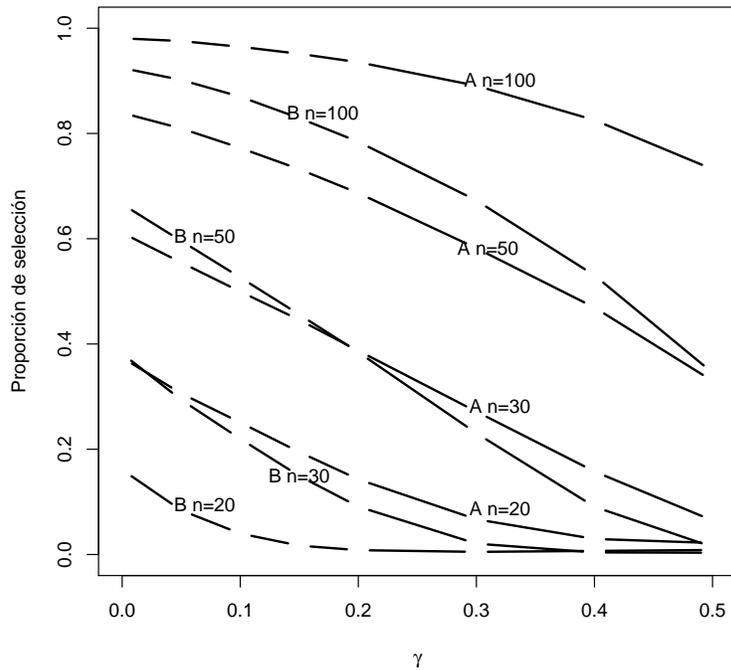


Figura 3.2. Proporción de veces que el AIC (A) y BIC (B) eligen a la distribución WEI(1.108, 1.5) como mejor modelo, siendo que de esta se generan los datos.

Para un segundo escenario los datos se generan de una distribución RAY(0.7979). Los resultados obtenidos se muestran en los Cuadros 3.5–3.8. De estos cuadros se tiene que para $n = 20$ y 30 , las proporciones de veces que el AIC elige al modelo correcto (Rayleigh) se mantienen aproximadamente constantes en 0.82, mientras que para $n = 50$ y 100 estas proporciones se mantienen aproximadamente constantes en 0.84. Así, de lo anterior se puede decir que para este caso del AIC el tamaño de muestra no tiene un efecto significativo en la proporción de selección del modelo correcto, ya que estas proporciones se mantienen muy cercas para todos los tamaños de muestra en estudio sin importar la fracción de censura considerada.

Para el BIC se tiene un comportamiento similar al descrito para el AIC, solo que las proporciones de selección del modelo correcto comienzan a “decrecer” de valores superiores para todos los tamaños de muestra. Para $n = 20$ dicha proporción decrece de 0.9018, que es cuando se tiene censura nula, a 0.8387, cuando se tiene un 50% de censura. Para $n = 30$

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

comienza en 0.9251 ($\gamma = 0$) hasta crecer a 0.9302 en $\gamma = 0.15$, y luego decrece a 0.8930 cuando se alcanza el valor de $\gamma = 0.5$. Para $n = 50$ y 100 las proporciones de selección del modelo correcto se mantienen constantes alrededor de 0.94 y 0.96 respectivamente, para todas las fracciones de censura en estudio. De esto se tiene que en este caso el desempeño de BIC es superior al del AIC, lo cual se refleja notoriamente en la Figura 3.3.

$n = 20$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.1751	0.1652	0.1556	0.1435	0.1322	0.1151	0.1056	0.1046
	Rayleigh	0.8231	0.8261	0.8272	0.8309	0.8313	0.8253	0.8142	0.7921
	Exponencial	0.0018	0.0088	0.0172	0.0256	0.0365	0.0596	0.0802	0.1033
BIC	Weibull	0.0916	0.0749	0.0673	0.0629	0.0606	0.0591	0.0574	0.0578
	Rayleigh	0.9018	0.9030	0.9000	0.8975	0.8921	0.8769	0.8613	0.8387
	Exponencial	0.0066	0.0222	0.0326	0.0397	0.0473	0.0640	0.0813	0.1035

Cuadro 3.5. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución RAY($\sigma = 0.7979$) considerando $n = 20$.

$n = 30$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.1692	0.1690	0.1618	0.1599	0.1564	0.1462	0.1269	0.1102
	Rayleigh	0.8307	0.8304	0.8366	0.8369	0.8374	0.8371	0.8364	0.8294
	Exponencial	0.0001	0.0006	0.0016	0.0032	0.0062	0.0168	0.0367	0.0604
BIC	Weibull	0.0743	0.0695	0.0622	0.0566	0.0514	0.0445	0.0428	0.0412
	Rayleigh	0.9251	0.9268	0.9296	0.9302	0.9286	0.9211	0.9078	0.8930
	Exponencial	0.0006	0.0037	0.0081	0.0132	0.0200	0.0344	0.0494	0.0658

Cuadro 3.6. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución RAY($\sigma = 0.7979$) considerando $n = 30$.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 50$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.1647	0.1641	0.1633	0.1628	0.1605	0.1593	0.1560	0.1471
	Rayleigh	0.8353	0.8359	0.8367	0.8372	0.8393	0.8399	0.8407	0.8418
	Exponencial	0.0000	0.0000	0.0000	0.0001	0.0002	0.0009	0.0033	0.0112
BIC	Weibull	0.0523	0.0518	0.0511	0.0503	0.0490	0.0446	0.0366	0.0294
	Rayleigh	0.9478	0.9482	0.9487	0.9490	0.9494	0.9493	0.9475	0.9415
	Exponencial	0.0000	0.0001	0.0002	0.0007	0.0017	0.0062	0.0159	0.0291

Cuadro 3.7. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución RAY($\sigma = 0.7979$) considerando un tamaño de muestra $n = 50$.

$n = 100$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.1577	0.1600	0.1591	0.1592	0.1604	0.1607	0.1587	0.1575
	Rayleigh	0.8423	0.8400	0.8409	0.8408	0.8396	0.8394	0.8413	0.8425
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
BIC	Weibull	0.0313	0.0336	0.0326	0.0331	0.0335	0.0326	0.0320	0.0304
	Rayleigh	0.9687	0.9664	0.9674	0.9669	0.9665	0.9674	0.9680	0.9682
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0014

Cuadro 3.8. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución RAY($\sigma = 0.7979$) considerando un tamaño de muestra $n = 100$.

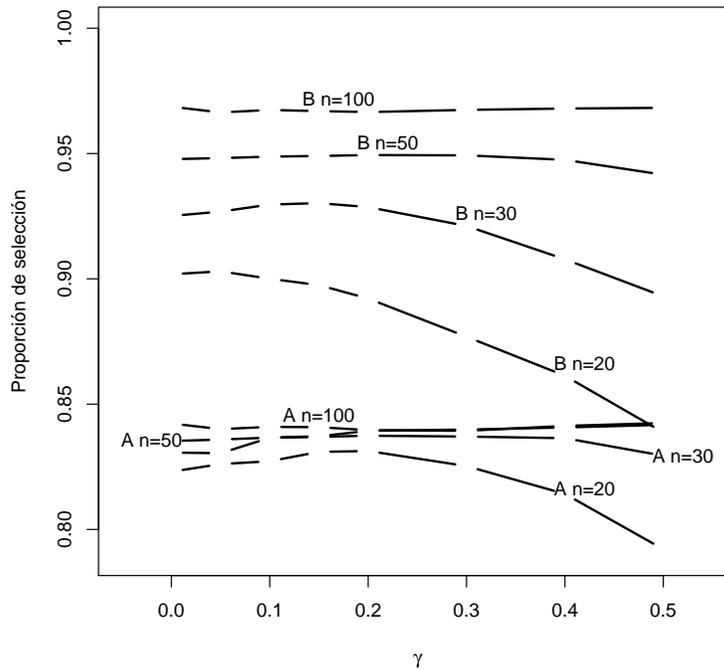


Figura 3.3. Proporción de veces que el AIC (A) y BIC (B) eligen a la distribución RAY($\sigma = 0.7979$) como mejor modelo, siendo que de esta se generan los datos.

Para un tercer escenario los datos se generan de una distribución EXP(1). Los resultados obtenidos se presentan en los Cuadros 3.9–3.12, donde se puede apreciar que estos no difieren mucho a los obtenidos en el caso Rayleigh, solo que aquí la distribución exponencial es la mayormente favorecida. La Figura 3.4 muestra, para las diferentes combinaciones de n y γ consideradas en el estudio, la proporción de las veces que el AIC y BIC eligen como mejor al modelo exponencial. De la figura o de los cuadros correspondientes, se observa que prácticamente para el AIC los resultados son análogos a los obtenidos en el segundo escenario, mientras que para el BIC en $n = 20$ es donde hay una pequeña diferencia; en este tercer escenario se tienen un decrecimiento ligeramente más rápido de la proporción de selección del modelo correcto. En este escenario también el BIC es superior al AIC, aunque hay que remarcar que el desempeño de este último criterio podría considerarse que no es malo, debido a que para todas las fracciones de censura aquí especificadas la proporción de selección del modelo correcto es mayor que 0.80.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 20$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.1551	0.1531	0.1457	0.1338	0.1194	0.0933	0.0711	0.0646
	Rayleigh	0.0214	0.0222	0.0263	0.0351	0.0473	0.0749	0.1126	0.1457
	Exponencial	0.8235	0.8247	0.8279	0.8311	0.8333	0.8319	0.8164	0.7898
BIC	Weibull	0.0620	0.0601	0.0516	0.0422	0.0349	0.0326	0.0314	0.0321
	Rayleigh	0.0359	0.0373	0.0445	0.0550	0.0668	0.0864	0.1171	0.1472
	Exponencial	0.9021	0.9027	0.9039	0.9028	0.8983	0.8810	0.8515	0.8207

Cuadro 3.9. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución EXP(1) considerando un tamaño de muestra $n = 20$.

$n = 30$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.1681	0.1688	0.1632	0.1591	0.1555	0.1426	0.1169	0.0881
	Rayleigh	0.0020	0.0023	0.0030	0.0053	0.0086	0.0217	0.0453	0.0810
	Exponencial	0.8299	0.8289	0.8338	0.8356	0.8359	0.8357	0.8378	0.8309
BIC	Weibull	0.0681	0.0684	0.0621	0.0565	0.0495	0.0345	0.0267	0.0245
	Rayleigh	0.0053	0.0063	0.0087	0.0131	0.0201	0.0404	0.0633	0.0905
	Exponencial	0.9266	0.9253	0.9292	0.9304	0.9303	0.9251	0.9100	0.8850

Cuadro 3.10. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución EXP(1) considerando un tamaño de muestra $n = 30$.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 50$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.1640	0.1646	0.1634	0.1608	0.1593	0.1583	0.1535	0.1431
	Rayleigh	0.0000	0.0001	0.0000	0.0001	0.0002	0.0013	0.0052	0.0163
	Exponencial	0.8360	0.8354	0.8365	0.8391	0.8406	0.8404	0.8414	0.8406
BIC	Weibull	0.0512	0.0510	0.0508	0.0494	0.0493	0.0437	0.0341	0.0230
	Rayleigh	0.0001	0.0001	0.0002	0.0007	0.0011	0.0052	0.0161	0.0350
	Exponencial	0.9487	0.9489	0.9490	0.9500	0.9495	0.9511	0.9498	0.9420

Cuadro 3.11. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución EXP(1) considerando un tamaño de muestra $n = 50$.

$n = 100$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	Weibull	0.1620	0.1601	0.1621	0.1610	0.1570	0.1583	0.1613	0.1593
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002
	Exponencial	0.8380	0.8399	0.8379	0.8390	0.8430	0.8417	0.8387	0.8405
BIC	Weibull	0.0335	0.0337	0.0328	0.0334	0.0322	0.0318	0.0334	0.0317
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0016
	Exponencial	0.9665	0.9663	0.9672	0.9666	0.9678	0.9681	0.9664	0.9667

Cuadro 3.12. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación (Weibull, Rayleigh y Exponencial), siendo que los datos son generados de una distribución EXP(1) considerando un tamaño de muestra $n = 100$.

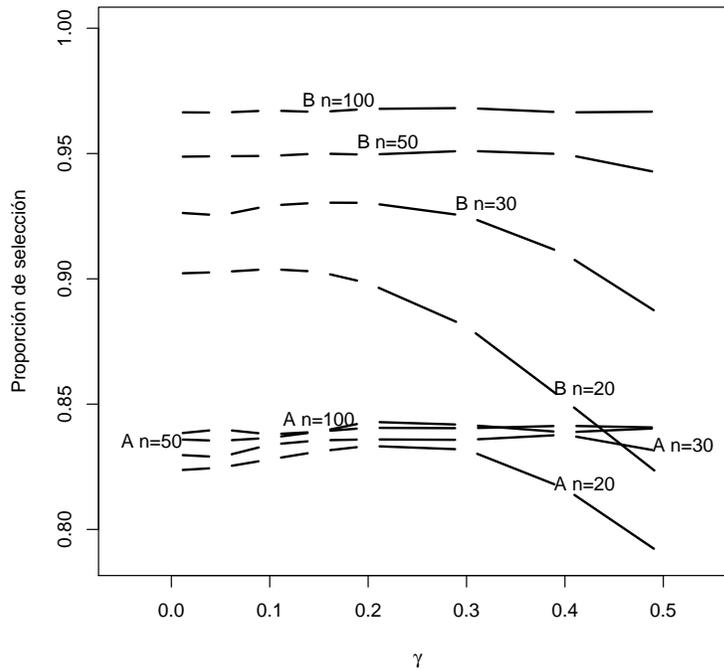


Figura 3.4. Proporción de veces que el AIC (A) y BIC (B) eligen a la distribución EXP(1) como mejor modelo, siendo que de esta se generan los datos.

En esta sección ya no se presentan más tablas de resultados, ya que por un lado se mencionó que los resultados son invariantes a la elección del parámetro de escala con el cual se generen los datos, y entonces únicamente se tienen resultados más representativos variando el parámetro de forma de la distribución Weibull. Por otro, al variar el parámetro de forma de la distribución Weibull el comportamiento de los resultados para valores de β al rededor de 1.5 es muy similar al caso ya discutido. Con valores de β bastante cercanos a 1, ya sea por arriba o por abajo, para tamaños de muestra pequeños y todas las fracciones de censura consideradas, la distribución exponencial es la más elegida y la segunda la Weibull pero en las fracciones de censura más pequeñas. Conforme el tamaño de muestra crece, la preferencia sobre el modelo Weibull aumenta, siendo más notable con el AIC, además que para que el modelo correcto sea el más elegido, entre más cerca se encuentre β de 1, más grande se necesita el tamaño de muestra, sin dejar de observar el comportamiento esperado que a medida que la fracción de censura crece, la proporción de selección del modelo correcto decrece. Cuando el parámetro β se varia en vecindades pequeñas alrededor de 2 se observan resul-

tados análogos (pero en vez de la exponencial la distribución Rayleigh es la más elegida), mientras que a medida que β se encuentra cada vez más lejos por debajo de 1 o por arriba de 2, la precisión tanto del AIC como del BIC va siendo mucho mejor y superior a los casos expuestos anteriormente.

3.2.2. Distribuciones de dos parámetros

Otra situación de interés es cuando se requiere seleccionar el mejor modelo que describa a un conjunto de datos, pero dentro de un conjunto de modelos con solamente dos parámetros. Para ello las distribuciones que aquí se consideran son: Weibull, Lognormal, Loglogística e Inversa Gausiana. Notemos que en estas circunstancias el AIC y BIC siempre eligen a un mismo modelo como el mejor. Esto se debe a que como los 4 modelos tienen dos parámetros, el segundo sumando de (2.21) y (2.25), respectivamente, es común para todos ellos, y así ambos criterios (AIC y BIC) se reducen a elegir como mejor modelo el que tenga menor $-\ell(\hat{\theta}_n)$, o equivalentemente el que tenga mayor log-verosimilitud, $\ell(\hat{\theta}_n)$.

Del Apéndice A se concluye que para cada una de las 4 distribuciones aquí consideradas, el CV queda completamente determinado por su correspondiente parámetro de forma. Así, para un coeficiente de variación dado el parámetro de forma está fijo, y en consecuencia la dependencia de la esperanza $E(T)$ solamente queda del parámetro de escala (ver Apéndice A). Pero como en el Apéndice C se muestra que para muestras censuradas por tiempo, el AIC y BIC son independientes del parámetro de escala de la distribución de la que se genera los datos, entonces para un CV fijo no es necesario variar $E(T)$, y por lo tanto el estudio se hace eligiendo diferentes valores de CV, lo que nos lleva tener resultados más generales.

Para un primer escenario, la elección de los parámetros se hizo de tal forma que todas las distribuciones compartieran una media $E(T) = 50$ y un coeficiente de variación $CV = 0.90$. Las gráficas de probabilidad de las distribuciones con tales especificaciones se muestran en la Figura 3.5.

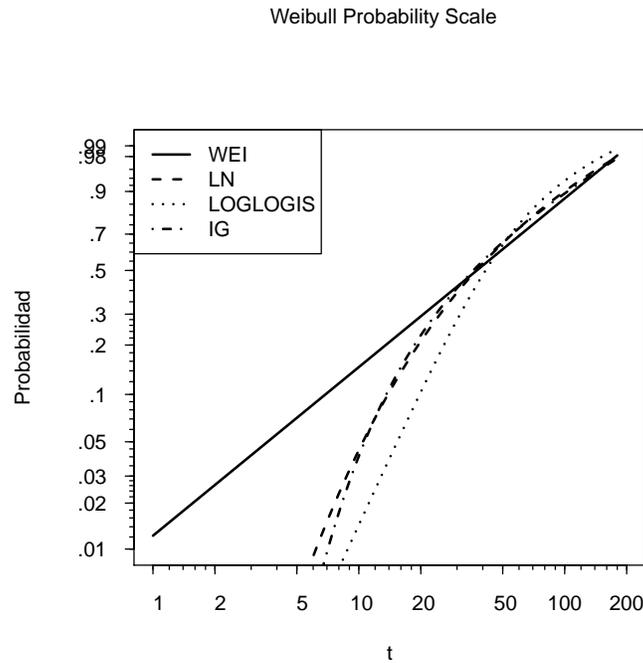


Figura 3.5. Gráficas de probabilidad Weibull, Lognormal, Loglogística e Inversa Gausiana con media $E(T) = 50$ y coeficiente de variación $CV = 0.90$

Al generar muestras de una distribución Weibull, con las especificaciones anteriores, se obtienen los resultados que se muestran en los Cuadros 3.13–3.16. De estos se observa que para cada tamaño de muestra (n), conforme el porcentaje de censura aumenta (γ) el porcentaje de veces que el AIC y BIC eligen como mejor modelo al Weibull va decreciendo, siendo este del cual se generaron los datos (el verdadero modelo). Además notemos que en los 4 casos tal decremento es muy parecido, con la diferencia de que para los tamaños de muestra más grande tal disminución comienza de valores más grandes, por ejemplo, para $n = 100$ comienza de 0.9661 hasta llegar a 0.5636, mientras que para $n = 50$ comienza de 0.8897 y decrece a 0.4872, que es cuando se tiene un 50 % de censura. Así, la precisión del AIC y BIC en esta situación se ve bastante afectada por la fracción de censura presente en la muestra, pero a medida que n crece esta mejora. Lo anterior gráficamente también se puede apreciar en la Figura 3.6.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 20$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.7489	0.6471	0.5980	0.5623	0.5273	0.4717	0.4224	0.3741
Lognormal	0.0708	0.0988	0.1178	0.1317	0.1490	0.1744	0.2036	0.2433
Loglogística	0.0387	0.1014	0.1146	0.1196	0.1169	0.1089	0.0937	0.0734
Inversa Gausiana	0.1416	0.1527	0.1696	0.1864	0.2068	0.2450	0.2803	0.3091

Cuadro 3.13. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación, siendo que los datos son generados de una distribución Weibull con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 20$.

$n = 30$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.8143	0.7161	0.6656	0.6246	0.5858	0.5273	0.4730	0.4244
Lognormal	0.0771	0.1071	0.1262	0.1444	0.1585	0.1941	0.2258	0.2647
Loglogística	0.0363	0.0993	0.1206	0.1312	0.1409	0.1355	0.1253	0.1051
Inversa Gausiana	0.0723	0.0776	0.0876	0.0998	0.1148	0.1431	0.1759	0.2058

Cuadro 3.14. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución Weibull con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 30$.

$n = 50$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.8897	0.7983	0.7500	0.7070	0.6685	0.5999	0.5369	0.4872
Lognormal	0.0632	0.0918	0.1070	0.1231	0.1411	0.1795	0.2215	0.2683
Loglogística	0.0261	0.0887	0.1181	0.1391	0.1539	0.1673	0.1656	0.1459
Inversa Gausiana	0.0210	0.0213	0.0248	0.0308	0.0365	0.0534	0.0761	0.0986

Cuadro 3.15. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución Weibull con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 50$.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 100$ Distribución	Fracción de censura γ							
	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.9661	0.9062	0.8640	0.8226	0.7789	0.7006	0.6277	0.5636
Lognormal	0.0231	0.0377	0.0475	0.0581	0.0715	0.1051	0.1525	0.2118
Loglogística	0.0097	0.0551	0.0873	0.1175	0.1467	0.1888	0.2084	0.2053
Inversa Gausiana	0.0011	0.0010	0.0012	0.0018	0.0029	0.0055	0.0114	0.0193

Cuadro 3.16. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución Weibull con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 100$.

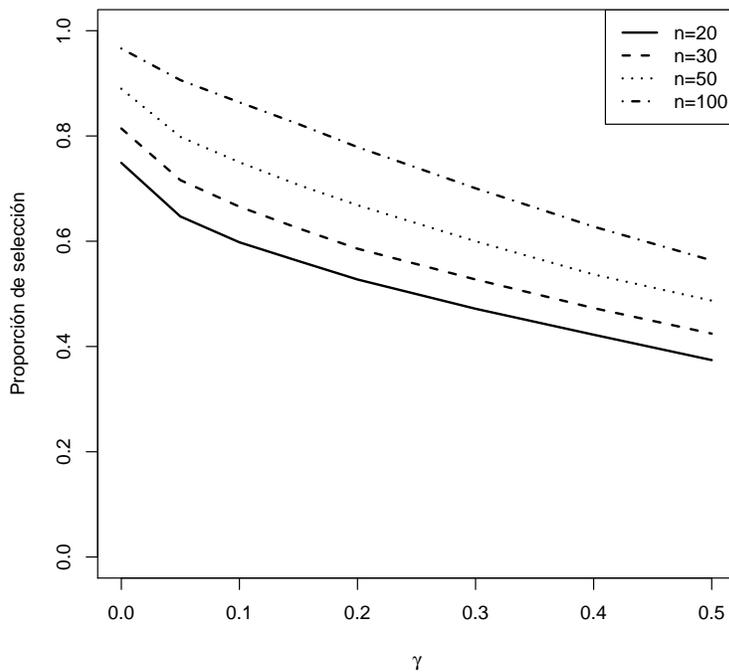


Figura 3.6. Proporción de veces que el AIC y BIC eligen al modelo correcto, siendo que los datos se generan de una distribución Weibull con media $E(T) = 50$ y coeficiente de variación $CV = 0.90$.

Para cuando la distribución Lognormal con media $E(T) = 50$ y coeficiente de variación $CV = 0.90$ es la que genera los datos, el comportamiento de los resultados difieren al obtenido en el caso Weibull, esto se puede observar en los Cuadros 3.17-3.20. Únicamente para $n = 100$ y $\gamma = 0$ la distribución Lognormal resulta la más adecuada, y para las restantes combinaciones de n y γ el AIC y BIC indican a la distribución inversa gaussiana como la mejor. Cuando solamente se considera $n = 20$ (Cuadro 3.17), excepto en $\gamma = 0.5$, el más

peor de los cuatro en competencia es el modelo Lognormal, manteniéndose la proporción de selección correspondiente aproximadamente constante en 0.09. Para $n = 30, 50$ y 100 los criterios ya no señalan al Lognormal como el peor, salvo cuando no hay censura y $n = 30$. Para los 3 últimos tamaños de muestra se tiene que el porcentaje de selección del modelo correcto decrece ligeramente conforme la fracción de censura aumenta, no obstante el mayor cambio se observa de $\gamma = 0$ a $\gamma = 0.05$. Ver Figura 3.20. Además, si bien conforme el tamaño de muestra aumenta la proporción de selección del modelo correcto crece, tal proporción es relativamente baja (< 0.5) aún cuando se alcanza $n = 100$ con muestras sin censura. Así, en esta situación el desempeño del AIC y BIC se ve afectado con la presencia de censura al menos con los tamaños de muestra aquí consideradas.

$n = 20$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.2061	0.1814	0.1659	0.1614	0.1629	0.1619	0.1665	0.1725
Lognormal	0.0902	0.0882	0.0922	0.0942	0.0954	0.0955	0.0915	0.0839
Loglogística	0.1470	0.1824	0.1677	0.1530	0.1353	0.1133	0.0915	0.0749
Inversa Gausiana	0.5568	0.5481	0.5742	0.5914	0.6064	0.6293	0.6505	0.6688

Cuadro 3.17. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación, siendo que los datos son generados de una distribución Lognormal con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 20$.

$n = 30$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.1524	0.1342	0.1291	0.1303	0.1328	0.1417	0.1515	0.1609
Lognormal	0.1505	0.1390	0.1428	0.1427	0.1459	0.1478	0.1368	0.1291
Loglogística	0.1688	0.2000	0.1846	0.1667	0.1509	0.1261	0.1048	0.0828
Inversa Gausiana	0.5284	0.5268	0.5435	0.5604	0.5704	0.5843	0.6068	0.6272

Cuadro 3.18. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución Lognormal con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 30$.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 50$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.0882	0.0802	0.0829	0.0875	0.0925	0.1038	0.1167	0.1338
Lognormal	0.2492	0.2272	0.2277	0.2276	0.2262	0.2268	0.2165	0.1955
Loglogística	0.1829	0.2112	0.2005	0.1849	0.1731	0.1484	0.1266	0.1080
Inversa Gausiana	0.4797	0.4815	0.4889	0.5000	0.5082	0.5210	0.5402	0.5627

Cuadro 3.19. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución Lognormal con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 50$.

$n = 100$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.0252	0.0268	0.0307	0.0354	0.0415	0.0525	0.0660	0.0841
Lognormal	0.4138	0.3764	0.3706	0.3663	0.3657	0.3593	0.3437	0.3153
Loglogística	0.1657	0.2020	0.1977	0.1880	0.1814	0.1640	0.1479	0.1351
Inversa Gausiana	0.3952	0.3948	0.4010	0.4103	0.4115	0.4243	0.4424	0.4655

Cuadro 3.20. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución Lognormal con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 100$.

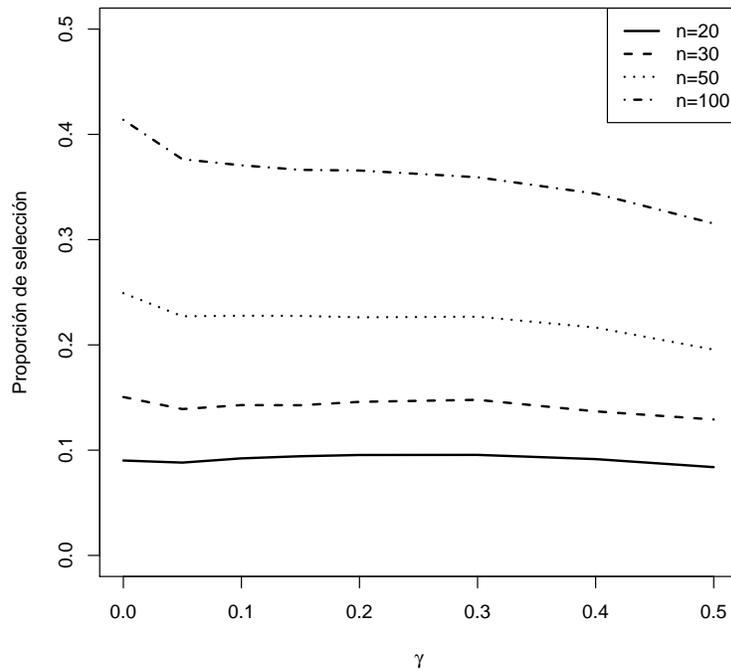


Figura 3.7. Proporción de veces que el AIC y BIC eligen al modelo correcto, siendo que los datos se generan de una distribución Lognormal con media $E(T) = 50$ y coeficiente de variación $CV = 0.90$.

Para seguir con el estudio ahora se asume que la distribución que genera los datos es Loglogística con media $E(T) = 50$ y $CV = 0.90$. Los resultados para esta situación se muestran en los Cuadros 3.21–3.24. El primer comportamiento general que se observa de tales cuadros es que conforme la fracción censura aumenta, la proporción de elegir el modelo correcto decrece, tal como lo anticipa el análisis teórico. Además, se puede observar que a medida que el tamaño de muestra se hace más grande tal decrecimiento comienza de valores cada vez superiores, aunque también es cada vez más rápido; para $n = 20$ comienza en 0.2942 hasta llegar a 0.1660 en $\gamma = 0.5$, para $n = 30$ de 0.3847 a 0.1561, para $n = 50$ de 0.5052 a 0.1818, y para $n = 100$ de 0.6633 a 0.2775. Adicionalmente del Cuadro 3.21 se tienen que para todos los valores de γ la distribución loglogística no es la mayormente elegida. Para $n = 30$ (Cuadro 3.21) en $\gamma = 0$ y 0.05 el modelo loglogístico es el más elegido, para $n = 50$ en $\gamma = 0$, 0.05, 0.10, 0.15 y 0.20, y para $n = 100$ en todos los valores de γ aquí considerados (Cuadro 3.24) es mayor la proporción que se selecciona el modelo correcto, salvo cuando la fracción de censura es del 50 % donde la distribución Weibull es la más elegida (0.3044). Obsérvese también que la mayor proporción de selección del modelo correcto es menor a 0.70, la cual se encuentra en $n = 100$ y $\gamma = 0$ y representa la combinación, de las aquí consideradas,

con la que se cuenta con mayor información. De esta manera, se tiene que este escenario el desempeño del AIC y BIC no es bueno aún con tamaños de muestra más grandes los cuales frecuentemente son usados en la práctica (50, 100) y pequeña fracción de censura, aunado que a medida que esta última crece, afecta significativamente la capacidad de selección del verdadero modelo.

El comportamiento de las proporciones de selección del modelo correcto, con mayor detalle se puede observar gráficamente en la Figura 3.8.

$n = 20$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.2233	0.2113	0.2153	0.2282	0.2387	0.2610	0.2676	0.2648
Lognormal	0.0595	0.0549	0.0586	0.0623	0.0617	0.0645	0.0631	0.0596
Loglogística	0.2942	0.3156	0.2807	0.2546	0.2367	0.2081	0.1866	0.1660
Inversa Gausiana	0.4230	0.4182	0.4454	0.4549	0.4629	0.4664	0.4827	0.5097

Cuadro 3.21. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación, siendo que los datos son generados de una distribución Loglogística con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 20$.

$n = 30$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.1636	0.1607	0.1818	0.2041	0.2276	0.2603	0.2799	0.2917
Lognormal	0.0961	0.0898	0.0912	0.0949	0.0970	0.1023	0.0991	0.0946
Loglogística	0.3847	0.3753	0.3266	0.2913	0.2596	0.2205	0.1894	0.1561
Inversa Gausiana	0.3556	0.3742	0.4004	0.4097	0.4157	0.4169	0.4317	0.4576

Cuadro 3.22. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución Loglogística con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 30$.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 50$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.0944	0.1036	0.1368	0.1701	0.2010	0.2503	0.2815	0.3070
Lognormal	0.1449	0.1328	0.1352	0.1439	0.1469	0.1520	0.1507	0.1441
Loglogística	0.5052	0.4640	0.4082	0.3662	0.3303	0.2717	0.2267	0.1818
Inversa Gausiana	0.2555	0.2996	0.3198	0.3199	0.3218	0.3260	0.3411	0.3671

Cuadro 3.23. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución Loglogística con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 50$.

$n = 100$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.0289	0.0417	0.0745	0.1095	0.1433	0.2082	0.2625	0.3044
Lognormal	0.1858	0.1771	0.1888	0.1978	0.2029	0.2066	0.2059	0.2046
Loglogística	0.6633	0.6081	0.5542	0.5103	0.4739	0.4027	0.3391	0.2775
Inversa Gausiana	0.1220	0.1732	0.1825	0.1824	0.1798	0.1825	0.1925	0.2135

Cuadro 3.24. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución Loglogística con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 100$.

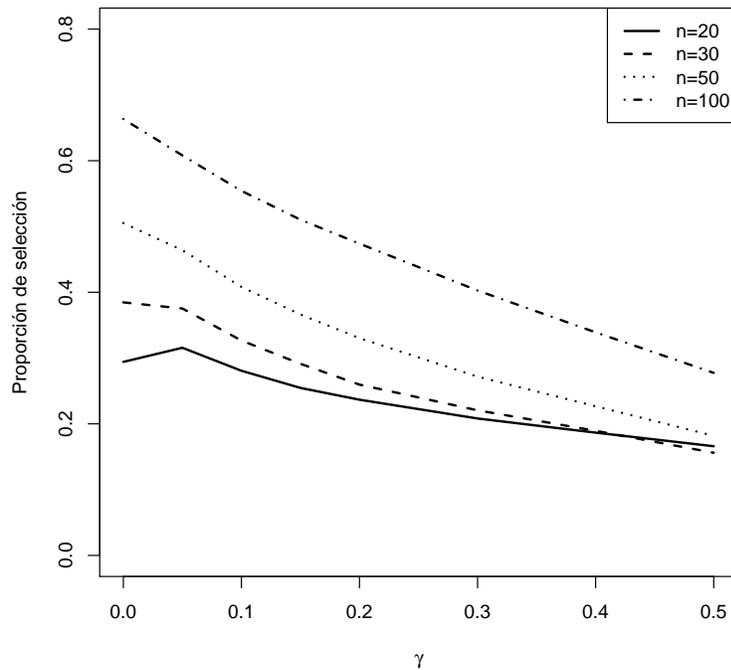


Figura 3.8. Proporción de veces que el AIC y BIC eligen al modelo correcto, siendo que los datos se generan de una distribución Loglogística con media $E(T) = 50$ y coeficiente de variación $CV = 0.90$.

Generando los datos de la distribución inversa gaussiana con media $E(T) = 50$ y $CV = 0.90$ se obtienen los resultados que se muestran en los Cuadros 3.25–3.28. Aquí el comportamiento de tales resultados es bastante diferente a lo obtenido cuando los datos son generados de la distribución Weibull, Lognormal ó Loglogística con las mismas especificaciones. Se observa que para todos los valores de n y γ , la distribución gaussiana inversa es la más elegida. Para cada uno de los primeros tres tamaños de muestra que se consideran ($n = 20, 30$ y 50), el comportamiento esperado no se aprecia, debido a que a partir de $\gamma = 0.05$ la proporción de selección del modelo correcto comienza a crecer en vez de decrecer (es decir, según ambos criterios, AIC y BIC, el modelo verdadero mejora entre más censura se tenga), además de que el valor al que crece es superior para los tamaños de muestra más pequeños. Lo que si se observa es que para cada γ inferior a 0.3, el AIC y BIC son mejores entre más grande sea el tamaño de muestra. Para $n = 100$ se observa un comportamiento más inclinado a lo intuitivamente esperado, ya que a medida que la fracción censura crece, la cantidad de interés tiende, aunque ligeramente, a decrecer. Así, a pesar de que el comportamiento de los resultados no es del todo esperado, en esta situación el desempeño del AIC y BIC no se podría juzgar de forma negativa.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 20$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.1378	0.1192	0.1034	0.0990	0.0948	0.0978	0.1039	0.1132
Lognormal	0.0709	0.0683	0.0753	0.0742	0.0758	0.0753	0.0713	0.0633
Loglogística	0.1052	0.1385	0.1231	0.1071	0.0951	0.0776	0.0652	0.0583
Inversa Gaussiana	0.6862	0.6740	0.6982	0.7198	0.7343	0.7493	0.7597	0.7652

Cuadro 3.25. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las tres en comparación, siendo que los datos son generados de una distribución inversa gaussiana con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 20$.

$n = 30$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.0830	0.0722	0.0643	0.0636	0.0654	0.0701	0.0801	0.0936
Lognormal	0.1101	0.1045	0.1095	0.1094	0.1100	0.1084	0.1053	0.0962
Loglogística	0.1008	0.1250	0.1112	0.0956	0.0849	0.0707	0.0604	0.0526
Inversa Gaussiana	0.7062	0.6983	0.7150	0.7314	0.7398	0.7509	0.7542	0.7576

Cuadro 3.26. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución inversa gaussiana con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 30$.

$n = 50$	Fracción de censura γ							
Distribución	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.0354	0.0290	0.0278	0.0291	0.0314	0.0371	0.0461	0.0592
Lognormal	0.1523	0.1484	0.1491	0.1542	0.1561	0.1578	0.1514	0.1402
Loglogística	0.0779	0.0987	0.0855	0.0740	0.0659	0.0571	0.0516	0.0502
Inversa Gaussiana	0.7345	0.7239	0.7377	0.7428	0.7466	0.7480	0.7509	0.7504

Cuadro 3.27. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución inversa gaussiana con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 50$.

$n = 100$ Distribución	Fracción de censura γ							
	0	0.05	0.10	0.15	0.20	0.30	0.40	0.50
Weibull	0.0043	0.0039	0.0040	0.0045	0.0058	0.0090	0.0130	0.0217
Lognormal	0.1872	0.1797	0.1862	0.1881	0.1933	0.1990	0.1946	0.1917
Loglogística	0.0382	0.0509	0.0417	0.0340	0.0314	0.0279	0.0305	0.0351
Inversa Gausiana	0.7703	0.7656	0.7681	0.7735	0.7694	0.7642	0.7620	0.7514

Cuadro 3.28. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 4 en comparación, siendo que los datos son generados de una distribución inversa gaussiana con media $E(T) = 50$ y $CV = 0.90$, considerando un tamaño de muestra $n = 100$.

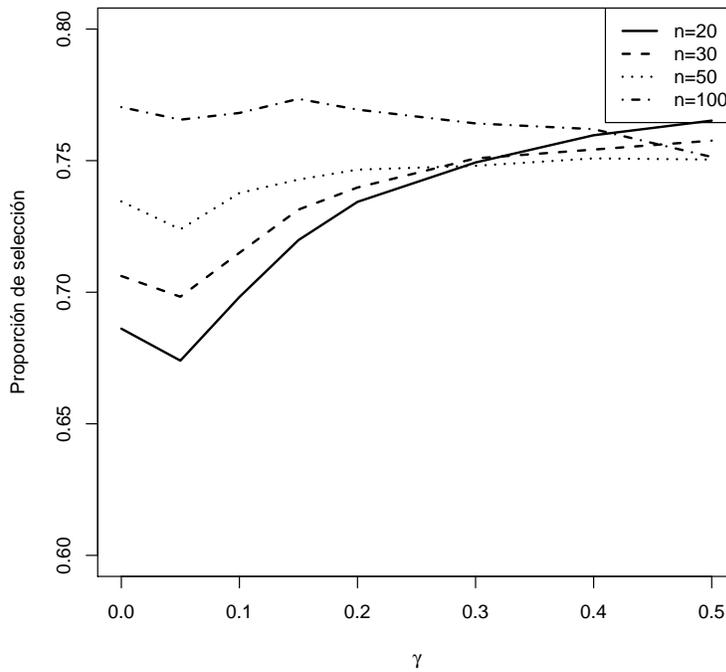


Figura 3.9. Proporción de veces que el AIC y BIC eligen al modelo correcto, siendo que los datos se generan de una distribución inversa gaussiana con media $E(T) = 50$ y coeficiente de variación $CV = 0.90$.

Para verificar la consistencia del AIC en las proporciones de selección de cada uno de los modelos que en esta sección se consideran, el Cuadro 3.29 muestra la mínima distancia K-L del modelo candidato $f(x|\theta)$ al modelo que genera los datos $g(x)$ (verdadero modelo). La mínima distancia K-L entre $f(x|\theta)$ y $g(x)$ esta dada por $I[g(\cdot), f(\cdot|\theta_0)]$, donde $\theta_0 = \arg \min_{\theta \in \Theta} I[g(\cdot), f(\cdot|\theta)]$. Como de (2.4) se tiene que el estimador de máxima verosimili-

tud de θ del modelo $f(x|\theta)$, $\hat{\theta}_n$, converge en probabilidad a θ_0 , entonces se espera que para n “grande”, la jerarquía del modelo más adecuado (de acuerdo a la distancia K-L) señalado en el Cuadro 3.29 se vea reflejado en los cuadros de resultados con tamaños de muestra mayores, al menos para cuando la fracción de censura es nula. Así, para cuando los datos son generados de la distribución WEI(52.0134, 1.1128), de tal cuadro se tiene que el modelo más adecuado es el Weibull, seguido por el loglogístico, lognormal e inverso gaussiano, lo cual hace consistente a lo contenido en el Cuadro 3.16, aunque para $\gamma = 0$ no exactamente la distribución loglogística sea la segunda más elegida. Esto es razonable ya que la distancia de este modelo y del modelo lognormal al verdadero modelo, respectivamente, están muy cercanas, 0.0632 y 0.0905. En $\gamma = 0.05, 0.10, 0.15, 0.20, 0.30$ y 0.40 de (3.16) el orden de jerarquía se mantiene, mientras que en $\gamma = 0.50$ pasa lo mismo que para cuando se tiene nula fracción censura. Algo similar sucede para $n = 50$.

Para cuando el modelo verdadero $g(x)$ es LOGNOR(3.6154, 0.7703) la jerarquía del mejor modelo de acuerdo a la distancia de K-L es la siguiente: lognormal, inversa gaussiana, loglogística y Weibull. Esta misma jerarquía se tiene para $n = 100$ (Cuadro 3.20), excepto que para los valores de γ diferentes de 0 el orden de los modelos lognormal e inverso gaussiano intercambia, resultando este último el más elegido. Algo similar ocurre para $n = 50$.

Considerando el modelo LOGLOGIS(3.7197, 0.3353) es el que genera los datos, del Cuadro 3.24 se observa que para $n = 100$ y $\gamma = 0$ la proporción de selección del modelo correcto es la más grande, luego la correspondiente al modelo lognormal, después la del inverso gaussiano y por último la de la distribución weibull. Esta jerarquía es la marcada por la distancia de K-L que se muestra en el tercer fila del Cuadro 3.29. Además, este mismo orden de preferencia se tiene para el resto de las γ 's, salvo para $\gamma = 0.50$ donde la posición del modelo lognormal e inverso gaussiano cambian y la de los dos restantes se mantienen.

Si la distribución IG(50, 1.2346) es la verdadera, comportamiento similares se obtienen a los observados en el caso Weibull, pero adicionalmente en esta situación la jerarquía que sugiere la distancia de K-L del Cuadro 3.29 también se aprecia para $n = 50$, salvo en $\gamma = 0.5$, donde la posición de los modelos Weibull y loglogístico se intercambian.

Modelo verdadero $g(x)$	Modelo candidato $f(x \theta)$			
	WEI	LOGNOR	LOGLOGIS	IG
WEI(52.0134, 1.1128)	0	0.0905	0.0632	0.5559
LOGNOR(3.6154, 0.7703)	0.0807	0	0.0095	0.0073
LOGLOGIS(3.7197, 0.3353)	0.1342	0.0138	0	0.0353
IG(50, 1.2346)	0.0944	0.0051	0.0223	0

Cuadro 3.29. Mínima distancia K-L entre $f(x|\theta)$ y $g(x)$, donde el modelo verdadero $g(x)$ tiene $E(T) = 50$ y $CV = 0.9$.

Más escenarios se estudiaron considerando coeficientes de variación $CV = 0.5, 2$ y 3 , con los cuales se obtuvo un panorama más general del estudio, ya que como se mencionó, los resultados solamente dependen del parámetro de forma de la distribución (ó coeficiente de variación) que genere los datos y no del de escala que se elija. Para cuando se generaron muestras de distribuciones con un coeficiente de variación menor al que ya se discutió, $CV = .5$, los resultados de selección de modelo correcto son bastante similares, manteniéndose aproximadamente el mismo problema de selección del AIC y BIC cuando el modelo Lognormal es el que genera los datos. Una posible explicación a esto es que la distribución lognormal y la inversa gaussiana son muy similares si el coeficiente de variación es menor que 1 (Takagi *et al.*, 1997), además de que numéricamente la mínima distancia de K-L de uno hacia el otro es relativamente “pequeña” y se alcanza en los mismo parámetros que hacen que ambos modelos tengan misma media y mismo coeficiente de variación. Por ejemplo, si $E(T) = 50$ y $CV = 0.5$ la mínima distancia K-L de la familia gaussiana inversa hacia el modelo lognormal con tales especificaciones es 0.001036 y se alcanza en $\theta_0 = 50$ y $\beta_0 = 4$, lo cuales también corresponden a la distribución gaussiana inversa con media 50 y coeficiente de variación 0.5. Resultados similares se obtienen si considera a la distribución gaussiana inversa con misma media y mismo coeficiente de variación como la verdadera, solo que ahora la mínima distancia K-L de la familia lognormal hacia este es 0.000869 y el correspondiente parámetro es $\mu_0 = 3.7997$ y $\sigma_0 = 0.4731$, los cuales también hacen que el modelo lognormal posea $E(T) = 50$ y $CV = 0.5$.

Para cuando se generan muestras de distribuciones con un coeficiente de variación mayor que 1 se obtienen resultados que difieren al caso anterior. A medida que el coeficiente de variación se aleja de 1, el problema de selección presenciado en la situación anterior cuando

los datos son generados de la distribución lognormal va desapareciendo, algo esperado por lo discutido en el párrafo anterior. Por la misma razón, cuando la distribución gaussiana inversa es de la que se obtienen las muestras, la eficiencia del AIC y BIC también tiende a mejorar. Con el modelo loglogístico el comportamiento de los resultados se mantienen prácticamente iguales al variar el CV, mientras que para el caso Weibull si cambian, obteniendo proporciones de selección de modelo correcto ligeramente más bajas que cuando se tomo un $CV < 1$. Por ejemplo, con $E(T) = 50$ y $CV = 2.5$ las proporciones de selección del modelo correcto generando los datos de la distribución WEI, LOGNOR, LOGLOGIS y IG con estas características, respectivamente se pueden observar en las Figuras 3.10-3.13

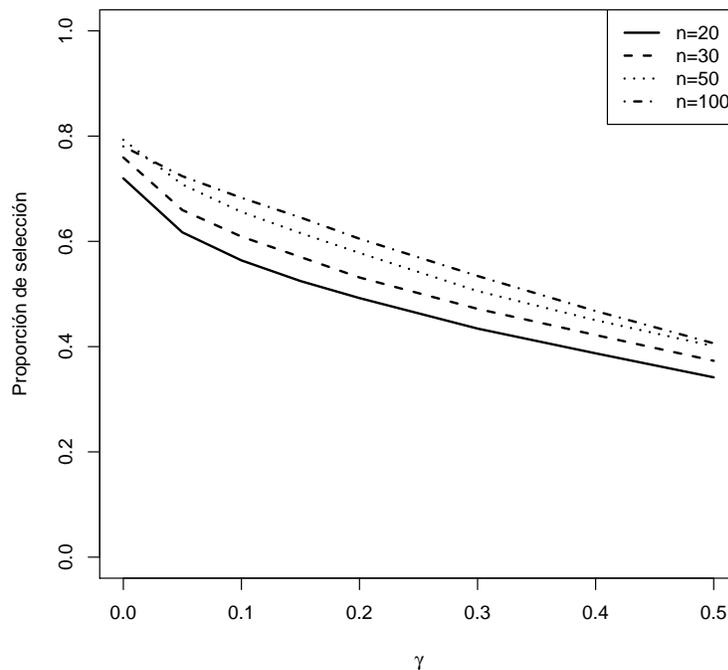


Figura 3.10. Proporción de veces que el AIC y BIC eligen al modelo correcto, siendo que los datos se generan de la distribución Weibull con media $E(T) = 50$ y coeficiente de variación $CV = 2.5$.

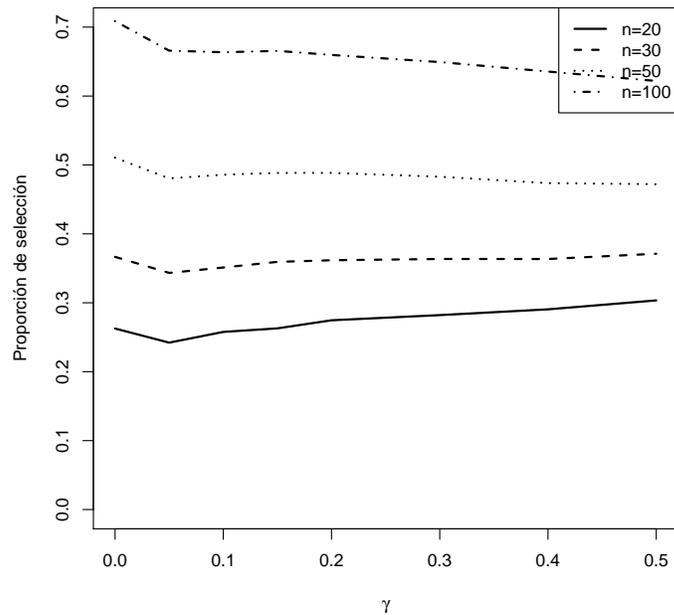


Figura 3.11. Proporción de veces que el AIC y BIC eligen al modelo correcto, siendo que los datos se generan de la distribución Lognormal con media $E(T) = 50$ y coeficiente de variación $CV = 2.5$.

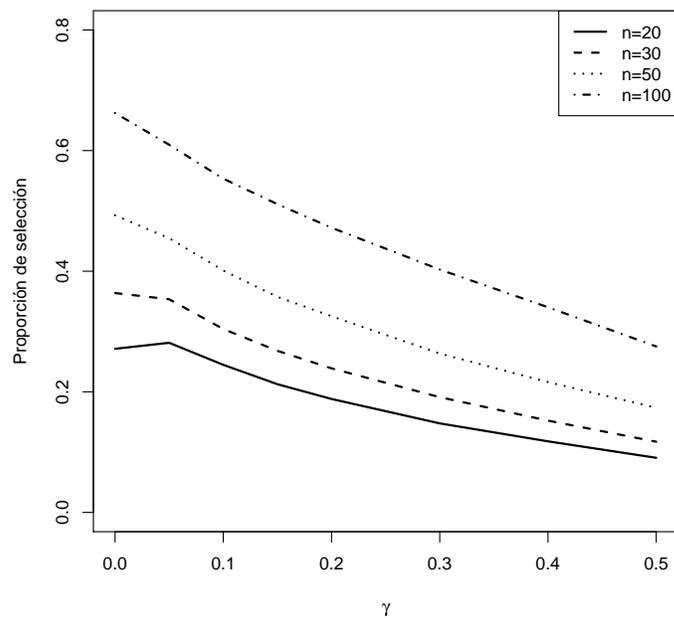


Figura 3.12. Proporción de veces que el AIC y BIC eligen al modelo correcto, siendo que los datos se generan de la distribución loglogística con media $E(T) = 50$ y coeficiente de variación $CV = 2.5$.

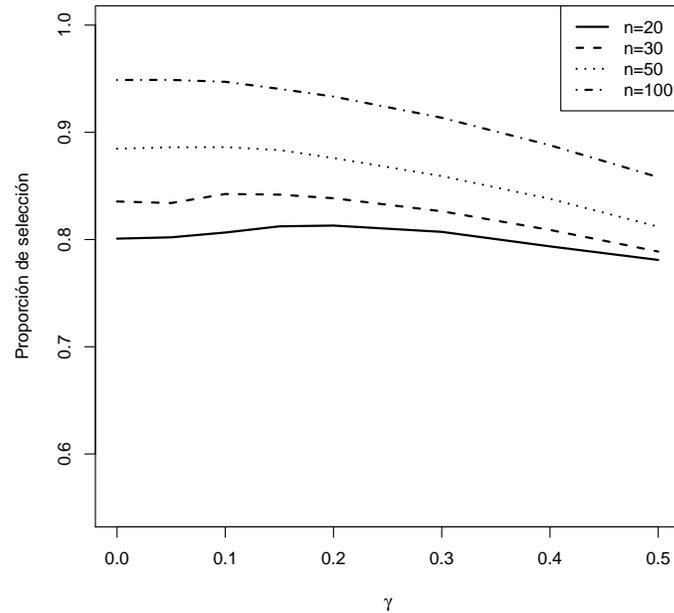


Figura 3.13. Proporción de veces que el AIC y BIC eligen al modelo correcto, siendo que los datos se generan de la distribución gaussiana inversa con media $E(T) = 50$ y coeficiente de variación $CV = 2.5$.

3.2.3. Distribuciones de uno, dos y tres parámetros

En esta parte se estudia el BIC y AIC pero ahora cuando se tiene el problema de discriminar entre las distribuciones Gama Generalizada, Weibull, Rayleigh, Exponencial y Log-normal. Como adelanto de los resultados que se obtienen en este escenario, de forma general se puede decir que se obtienen comportamientos bastante similares a los expuestos en las secciones anteriores.

Las distribuciones que aquí se consideran todas tienen un parámetro de escala, así que por el Apéndice C se sigue que la elección de éste no altera las proporciones de elección tanto del AIC como del BIC, esto es, los resultados son invariantes a la elección del parámetro de escala en la distribución que genere los datos. Los valores de los parámetros de las distribuciones con los cuales se experimentan en este apartado son los siguientes: $GG(\theta = 64, \beta = 5, \kappa = 3)$, $WEI(\beta = 1.5, \eta = 64)$, $RAY(\sigma = 45.6084)$, $EXP(\eta = 64.5981)$ y $LOGNOR(\mu = 4.839123, \sigma = 0.0365) \approx GG(\theta = 64, \beta = 5, \kappa = 30)$. La gráfica de probabilidad correspondiente a estas 5 distribuciones se muestra en la Figura 3.14.

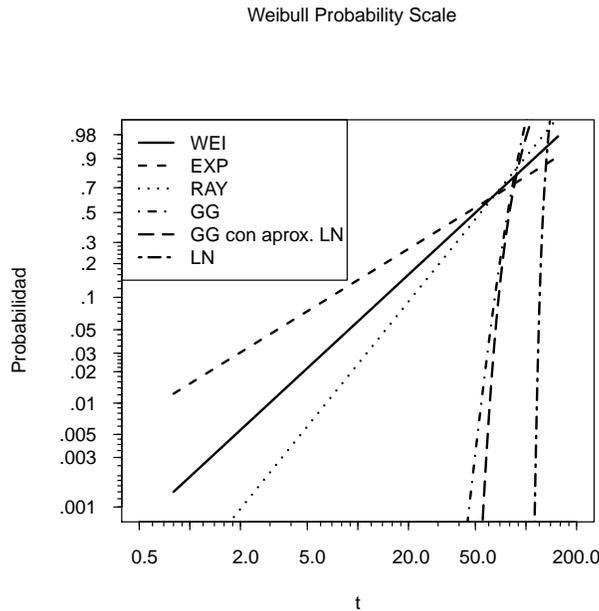


Figura 3.14. Gráficas de probabilidad GG, WEI, RAY, EXP y LOGNOR

Para cuando los datos se generan de las distribuciones $WEI(\beta = 1.5, \eta = 64)$, $RAY(\sigma = 45.6084)$, $EXP(\eta = 64.5981)$ y $LOGNOR(\mu = 4.839123, \sigma = 0.0365)$, las tablas de resultados como desplegadas en los casos de las Subsecciones anteriores no se presentan, pero sí las gráficas que resumen las probabilidades de selección del modelo correcto, esto debido a que el comportamiento de los resultados son análogos a los obtenidos en la Subsección 3.2.1. Por estas razones para estas distribuciones los comentarios que se hacen son más breves y hacen referencia a otros. Estas gráficas para las distribuciones WEI, RAY, EXP y LOGNOR respectivamente se encuentran en las Figuras 3.15–3.17. Para cuando los datos se generan de la distribución $WEI(\beta = 1.5, \eta = 64)$ (Figura 3.15) se observan resultados similares al caso estudiado en la Subsección 3.2.1. Se observa que para cada fracción de censura γ , a medida que n crece, la proporción de selección de ambos criterios (AIC y BIC) aumenta, y para cada n esta decrece conforme al fracción de censura tiende a crecer. Aquí el desempeño del AIC es superior al del BIC, aunque el desempeño de ambos es malo en los tamaños de muestra más pequeños. Observaciones similares se tienen cuando los datos son generados de la distribución $RAY(\sigma = 45.6084)$ y $EXP(\eta = 64.5981)$, solo que en estos casos el desempeño del BIC es mejor que del AIC y la presencia de censura no tiene un efecto negativo significativo en ambos criterios, al menos para las fracciones de censura que aquí se experimentan (Figuras

3.16 y 3.17).

Generando los datos del modelo LOGNOR($\mu = 4.839123, \sigma = 0.0365$), las proporciones de selección del modelo correcto se presentan en la Figura 3.17 en la cual se aprecian cosas observadas en escenarios estudiados: para cada tamaño de muestra, conforme la fracción de censura presente es cada vez mas grande, las proporciones de selección tienden a decrecer, y tal decrecimiento comienza de valores mayores a medida que n crece, aunque también se tiene que dicho decrecimiento es más rápido, teniendo un efecto casi nulo la presencia de censura para los tamaños de muestra mas pequeños. Aquí también el BIC es ligeramente superior al AIC.

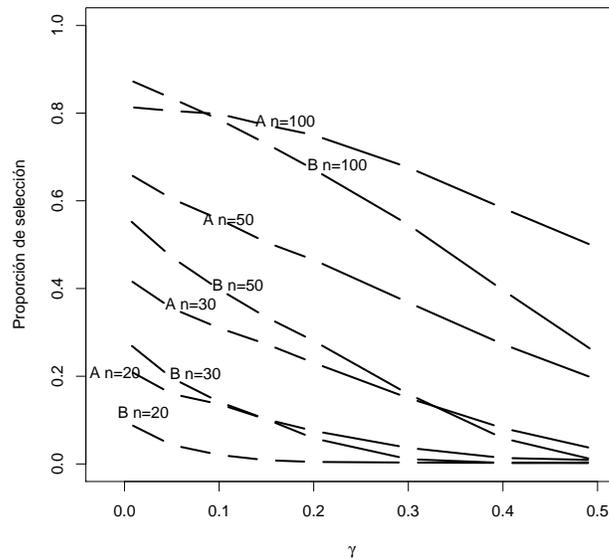


Figura 3.15. Proporción de veces que el AIC (A) y BIC (B) eligen a la distribución WEI como mejor modelo, siendo que de esta se generan los datos.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

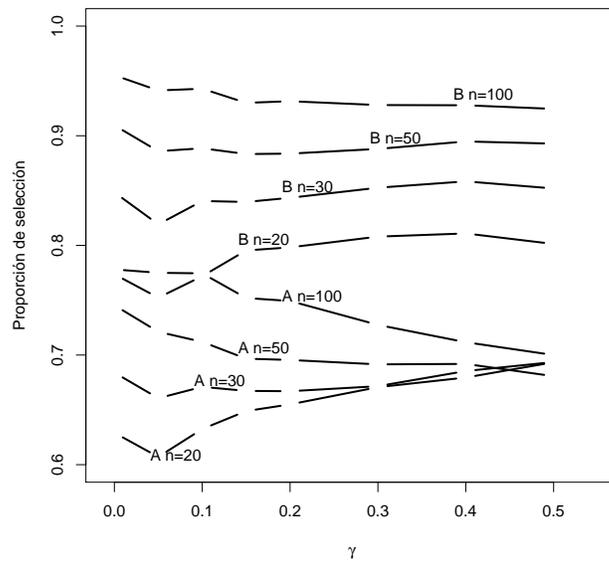


Figura 3.16. Proporción de veces que el AIC (A) y BIC (B) eligen a la distribución RAY como mejor modelo, siendo que de esta se generan los datos.

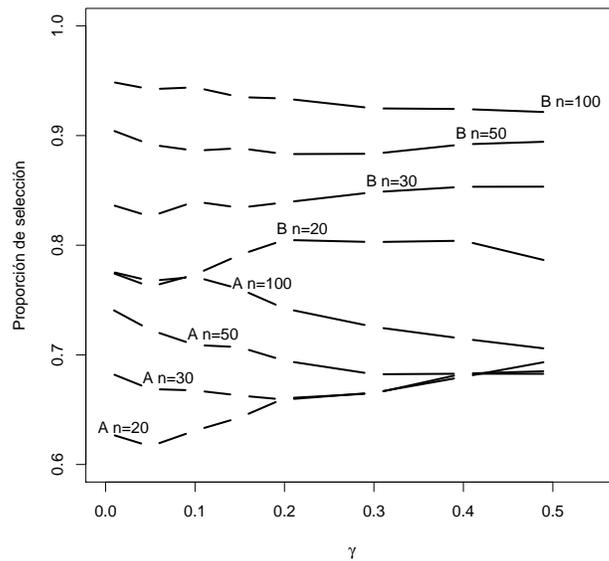


Figura 3.17. Proporción de veces que el AIC (A) y BIC (B) eligen a la distribución EXP como mejor modelo, siendo que de esta se generan los datos.

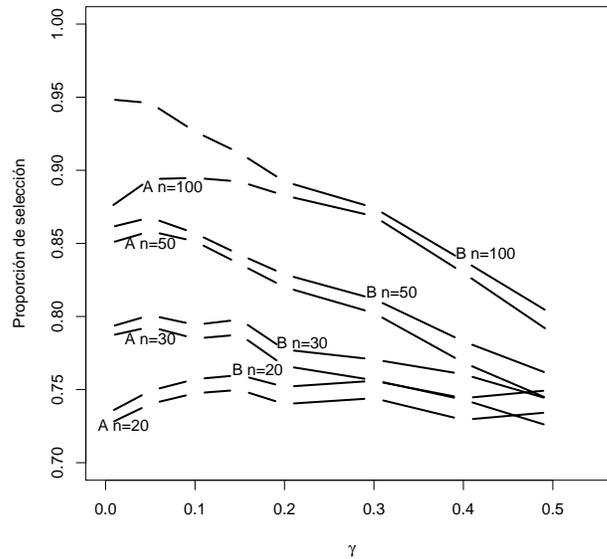


Figura 3.18. Proporción de veces que el AIC (A) y BIC (B) eligen a la distribución LOGNOR como mejor modelo, siendo que de esta se generan los datos.

Ahora, tomando a la distribución $GG(\theta = 64, \beta = 5, \kappa = 3)$ como la distribución verdadera, los resultados que se obtienen para tamaños muestra menores a $n = 100$ son bastantes desalentadores, por lo cual no se presentan. El desempeño de ambos criterios es bastante malo aun cuando la presencia de censura es nula y los tamaños de muestra son cercanos por debajo de 100. Sin embargo, para ilustrar como la fracción de censura afecta en este caso a la proporción de selección del modelo correcto, se experimentaron con tamaños de muestras mayores, $n = 100, 200, 300$ y 400 . Los resultados obtenidos con estos tamaño de muestra se presentan en los Cuadros 3.30–3.33. Para todos los tamaños de muestra y fracciones de censura considerados, se observa claramente la superioridad del AIC contra el BIC. Por ejemplo, para $n = 100$ las proporciones de selección del modelo correcto con el AIC comienzan a decrecer de 0.3737 hasta llegar a 0.0074, que es cuando se tiene 50% de censura, mientras que con el BIC tales proporciones decrecen de 0.0176 a 0. Con el AIC y fracciones de censura nula, el modelo Weibull es el más favorecido y el que genera los datos el segundo (GG), seguido por el modelo Lognormal, pero a medida que la fracción de censura γ se acerca a 0.5 este último modelo pasa a ocupar el primero, dejando al Weibull en segundo y desplazando al verdadero en el tercer lugar. Con el BIC algo similar ocurre solo que el modelo verdadero desde el comienzo ya es el tercero más elegido (Cuadro 3.30).

Al incrementar el tamaño de muestra a $n = 200$, del Cuadro 3.31 se observa que en ambos criterios las proporciones de selección del modelo correcto incrementan de forma muy significativa, pero a medida que la fracción de censura crece, estas proporciones tienden a decrecer rápidamente hasta casi alcanzar 0, llegando ambos criterios a favorecer más a la distribución Weibull y Lognormal, cuando la fracción de censura se acerca al 50 %. Para $n = 300$ también en ambos criterios se tiene que la proporción de selección del modelo correcto incrementa, aunque conforme la fracción de censura se aproxima a 0.5, esta tiende a decrecer más drásticamente que el caso anterior (Cuadro 3.32). Cuando $n = 400$ (Figura 3.33), con el BIC el decrecimiento de la proporción de selección del modelo correcto es todavía más rápido (de 0.7986 a 0), mientras que con el AIC se observa un decrecimiento ligeramente más lento (de 0.9712 a 0.1866) que el observado con $n = 300$ (de 0.9225 a 0.0555), aunque el desempeño de ambos criterios sigue siendo deficiente en la presencia de fracciones de censura cercanas al 50 %. Las proporciones de selección del modelo correcto gráficamente se pueden apreciar en la Figura 3.19, donde se refleja de mejor manera la superioridad que en este caso tiene el AIC hacia el BIC.

$n = 100$ Distribución		Fracción de censura γ							
		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	GG	0.3737	0.2228	0.1273	0.0598	0.0278	0.0120	0.0099	0.0074
	Weibull	0.4211	0.4532	0.5014	0.5152	0.5205	0.5026	0.4834	0.4646
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Lognormal	0.2052	0.3240	0.3713	0.4250	0.4517	0.4854	0.5067	0.5280
BIC	GG	0.0176	0.0007	0.0003	0.0005	0.0006	0.0012	0.0007	0.0000
	Weibull	0.6213	0.5799	0.5722	0.5524	0.5394	0.5128	0.4924	0.4718
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Lognormal	0.3611	0.4194	0.4275	0.4471	0.4600	0.4860	0.5069	0.5282

Cuadro 3.30. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 5 en comparación (GG, WEI, RAY, EXP y LOGNOR), siendo que los datos son generados de una distribución $GG(\theta = 64, \beta = 5, \kappa = 3)$ considerando un tamaño de muestra $n = 100$.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 200$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	GG	0.7831	0.6653	0.5607	0.4575	0.3531	0.1657	0.0410	0.0097
	Weibull	0.1808	0.2380	0.2958	0.3462	0.3984	0.4815	0.5307	0.5225
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Lognormal	0.0361	0.0967	0.1435	0.1963	0.2485	0.3528	0.4283	0.4678
BIC	GG	0.3437	0.1212	0.0305	0.0056	0.0011	0.0002	0.0001	0.0000
	Weibull	0.4892	0.5852	0.6206	0.6144	0.6054	0.5773	0.5564	0.5308
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Lognormal	0.1671	0.2936	0.3489	0.3800	0.3936	0.4225	0.4435	0.4692

Cuadro 3.31. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 5 en comparación (GG, WEI, RAY,EXP y LOGNOR), siendo que los datos son generados de una distribución $GG(\theta = 64, \beta = 5, \kappa = 3)$ considerando un tamaño de muestra $n = 200$.

$n = 300$		Fracción de censura γ							
Distribución		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	GG	0.9225	0.8550	0.7807	0.6962	0.6102	0.4167	0.2192	0.0555
	Weibull	0.0715	0.1165	0.1655	0.2180	0.2679	0.3710	0.4696	0.5321
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Lognormal	0.0060	0.0285	0.0539	0.0858	0.1219	0.2124	0.3112	0.4124
BIC	GG	0.6354	0.4198	0.2531	0.1156	0.0338	0.0012	0.0001	0.0000
	Weibull	0.3096	0.4348	0.5256	0.5932	0.6246	0.6190	0.5984	0.5657
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Lognormal	0.0550	0.1454	0.2213	0.2912	0.3416	0.3799	0.4016	0.4343

Cuadro 3.32. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 5 en comparación (GG, WEI, RAY,EXP y LOGNOR), siendo que los datos son generados de una distribución $GG(\theta = 64, \beta = 5, \kappa = 3)$ considerando un tamaño de muestra $n = 300$.

3.2. IMPLEMENTACIÓN DEL MECANISMO DE CENSURA TIPO I

$n = 400$ Distribución		Fracción de censura γ							
		0	0.05	0.1	0.15	0.2	0.3	0.4	0.5
AIC	GG	0.9712	0.9396	0.8927	0.8342	0.7645	0.5963	0.3982	0.1866
	Weibull	0.0278	0.0528	0.0892	0.1309	0.1770	0.2785	0.3867	0.4881
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Lognormal	0.0010	0.0076	0.0181	0.0349	0.0586	0.1252	0.2151	0.3253
BIC	GG	0.7986	0.6296	0.4763	0.3247	0.1836	0.0169	0.0004	0.0000
	Weibull	0.1859	0.3059	0.4081	0.4968	0.5690	0.6398	0.6241	0.5970
	Rayleigh	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Exponencial	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Lognormal	0.0156	0.0645	0.1156	0.1786	0.2474	0.3433	0.3755	0.4030

Cuadro 3.33. Proporción de veces que los criterios AIC y BIC indican que la distribución más adecuada es una de las 5 en comparación (GG, WEI, RAY, EXP y LOGNOR), siendo que los datos son generados de una distribución $GG(\theta = 64, \beta = 5, \kappa = 3)$ considerando un tamaño de muestra $n = 400$.

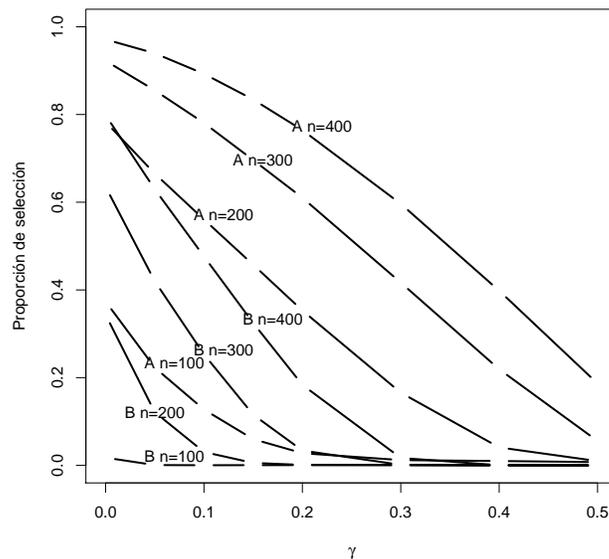


Figura 3.19. Proporción de veces que el AIC (A) y BIC (B) eligen a la distribución GG como mejor modelo, siendo que de esta se generan los datos.

CAPÍTULO 4

Conclusiones

Después de hacer un estudio de simulación para comparar el desempeño de los criterios AIC y BIC, en la selección de modelos, bajo la presencia de censura, extraemos las siguientes conclusiones.

Al hacer la discriminación entre los modelos WEI, EXP y RAY, se observa que generando los datos con uno de los modelos anidados en el modelo WEI, esto es EXP o RAY, la eficiencia del criterio BIC es superior al AIC. Esta superioridad va siendo cada vez mayor a medida que el tamaño de muestra aumenta, manteniéndose casi constante alrededor de 0.84 la proporción de selección del modelo correcto con el índice AIC, en todas las fracciones de censura que se consideran. Aquí las fracciones de censura que se estudian, no afectan significativamente el desempeño de ambos criterios. En la discriminación entre los mismos modelos considerados anteriormente, pero generando los datos con el modelo $WEI(\eta, 1.5)$, el desempeño del criterio AIC es superior al BIC, observando en ambos casos que a medida que el tamaño de muestra crece, la eficiencia de ambos criterios aumenta, siendo estos cada vez más semejantes. Para tamaños de muestra $n = 20, 30$ y 50 , la proporción de selección del modelo correcto por parte del criterio BIC es fuertemente afectada por la fracción de censura, ocurriendo lo mismo para el criterio AIC, pero solo para los dos tamaños de muestra más pequeños. Comportamientos similares se observan cuando el parámetro de forma se

elige cerca de 1.5, pero a medida que este se acerca a 1 o 2, ambos criterios se confunden cada vez más, necesitando tamaños de muestra más grandes para mejorarla identificación del modelo correcto. Conforme β se aleja de 1 por debajo o se aleja de 2 por arriba, la capacidad de selección de ambos criterios aumenta y va siendo cada vez menos afectada por la fracción de censura. Esto es algo esperado, ya que el modelo verdadero se aleja cada vez más de los modelos RAY y EXP. En este caso, se observa que ambos criterios son consistentes. En este caso, el criterio AIC es preferible para muestras pequeñas y el BIC lo es para muestras grandes. Otro caso de interés fue la discriminación entre cuatro modelos que frecuentemente compiten cuando se tienen datos de vida: Weibull, Lognormal, Loglogístico y Gaussiana inversa. Cuando se simularon datos de estas distribuciones con media $\mathbb{E}(T) = 50$ y coeficiente de variación $CV = 0.9$, la fracción de censura fue un factor importante que afectó significativamente la capacidad de selección del modelo correcto para ambos criterios, AIC y BIC. Este efecto fue aún mayor cuando se tomaron muestras pequeñas. En este estudio, el resultado más inesperado se observó cuando los datos se generaron con la distribución Lognormal, donde la distribución Gaussiana inversa fue mayormente favorecida para todas las combinaciones de n y γ consideradas en el estudio, salvo en $n = 100$ y $\gamma = 0$. Una explicación de esto, es que para valores de CV menores que 1.0, estas dos distribuciones tienden a ser muy parecidas. Cuando en el estudio se generaron los datos con un valor de CV igual a 2.5, con el modelo Loglogístico se obtuvo un comportamiento prácticamente análogo al caso $CV = 0.9$, mientras que con el modelo Weibull se obtuvieron resultados ligeramente menos favorables. Con el modelo Lognormal, al menos con $n = 50$, el problema anteriormente mencionado ya no se tiene, ya que aquí el modelo correcto es elegido con mayor frecuencia y se enfatiza más a medida que crece el valor del CV y el tamaño de muestra es mayor.

Al considerar el problema de discriminar entre los modelos GG, WEI, EXP, RAY y LOGNOR, se observa que conforme la presencia de fracción de censura γ se hace más grande, el porcentaje de selección del modelo correcto decrece, excepto cuando el modelo que genera los datos es el RAY o EXP, donde únicamente el tamaño de muestra, n , tiene un efecto positivo sobre las proporciones de selección a medida que este aumenta. Cuando los datos son generados de la distribución $WEI(\beta = 1.5, \eta = 64)$, el AIC resulta superior al BIC, no obstante con tamaños de muestra pequeños y fracciones de censura “grandes” el desempeño de ambos resulta bastante malo (Figura 3.15. Contrario al caso Weibull, cuando los datos se generaron del modelo RAY, EXP o LOGNOR el desempeño del modelo AIC resulta inferior

al del BIC. Cuando los datos fueron extraídos del modelo $GG(\theta = 64, \beta = 5, \kappa = 3)$ el desempeño observado para ambos criterios fue pésimo, con proporciones de selección cercanas a cero aún con las fracciones de censura más pequeñas y con los tamaños de muestra más grande, de las aquí consideradas. Sin embargo, al experimentar con tamaños de muestra más grandes, $n = 100, 200, 300$ y 400 , además de observar el efecto positivo del tamaño de muestra sobre las proporciones de selección del modelo correcto, también se observa el efecto negativo de las fracciones de censura sobre estas, incluso con el n más grande, $n = 400$ (Figura 3.19).

En general, en los escenarios estudiados se observó que la fracción de censura afecta significativamente el desempeño de ambos criterios AIC y BIC. Este efecto disminuye a medida que aumenta el tamaño muestral, aunque en muchos casos se requiere un incremento muy grande, hasta alcanzar tamaños de muestra que en la práctica con frecuencia no se tienen. Por ello se deben tener precauciones en el uso de estos índices como criterios de selección de modelos de datos de vida, en presencia de censura y tamaños de muestra pequeños, ya que nos pueden llevar a elegir un modelo incorrecto con alta probabilidad. Por esto, en situaciones similares a las aquí estudiadas, estos criterios pueden usarse como una herramienta complementaria, mas no como la única que determine la elección del modelo. Incluso, deben tomarse en cuenta otras fuentes de información, como por ejemplo la física o la química relacionada con el patrón de falla de los componentes que se estudian. Esto último es importante sobre todo, cuando nos interesa hacer extrapolaciones.

Además de extender el estudio anterior en la discriminación de otros modelos y otras fracciones de censura de interés, se podría estudiar el desempeño de ambos criterios en modelos de regresión de confiabilidad, ya sea en la elección del número de covariables a incluir, o elegir entre modelos no necesariamente anidados.

Algunas distribuciones de tiempo de vida

A.1. Distribución Weibull

Una variable aleatoria T tiene distribución Weibull con parámetro de escala $\eta > 0$ y parámetro de forma $\beta > 0$, $T \sim \text{WEI}(\eta, \beta)$, si su función de densidad y función de distribución están respectivamente dadas por

$$f(t; \eta, \beta) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^\beta\right]$$

$$F(t; \eta, \beta) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^\beta\right], \quad t \geq 0.$$

La utilidad práctica de esta distribución se deriva de su habilidad de modelar tiempos de falla en diferentes disciplinas.

La media y la varianza de la distribución Weibull están respectivamente dadas por

$$E(T) = \eta\Gamma(1 + 1/\beta)$$

$$\text{Var}(T) = \eta^2 [\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)]$$

donde $\Gamma(x) = \int_0^\infty z^{x-1} \exp(-z) dz$ es la función gama, $x > 0$.

Cuando $\beta = 2$, la distribución Weibull se reduce a la distribución Rayleigh con parámetro $\sigma = \eta/\sqrt{2}$, RAY(σ), con función de densidad y función de distribución dadas por

$$f(t; \sigma) = \frac{t}{\sigma^2} \exp \left[- \left(\frac{t}{\sigma\sqrt{2}} \right)^2 \right]$$

$$F(t; \sigma) = 1 - \exp \left[- \left(\frac{t}{\sigma\sqrt{2}} \right)^2 \right], \quad t \geq 0.$$

La media y la varianza de la distribución RAY(σ) son respectivamente, $E(T) = \sigma\sqrt{\pi}/\sqrt{2}$ y $\text{Var}(T) = (4 - \pi)\sigma^2/2$.

Ahora, si $\beta = 1$, la distribución Weibull se reduce a la distribución Exponencial, EXP(η), con función de densidad y función de distribución respectivamente dadas por

$$f(t; \eta) = \frac{1}{\eta} \exp \left(-\frac{t}{\eta} \right)$$

$$F(t; \eta) = 1 - \exp \left(-\frac{t}{\eta} \right), \quad t \geq 0.$$

La media y la varianza de una variable aleatoria $T \sim \text{EXP}(\eta)$ son iguales:

$$E(T) = \text{Var}(T) = \eta.$$

A.2. Distribución Lognormal

Una variable aleatoria T tiene distribución Lognormal con parámetros $-\infty < \mu < \infty$ y $\sigma > 0$, $T \sim \text{LOGNOR}(\mu, \sigma)$, si $\log(T) \sim \text{NOR}(\mu, \sigma)$, es decir, si su función de densidad y función de distribución están respectivamente dadas por

$$f(t; \mu, \sigma) = \frac{1}{\sigma t} \phi_{\text{nor}} \left[\frac{\log(t) - \mu}{\sigma} \right]$$

$$F(t; \mu, \sigma) = \Phi_{\text{nor}} \left[\frac{\log(t) - \mu}{\sigma} \right], \quad t \geq 0$$

donde $\phi_{\text{nor}}(\cdot)$ y $\Phi_{\text{nor}}(\cdot)$ denotan la función de densidad y función de distribución acumulada de una variable aleatoria normal estándar. En este caso $\exp(\mu)$ es un parámetro de escala, mientras que σ es un parámetro de forma.

La media y varianza de esta distribución son,

$$E(T) = \exp \left(\mu + \frac{1}{2}\sigma^2 \right)$$

$$\text{Var}(T) = \exp(2\mu + \sigma^2) [\exp(\sigma^2) - 1]$$

La distribución lognormal es un modelo común para tiempos de falla. Siguiendo el teorema del límite central, esta distribución puede justificarse por ejemplo para una variable que proviene del producto de un número grande de variables aleatorias positivas, independientes, e idénticamente distribuidas.

A.3. Distribución Loglogística

Una variable aleatoria T tiene distribución Loglogística, $T \sim \text{LOGLOGIS}(\mu, \sigma)$, si su función de densidad y función de distribución están respectivamente dadas por

$$f(t; \mu, \sigma) = \frac{1}{\sigma t} \phi_{\text{logis}} \left[\frac{\log(t) - \mu}{\sigma} \right]$$

$$F(t; \mu, \sigma) = \Phi_{\text{logis}} \left[\frac{\log(t) - \mu}{\sigma} \right], \quad t \geq 0$$

donde $-\infty < \mu < \infty$, $\sigma > 0$, $\phi_{\text{logis}}(z) = \exp(z) / [1 + \exp(z)]^2$ y $\Phi_{\text{logis}}(z) = \exp(z) / [1 + \exp(z)]$. $\exp(\mu)$ es un parámetro de escala y σ es un parámetro de forma.

La media y la varianza de esta distribución son respectivamente:

$$E(T) = \begin{cases} \exp(\mu)\Gamma(1 + \sigma)\Gamma(1 - \sigma), & \text{si } \sigma < 1 \\ \text{no existe,} & \text{si } \sigma \geq 1 \end{cases}$$

$$\text{Var}(T) = \begin{cases} \exp(2\mu)\Gamma(1 + 2\sigma)\Gamma(1 - 2\sigma) - E^2(T), & \text{si } \sigma < 1/2 \\ \text{no existe,} & \text{si } \sigma \geq 1/2. \end{cases}$$

La forma de la distribución loglogística es similar a la de la distribución lognormal.

A.4. Distribución Inversa Gaussiana

T tiene distribución Inversa Gaussiana, $T \sim \text{IG}(\theta, \beta)$, si su función de densidad y función de distribución están respectivamente dadas por

$$f(t; \theta, \beta) = \frac{1}{t} \phi_{\text{igau}} [\log(t/\theta); \beta]$$

$$F(t; \theta, \beta) = \Phi_{\text{igau}} [\log(t/\theta); \beta], \quad t \geq 0$$

donde $\theta > 0$ es parámetro de escala y $\beta > 0$ es un parámetro de forma sin unidades,

$$\begin{aligned}\phi_{\text{ligau}}(z; \beta) &= \frac{\sqrt{\beta}}{\exp(z/2)} \phi_{\text{nor}} \left\{ \sqrt{\beta} \left[\frac{\exp(z) - 1}{\exp(z/2)} \right] \right\} \\ \Phi_{\text{ligau}}(z; \beta) &= \Phi_{\text{nor}} \left\{ \sqrt{\beta} \left[\frac{\exp(z) - 1}{\exp(z/2)} \right] \right\} + \exp(2\beta) \Phi_{\text{nor}} \left\{ -\sqrt{\beta} \left[\frac{\exp(z) + 1}{\exp(z/2)} \right] \right\}.\end{aligned}$$

La media y la varianza para esta distribución son $E(T) = \theta$ y $\text{Var}(T) = \theta^2/\beta$, respectivamente.

A.5. Distribución Gama Generalizada

T tiene distribución Gama Generalizada, $T \sim \text{GG}(\theta, \beta, \kappa)$, si su función de densidad y función de distribución están respectivamente dadas por

$$\begin{aligned}f(t; \theta, \beta, \kappa) &= \frac{\beta}{\Gamma(\kappa)\theta} \left(\frac{t}{\theta}\right)^{\kappa\beta-1} \exp\left[-\left(\frac{t}{\theta}\right)^\beta\right] \\ F(t; \theta, \beta, \kappa) &= \Gamma_1\left[\left(\frac{t}{\theta}\right)^\beta; \kappa\right] \quad t \geq 0\end{aligned}$$

donde $\theta, \beta, \kappa > 0$ y $\Gamma_1(v; \kappa) = \frac{\int_0^v x^{\kappa-1} \exp(-x) dx}{\Gamma(\kappa)}$.

La media y varianza de esta distribución son:

$$\begin{aligned}E(T) &= \frac{\theta\Gamma(1/\beta + \kappa)}{\Gamma(\kappa)} \\ \text{Var}(T) &= \theta^2 \left[\frac{\Gamma(2/\beta + \kappa)}{\Gamma(\kappa)} - \frac{\Gamma^2(1/\beta + \kappa)}{\Gamma^2(\kappa)} \right].\end{aligned}$$

APÉNDICE B

Algunos resultados

Definición B.1

$$I(\theta) = \int g(x) \frac{\partial \log[f(x|\theta)]}{\partial \theta} \frac{\partial \log[f(x|\theta)]}{\partial \theta'} dx \quad (\text{B.1})$$

$$J(\theta) = - \int g(x) \frac{\partial^2 \log[f(x|\theta)]}{\partial \theta \partial \theta'} dx \quad (\text{B.2})$$

Resultado B.1 Si $g(x)$ se encuentra anidada en $\{f(x|\theta) : \theta \in \mathbf{R}^p\}$, es decir, $g(x) = f(x|\theta^*)$ para algún $\theta^* \in \mathbf{R}^p$, entonces

1. $\theta_0 = \theta^*$
2. y bajo condiciones de regularidad, $I(\theta_0) = J(\theta_0)$.

Obsérvese que bajo los mismos supuestos del resultado (B.1), la ecuación (B.2) corresponde a la forma alternativa de calcular la información esperada de Fisher, $I(\theta_0)$.

Resultado B.2 Si X_1, \dots, X_n es una muestra aleatoria de $g(x)$, $\hat{\theta}_n$ es el emv de θ , y $\theta_0 = \arg \min_{\theta \in \Theta} I[g(\cdot), f(\cdot|\theta)]$, entonces bajo condiciones de regularidad

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{d} \text{MVN}_p[0, J^{-1}(\theta_0)I(\theta_0)J^{-1}(\theta_0)]. \quad (\text{B.3})$$

Notése que si se tienen las condiciones del Resultado (B.1), este resultado es el correspondiente a la normalidad del emv de θ , $\hat{\theta}_n$, en la teoría clásica de máxima verosimilitud en la que se supone que el verdadero modelo es $f(x|\theta_0)$. Una prueba de este resultado se puede ver en Konishi y Kitagawa (2008). Otra en Claeskens y Hjort (2008).

APÉNDICE C

Anexos

Supongamos que se tiene una variable aleatoria continua T con función de distribución $G(t|\alpha, \theta)$ y función de densidad $g(t|\alpha, \theta)$, donde α y θ son sus parámetros. Supongamos además que el primero de ellos, α , es un parámetro de escala, esto es, la densidad (g) y función de distribución (G) de X pueden escribirse respectivamente como

$$g(t|\alpha, \theta) = \frac{1}{\alpha}g(t/\alpha|1, \theta) \quad (\text{C.1})$$

$$G(t|\alpha, \theta) = G(t/\alpha|1, \theta) \quad (\text{C.2})$$

Sean t_1, \dots, t_r los tiempos de falla ordenados observados de T hasta el tiempo de censura $t_{1-\gamma}$, de una muestra de tamaño n . Sean $f_1(t|\alpha_1, \theta_1), \dots, f_p(t|\alpha_p, \theta_p)$ y $F_1(t|\alpha_1, \theta_1), \dots, F_p(t|\alpha_p, \theta_p)$ respectivamente las densidades y funciones de distribución, de p modelos alternativos para modelar estos datos, donde para el j -ésimo modelo ($j = 1, \dots, p$), α_j representa un parámetro de escala y θ_j sus restantes parámetros.

Observemos que si $\hat{\alpha}_j$ y $\hat{\theta}_j$ son los EMV de α_j y θ_j , respectivamente, entonces de (C.1) y (C.2) la verosimilitud maximizada para el j -ésimo modelo se puede expresar como

$$\begin{aligned}
L_j(\widehat{\alpha}_j, \widehat{\theta}_j; \mathbf{t}) &= \max_{\alpha_j, \theta_j} \left[\prod_{i=1}^r f_j(t_i | \alpha_j, \theta_j) \right] [1 - F_j(t_{1-\gamma} | \alpha_j, \theta_j)]^{n-r} \\
&= \max_{\alpha_j, \theta_j} \left[\prod_{i=1}^r \frac{1}{\alpha_j} f_j(t_i / \alpha_j | 1, \theta_j) \right] [1 - F_j(t_{1-\gamma} / \alpha_j | 1, \theta_j)]^{n-r} \\
&= \max_{\alpha_j, \theta_j} \left[\prod_{i=1}^r \frac{1/\alpha}{\alpha_j/\alpha} f_j[(t_i/\alpha)/(\alpha_j/\alpha) | 1, \theta_j] \right] \{1 - F_j[(t_{1-\gamma}/\alpha)/(\alpha_j/\alpha) | 1, \theta_j]\}^{n-r} \\
&= (1/\alpha)^r \max_{\alpha_j, \theta_j} \left[\prod_{i=1}^r f_j(t_i/\alpha | \alpha_j/\alpha, \theta_j) \right] [1 - F_j(t_{1-\gamma}/\alpha | \alpha_j/\alpha, \theta_j)]^{n-r} \\
&= \alpha^{-r} \max_{\alpha_j^*, \theta_j} \left[\prod_{i=1}^r f_j(t_i^* | \alpha_j^*, \theta_j) \right] [1 - F_j(t_{1-\gamma}^* | \alpha_j^*, \theta_j)]^{n-r} \\
&= \alpha^{-r} L_j(\widehat{\alpha}_j^*, \widehat{\theta}_j; \mathbf{t}^*)
\end{aligned}$$

donde $\alpha_j^* = \alpha_j/\alpha$, $t_i^* = t_i/\alpha$, $t_{1-\gamma}^* = t_{1-\gamma}/\alpha$ y $\widehat{\alpha}_j^*$ es el EMV del parámetro de escala bajo el modelo j , α_j , pero con una muestra t_1^*, \dots, t_r^* que proviene de una variable aleatoria $T^* \stackrel{d}{=} T/\alpha$, esto es, con densidad $g(t|1, \theta)$ y función de distribución $G(t|1, \theta)$, donde el parámetro de escala es conocido e igual a 1. Además obsérvese que $G(t_{1-\gamma}^* | 1, \theta) = G(t_{1-\gamma}/\alpha | 1, \theta) = G(t_{1-\gamma} | \alpha, \theta) = 1 - \gamma$. De este modo, las log-verosimilitudes de los p modelos alternativos pueden expresarse de la siguiente manera

$$l_j(\widehat{\alpha}_j, \widehat{\theta}_j; \mathbf{t}) = l_j(\widehat{\alpha}_j^*, \widehat{\theta}_j; \mathbf{t}^*) - r \log(\alpha), \quad j = 1, \dots, p$$

donde \mathbf{t}^* es el vector que contiene a las t_i^* , las cuales corresponden a observaciones de una variable aleatoria T^* con densidad y función de distribución $g(t|1, \theta)$ y $G(t|1, \theta)$, respectivamente. Así, ya que $r \log(\alpha)$ es un término común en todos los modelos, se sigue que tanto en el AIC como en el BIC la elección del mejor modelo es independiente del parámetro de escala de la distribución de la cual provienen los datos, bajo las condiciones en las cuales tanto el modelo que genera los datos como los modelos alternativos tienen un parámetro de escala. Esto es, repeticiones independientes de las simulaciones para dos casos en el que únicamente el parámetro de forma difiere en el modelo que genera los datos, proveen valores del AIC y BIC que son versión escalada uno del otro.

Bibliografía

- Abreu G. C. G., Edwards D., Labouriau. (2010). High-Dimensional Graphical Model Search with the gRapHD R Package. *Journal of Statistical Software*, 37(1), 1–18.
- Akaike, H. (1974). A new look at the statistical identification model, *IEEE Transactions on Automatic Control*, 6, 716–723.
- Anderson, D. R., Burnham, K. P. and White, G. C. (1994). AIC model selection in overdispersed capture-recapture data. *Ecology*, 75, 1780–1793.
- Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information–Theoretic Approach*, Springer, New York.
- Claeskens, G. and Hjort, G., L. (2008). *Model Selection and Model Averaging*, Cambridge: Cambridge University Press.
- Dennis B., Ponciano J. M., Lele S., Taper M. and D. Staples. (2006). Estimating density dependence, process noise and observation error. *Ecological Monographs*, 76(3), 323-341.
- Edwards D., Abreu G. C. G., Labouriau R (2010). Selecting High-dimensional Mixed Graphical Models Using Minimal AIC or BIC Forests. *BMC Bioinformatics*, 11(1), 18.
- Johnson, J. B., and K. S. Omland. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19, 101-108.

- Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*, Springer Series in Statistics. Springer Verlag, New York.
- Kullback, S. (1997). *Information Theory and Statistics*, Dover Publications, New York.
- Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*, John Wiley and Sons, New York.
- P. Lahiri, W. (2001). *Model Selection*, Institute of Mathematical Statistics Lecture Notes—Monograph Series volume 38, pp.135-207, Beachwood Ohio.
- Ponciano, J. M., Mark L. Taper, Brian C. Dennis and Subhash R. Lele. (2009). Hierarchical models in Ecology: Confidence intervals, hypothesis testing and model selection using data cloning. *Ecology*, 90(2), 356-362.
- Shibata, R. (1995). Bootstrap Estimate of Kullback-Leibler Information for Model Selection, *Technical Report No. 424*, Department of Statistical University of California Berkeley, California.
- Takagi, K., Kumagai S., Matsunaga I. and Kusaka Y. (1997). Application of inverse gaussian distribution to occupational exposure data, *Ann Occup Hyg*, 41, 505–514.
- Wasserman, L. A. (2004). *All of Statistics: A Concise Course in Statistical Inference*, Springer, New York.