

ENFOQUE DEL MODELO DE REGRESIÓN LOGÍSTICA USANDO EL MUESTREADOR DE GIBBS.

Vivian Sistachs Vega¹ y María Victoria Mederos²

Dpto. de Matemática Aplicada, Facultad de Matemática y Computación, Universidad de La Habana
Miguel A. Díaz Martínez, Dpto. de Matemática General, Facultad de Ingeniería Industrial,
Instituto Superior Politécnico "José A. Echevarría" (ISPJAE)

RESUMEN

El modelo de Regresión Logística es un modelo muy utilizado en diferentes ramas de la ciencia donde se estudian problemas de pronóstico cuando la variable respuesta es binaria. Este modelo puede ser estudiado bajo el enfoque bayesiano, en el cual los cálculos pueden ser complicados incluso utilizando medios de cómputo; es por esto que tendremos en cuenta el algoritmo Muestreador de Gibbs, que es un método iterativo para obtener de forma aproximada la distribución posterior de los parámetros en el modelo. Para la obtención de la distribución con este método podemos utilizar el software Winbugs, que no muestra claramente cómo se obtienen los resultados, y es por ello que en este trabajo hacemos una comparación de su empleo a través del software Winbugs 1.4 y de la implementación del algoritmo en MATLAB. Esta comparación se realiza a través de un ejemplo de pronóstico bioclimático de la frecuencia de ocurrencia de la enfermedad ASMA, donde se analizan las salidas del Winbugs y del programa en MATLAB para dicho ejemplo, y también a través de un estudio de simulación.

ABSTRACT

The logistic regression model is very useful in different areas of the science for studying forecast problems when the response variable is binary. This model might be studied by the Bayesian approach, but in this case the computations can be very difficult. The Gibbs Sampler algorithm supported on the Winbugs software is an alternative method to obtain of posterior distribution of the parameters of the model. This software does not show how the results are obtained. For this reason in this work a new implementation in MATLAB of this algorithm is proposed. A comparison of both implementation was made through a bioclimatic forecast example of the frequency of the ASMA disease and a simulation study.

Key words: Bayesian approach, forecasting, Winbugs.

MSC: 62J05.

1. INTRODUCCIÓN

La Regresión Logística es uno de los métodos estadísticos más expresivos y versátiles disponibles para el análisis de datos. Muchos especialistas de diferentes ramas han trabajado con ella para predecir o pronosticar una variable respuesta binaria o dicotómica, donde las variables independientes pueden ser de cualquier naturaleza, convirtiéndola en un método estándar para el análisis de regresión cuando los datos son binarios (Hosmer and Lemeshow, 1989, Silva, 1994).

El objetivo de la Regresión Logística es encontrar el mejor ajuste del modelo con el menor número de parámetros y describir la relación entre la variable respuesta y un conjunto de variables (covariables) explicatorias independientes.

El modelo de Regresión Logística se define como

$$\hat{p}(y = 1 / X) = \left\{ 1 + \exp \left[- \left(\hat{\beta}_0 + \sum_{i=1}^d \hat{\beta}_i x_i \right) \right] \right\}^{-1}$$

donde $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_d$ son los coeficientes de la regresión y x_1, x_2, \dots, x_d son las componentes de X .

E-mail: ¹yivian@matcom.uh.cu
²mmederos@matcom.uh.cu

La estadística bayesiana proporciona un enfoque diferente al problema del desconocimiento de los parámetros del modelo, ya que en vez de considerar un único valor de los parámetros del modelo (como lo hace la estimación máximo verosímil) la inferencia bayesiana expresa la incertidumbre de los parámetros en términos de distribución de probabilidad.

En el trabajo se formula el modelo de Regresión Logística bajo el paradigma bayesiano, utilizando una distribución inicial no informativa y el algoritmo del Muestreador de Gibbs a través del cual se determina la distribución final de los parámetros del modelo.

Para la obtención de la distribución con este algoritmo podemos utilizar el software Winbugs que no muestra claramente, cómo se llega a los resultados y es por ello que haremos una comparación del empleo del Muestreador de Gibbs a través del software Winbugs 1.4 y de la implementación de este algoritmo en MATLAB.

El problema del modelo de regresión logística bajo el paradigma bayesiano se abordará en el epígrafe 2 y a través de un ejemplo de pronóstico bioclimático se realiza en el epígrafe 3 el análisis del muestreador de Gibbs usando el Winbugs y la implementación en MATLAB. En el epígrafe 4 se realiza la comparación general por simulación del algoritmo para la implementación del mismo en MATLAB.

2. PARADIGMA BAYESIANO PARA EL MODELO DE REGRESIÓN LOGÍSTICA

La base del paradigma bayesiano es encontrar la distribución posterior del o de los parámetros o cantidades de interés en el modelo, donde resulta interesante el hecho de que los parámetros son tratados como variables aleatorias, o sea son una descripción de la incertidumbre del modelo cuantificada a través de la probabilidad (Bernardo, 2003).

Como problema central de este paradigma está el hecho de proporcionar una metodología que permita asimilar la información que se tiene con el objetivo de mejorar el conocimiento del mundo real.

La metodología del paradigma bayesiano consta de:

- Proceso de aprendizaje, el cual constituye la base de todo problema de inferencia sobre el valor del parámetro, y se reduce básicamente a determinar su distribución posterior o final.
- Distribución predictiva, la cual es utilizada para poder describir la información que se posee sobre posibles valores de las observaciones.
- Comportamiento asintótico, análisis que se realiza y es más preciso en la medida en que se disponga de una mayor cantidad de datos.

En nuestro trabajo abordaremos sólo el problema del proceso de aprendizaje ya que nos interesa analizar la estimación de los parámetros de la distribución del modelo de regresión logística.

2.1. Proceso de aprendizaje

Este proceso permite incorporar la información proporcionada por los datos y consta de:

1. Distribución a priori (información inicial).
2. Función de verosimilitud (información de los datos).
3. Distribución posterior (El cómputo lo facilita el teorema de Bayes mediante los métodos MCMC).

En la práctica el cálculo de la distribución posterior se hace a través del teorema de Bayes, y se utiliza el equivalente que considera la distribución posterior proporcional al producto de la verosimilitud por la a priori.

$$p(\theta/x) \propto f(x/\theta) p(\theta)$$

Hay algunas observaciones que debemos destacar como son:

- Desafortunadamente la implementación de las técnicas Bayesianas usualmente requiere de un esfuerzo computacional muy alto.
- Excepto en aplicaciones muy sencillas tanto $f(x/\theta)$ como $p(\theta)$ pueden llegar a tener formas muy complicadas.
- En la gran mayoría de los problemas las integrales requeridas no pueden resolverse analíticamente.
- Es necesario contar con métodos numéricos eficientes que permitan calcular o aproximar integrales en varias dimensiones.

Carlin 1996 plantea varias vías para obtener la distribución posterior que es en la que se involucran las integrales, estos métodos son:

- Métodos Asintóticos. (Aproximación normal y método de Laplace). Para que sea válida la distribución posterior debe ser unimodal.
- Métodos Monte Carlo no iterativos. (muestreo directo y muestreo indirecto).
- Métodos MCMC. (Metropolis-Hastings y Muestreo de Gibbs).

Como se había planteado las distribuciones a posteriori necesarias no siempre pueden ser obtenidas analíticamente y las aproximaciones asintóticas son sólo apropiadas bajo ciertas circunstancias. Por ello se utilizan métodos Cadena de Markov-Integración Monte Carlo (MCMC), específicamente el algoritmo Muestreador de Gibbs (Gibbs Sampler) para simular las realizaciones de la distribución posterior conjunta.

La implementación de las técnicas de Monte Carlo vía Cadenas de Markov (MCMC) permiten generar, de manera iterativa, observaciones de distribuciones multivariadas que difícilmente podrían simularse utilizando métodos directos. (Gelman, 1995, Carlin, y Louis, 1996, Spielhalter, 1996 y Lee, 1997).

La idea básica es muy simple: construir una cadena de Markov que sea fácil de simular y cuya distribución de equilibrio corresponda a la distribución a posteriori que nos interesa, garantizando que esta sea homogénea, irreducible y aperiódica.

En el Muestreador de Gibbs definimos:

Sea $Q(\theta^*/\theta)$ una distribución de transición (arbitraria) y

$$\alpha(\theta^*, \theta) = \min \left\{ \frac{p(\theta^*/x)Q(\theta/\theta^*)}{p(\theta/x)Q(\theta^*/\theta)}, 1 \right\}$$

Aquí se simula una cadena de Markov con distribución de equilibrio $p(\theta/x)$ pero en este caso $\alpha(\theta^*, \theta) = 1$ siempre, por lo que el algoritmo es fácil de aplicar ya que la probabilidad de aceptar un candidato no depende de la distribución de transición.

Algoritmo

Dado un valor inicial $\theta^{(0)} = (\theta_1(0), \dots, \theta_k(0))$, se simula una cadena de Markov en la que $\theta(t+1)$ se obtiene a partir de $\theta(t)$ de la siguiente manera:

Se genera una observación $\theta_1^{(t+1)}$ de $p(\theta_1/\theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_k^{(t)}, x)$;

Se genera una observación $\theta_2^{(t+1)}$ de $p(\theta_2/\theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_k^{(t)}, x)$;

⋮

Se genera una observación $\theta_k^{(t+1)}$ de $p(\theta_k/\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}, x)$.

La sucesión $\theta^{(1)}, \theta^{(2)}, \dots$ así obtenida es entonces una realización de una cadena de Markov cuya distribución de transición está dada por:

$$p(\theta^{(t+1)}|\theta^{(t)}) = \prod_{i=1}^k p(\theta_i^{(t+1)}|\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_{i+1}^{(t)}, \dots, \theta_k^{(t)}, \mathbf{x})$$

En general no es fácil determinar en qué momento la(s) cadena(s) ha(n) convergido.

En general, para los algoritmos MCMC no es fácil determinar en qué momento la(s) cadena(s) ha(n) convergido (Gutiérrez, 2003). Un método empírico comúnmente utilizado, consiste en graficar los promedios ergódicos de algunas funciones del parámetro contra el número de iteraciones y elegir el valor de T, a partir del cual las gráficas se estabilizan. Es frecuente en este caso omitir los primeros valores al calcular los promedios ergódicos. La idea de este periodo de calentamiento es permitir que la(s) cadena(s) salga(n) de una primera fase de inestabilidad.

3. MUESTREADOR DE GIBBS USANDO EL WINBUGS Y LA IMPLEMENTACIÓN EN MATLAB

3.1. Planteamiento del problema de pronóstico bioclimático

El Instituto de Meteorología del Ministerio de Ciencia, Tecnología y Medio Ambiente en colaboración con el Ministerio de Salud Pública y la Universidad de La Habana, trabajaron en una investigación bajo la nominación de “Clima y Salud Humana” enmarcado en un problema principal de las Ciencias Naturales y Exactas bajo el título “Perfeccionamiento de los conocimientos sobre el clima de Cuba” donde se trataba de corroborar el hecho de que la salud y capacidad de trabajo del hombre dependen en gran medida de las condiciones climáticas de su lugar de residencia. (Sistachs, 1996).

Para el pronóstico de la enfermedad ASMA, en los niños, se tomaron las frecuencias de ocurrencia diaria de la enfermedad durante 1982. Los datos provenían de los controles de asistencia al Cuerpo de Guardia del Departamento de Estadística de los hospitales Pediátrico de Centro Habana, William Soler y Marianao.

Para el análisis de las variables climáticas se utilizó el meteorograma mensual de la estación de Casablanca en La Habana, que contenía la información de dichas variables. Se tuvo en cuenta además cuatro periodos de tiempo, dos de transición invierno-verano y verano-invierno, el invierno y el verano. Aquí solo ilustraremos el caso de periodo de transición del invierno al verano.

Como plantea Spiegelhalter (1996) para ajustar el modelo usando el Muestreador de Gibbs se deben tener en cuenta cuatro pasos:

1. Valores iniciales para todos los nodos no observados.
2. Distribución condicional completa para cada nodo no observado .
3. Monitorear la salida (esto es, para decidir la longitud total de la corrida un aspecto crucial es el análisis de la convergencia ya que puede atascarse en ciertas regiones cuando la cadena es muy larga, y también si hay muchas modas separadas, la convergencia es lenta).

4. RESUMEN ESTADÍSTICO DE LOS PARÁMETROS DE SALIDA

En nuestro caso seguiremos estos pasos haciendo uso del Winbugs, el cual es un sistema muy utilizado actualmente para análisis estadístico bayesiano con los algoritmos MCMC.

El modelo en nuestro caso sería:

$$\text{logit}(p[i]) = \alpha_0 + \alpha_1 * X1[i] + \alpha_2 * X2[i] + b[i]$$

X1: Tipo de situación sinóptica (TSS)

X2: Tensión de vapor de agua. (TVA).

El Gráfico 1 muestra el modelo de regresión logística a través del enfoque bayesiano utilizando el Winbugs. En el Anexo 1 aparecen los resultados de las cadenas (Traza) y funciones de densidad respectivas de las distribuciones finales de los parámetros. En la Tabla 1 aparecen los resultados de las características numéricas de las distribuciones asociadas a los parámetros del modelo.

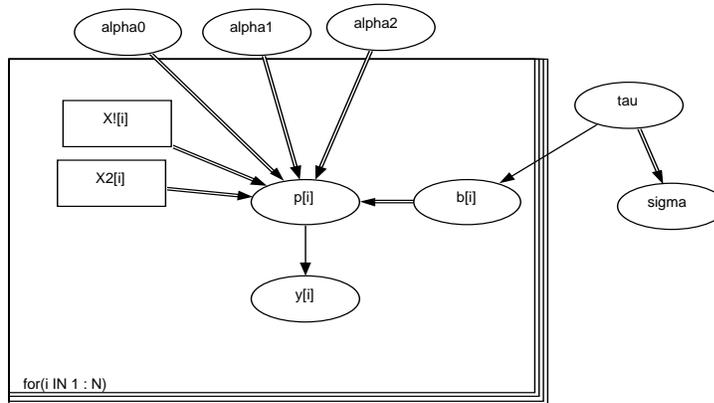


Gráfico 1. Representación grafica del modelo en el WinBUGS

Código de programación del modelo representado gráficamente.

```

model;
{
  for( i in 1 : N ) {
    y[i] ~ dbin(p[i],N)
  }
  for( i in 1 : N ) {
    logit(p[i]) <- alpha0 + alpha1 * X1[i] + alpha2 * X2[i] + b[i]
  }
  for( i in 1 : N ) {
    b[i] ~ dnorm( 0.0,tau)
  }
  tau ~ dgamma(0.001,0.001)
  sigma <- 1 / sqrt(tau)
  alpha0 ~ dnorm( 0.0,1.0E-6)
  alpha1 ~ dnorm( 0.0,1.0E-6)
  alpha2 ~ dnorm( 0.0,1.0E-6)
}

```

Tabla 1. Resumen Estadístico para el Muestreador de Gibbs usando el Winbugs (burn 10000).

| Parámetro | media | DS | Error Monte Carlo | start | sample |
|-----------|-----------|---------|-------------------|-------|--------|
| alpha0 | - 3.376 | 1.234 | 0.08571 | 10000 | 10001 |
| alpha1 | - 0.1558 | 0.4285 | 0.01603 | 10000 | 10001 |
| alpha2 | - 0.07848 | 0.06741 | 0.004661 | 10000 | 10001 |
| tau | 272.5 | 443.7 | 31.11 | 10000 | 10001 |

De aquí podemos concluir al evaluar el modelo en los parámetros, que la probabilidad de una alta frecuencia de ocurrencia de la enfermedad es 0.006, o sea muy pequeña, sin importar el tipo de situación sinóptica y el comportamiento de la tensión de vapor de agua.

Como el Winbugs al ser utilizado resulta una caja negra y es poco flexible pues no permite cambiar con facilidad el modelo pues tiene fijo ciertas distribuciones que no permiten ser modificadas, se decidió implementar el algoritmo haciendo uso del MATLAB.

Los detalles de la implementación aparecen en Mederos y Sistachs. (2003). Para el caso en estudio los resultados obtenidos se muestran en el Anexo 2 (trazas y densidades) y en la Tabla 2 aparece el resumen estadístico con esta implementación.

Tabla 2. Resumen estadístico para el Muestreador de Gibbs usando la implementación en MATLAB. (comenzando en la iteración 500 hasta la iteración 5000)

| Parámetro | media | DS | start | sample |
|-----------|---------|-------|-------|--------|
| alpha0 | - 4.013 | 2.48 | 500 | 5001 |
| alpha1 | 1.0308 | 0.35 | 500 | 5001 |
| alpha2 | 0.166 | 0.09 | 500 | 5001 |
| tau | 171.25 | 33.59 | 500 | 5001 |

De lo que podemos concluir al evaluar el modelo en los parámetros, que la probabilidad de una alta frecuencia de ocurrencia de la enfermedad en dependencia de las covariables es de 0.5836 si hay un tipo de situación sinóptica 1, es decir presencia de anticiclones tanto Atlántico como migratorio y una tensión de vapor de agua elevada. Observe que los resultados del comportamiento del algoritmo con el sistema Winbugs difieren de la implementación en MATLAB.

Dada esta situación se realizó un estudio de simulación con el objetivo de comparar el comportamiento de la estimación de los coeficientes del modelo en estudio, utilizando la implementación en MATLAB del Muestreador de Gibbs bajo el enfoque bayesiano con la obtenida usando el Winbugs. Esta comparación se hace para corroborar que en el caso de dos covariables, la estimación de los coeficientes del modelo se comporta mejor que la que reporta el Wingbugs, cuya manipulación no es cómoda.

5. COMPARACIÓN DE RESULTADOS

Para realizar este estudio se generaron 100 muestras de la variable respuesta binaria para el modelo con dos covariables. El modelo que mejor se ajustó a los datos muestrales fue el que se seleccionó como modelo real y partiendo de él se generaron muestras de la variable dependiente en cada ejemplo calculando en todo los casos el riesgo de la estimación para ver que modelos dan una estimación de los valores de los parámetros más cercanos al valor real.

Para lograr lo anterior se parte del supuesto de que se conoce en el modelo probabilístico o sea el valor de $\pi(x)$ teniendo fijos los coeficientes y las covariables, lo cual permite generar las muestras de la variable dependiente con distribución Bernoulli ($\pi(x)$) y entonces se utiliza la función que genera números aleatorios con esta distribución en el MATLAB.

Para realizar la comparación se calculó la estimación del riesgo en cada muestra generada, es decir la suma de los errores cuadráticos medios de la estimación de los parámetros $\left(\sum_{i=1}^n (\beta - \hat{\beta}_i)^2 \right)$. La estimación de los

parámetros que minimiza la función de pérdida (el riesgo calculado) coincide con los valores esperados de los parámetros del modelo en estudio para obtener el riesgo en cada procedimiento o enfoque estudiado se estimó la función de pérdida o riesgo como la suma de los riesgos en cada muestra generada, además se realizó una comparación de los valores obtenidos del riesgo en cada muestra para determinar la significación de la diferencia entre los procedimientos estudiados.

En la Tabla 3 aparecen los resultados de los cálculos de los riesgos de estimación para cada muestra, así como el cálculo del riesgo para los procedimientos o enfoques, y analizando los resultados podemos afirmar que el Gibbs implementado en el MATLAB tiene el menor riesgo de estimación. Además se confeccionaron los gráficos de caja y bigotes para analizar el comportamiento de los riesgos según el procedimiento empleado (ver Gráfico 2).

Tabla 3. Resultados de los cálculos de los riesgos de estimación en las muestras simuladas para dos covariables.

| muestra | clásico | Gibbs Matlab | Gibbs Winbugs | muestra | clásico | Gibbs Matlab | Gibbs Winbugs | muestra | clásico | Gibbs Matlab | Gibbs Winbugs |
|---------|---------|--------------|---------------|---------|---------|--------------|---------------|---------|---------|--------------|---------------|
| 1 | 69.51 | 6.67 | 27.98 | 35 | 7.30 | 12.40 | 5.52 | 69 | 16.57 | 1.62 | 3.42 |
| 2 | 106.89 | 63.90 | 27.31 | 36 | 2.35 | 3.62 | 50.44 | 70 | 96.27 | 19.24 | 20.98 |
| 3 | 33.79 | 3.12 | 1.90 | 37 | 0.26 | 15.48 | 22.04 | 71 | 0.10 | 0.08 | 16.03 |
| 4 | 6.83 | 9.20 | 82.78 | 38 | 3.09 | 1.08 | 32.93 | 72 | 0.14 | 2.25 | 16.46 |
| 5 | 1.23 | 3.42 | 24.34 | 39 | 129.29 | 7.72 | 41.09 | 73 | 131.65 | 12.93 | 70.57 |
| 6 | 8.82 | 0.38 | 22.84 | 40 | 1.16 | 2.19 | 6.52 | 74 | 34.07 | 14.42 | 5.61 |
| 7 | 86.19 | 2.71 | 25.57 | 41 | 8.48 | 6.79 | 0.90 | 75 | 91.12 | 32.18 | 27.90 |
| 8 | 0.75 | 1.75 | 10.27 | 42 | 85.60 | 9.85 | 42.77 | 76 | 1.94 | 1.73 | 12.97 |
| 9 | 76.75 | 1.85 | 30.38 | 43 | 30.40 | 1.81 | 13.04 | 77 | 7.16 | 0.65 | 5.70 |
| 10 | 1.96 | 2.38 | 40.10 | 44 | 7.80 | 1.64 | 7.30 | 78 | 12.04 | 17.29 | 2.01 |
| 11 | 33.04 | 5.40 | 1.56 | 45 | 2.36 | 5.59 | 4.78 | 79 | 16.11 | 0.22 | 89.45 |
| 12 | 15.49 | 22.77 | 5.27 | 46 | 11.92 | 14.23 | 14.92 | 80 | 1.70 | 0.67 | 8.88 |
| 13 | 9.10 | 9.34 | 2.69 | 47 | 2.69 | 0.17 | 10.54 | 81 | 33.65 | 1.33 | 4.12 |
| 14 | 27.23 | 1.35 | 7.42 | 48 | 84.17 | 26.79 | 32.57 | 82 | 19.27 | 1.91 | 6.27 |
| 15 | 48.79 | 32.75 | 15.41 | 49 | 20.03 | 3.85 | 32.57 | 83 | 7.12 | 4.07 | 4.35 |
| 16 | 2.82 | 10.47 | 9.43 | 50 | 52.24 | 19.24 | 19.06 | 84 | 6.23 | 2.05 | 11.66 |
| 17 | 72.91 | 0.34 | 5.57 | 51 | 96.27 | 12.92 | 20.98 | 85 | 1.35 | 7.02 | 101.34 |
| 18 | 26.08 | 3.43 | 1.08 | 52 | 131.65 | 47.02 | 70.57 | 86 | 14.65 | 1.86 | 2.90 |
| 19 | 3.72 | 2.76 | 27.02 | 53 | 0.32 | 1.73 | 7.02 | 87 | 2.85 | 1.88 | 4.15 |
| 20 | 0.89 | 7.73 | 18.52 | 54 | 1.94 | 16.82 | 12.97 | 88 | 82.80 | 5.15 | 25.38 |
| 21 | 168.01 | 19.72 | 25.38 | 55 | 13.22 | 14.10 | 6.83 | 89 | 1.88 | 9.16 | 19.67 |
| 22 | 4.22 | 8.44 | 19.67 | 56 | 7.30 | 12.40 | 5.52 | 90 | 46.54 | 6.28 | 8.84 |
| 23 | 2.57 | 23.68 | 8.84 | 57 | 0.28 | 3.64 | 13.99 | 91 | 39.80 | 33.38 | 13.26 |
| 24 | 0.00 | 2.33 | 13.26 | 58 | 3.09 | 1.88 | 32.93 | 92 | 15.39 | 11.55 | 17.57 |
| 25 | 4.08 | 3.58 | 17.57 | 59 | 129.29 | 7.72 | 41.09 | 93 | 0.51 | 2.19 | 11.79 |
| 26 | 43.20 | 37.05 | 11.79 | 60 | 40.23 | 3.24 | 20.93 | 94 | 0.39 | 2.26 | 21.13 |
| 27 | 88.69 | 10.48 | 21.13 | 61 | 7.80 | 1.81 | 7.30 | 95 | 56.20 | 0.31 | 47.09 |
| 28 | 9.88 | 3.17 | 47.09 | 62 | 2.36 | 1.64 | 4.78 | 96 | | 39.40 | 13.20 |
| 29 | 6.63 | 4.58 | 13.20 | 63 | 0.85 | 10.83 | 40.82 | 97 | 59.23 | 1.73 | 6.69 |
| 30 | 0.36 | 6.42 | 6.69 | 64 | 11.92 | 5.60 | 14.92 | 98 | 1.14 | 4.89 | 20.80 |
| 31 | 0.47 | 3.86 | 20.80 | 65 | 0.61 | 14.24 | 12.67 | 99 | 0.53 | 3.75 | 16.29 |
| 32 | 0.26 | 11.05 | 16.29 | 66 | 84.17 | 44.53 | 32.57 | 100 | 127.35 | 60.50 | 17.52 |
| 33 | 0.18 | 0.88 | 17.52 | 67 | 2.45 | 2.79 | 14.79 | | | | |
| 34 | 63.04 | 14.11 | 20.41 | 68 | 1.40 | 4.39 | 10.67 | | | | |
| | | | | | | | | Riesgo | 2934.41 | 984.75 | 2009.38 |

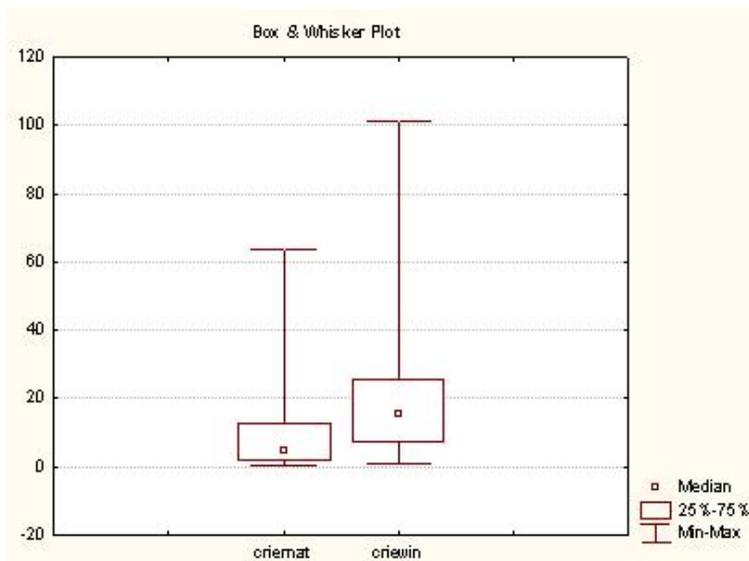


Gráfico 2. Comparación del riesgo entre el Gibbs MATLAB y el Gibbs Winbugs.

6. CONCLUSIONES

En el cálculo de la distribución posterior una dificultad que se presenta es la imposibilidad en muchos caso de encontrar la distribución posterior por vía analítica directa, pero con la aplicación de los métodos numéricos MCMC, en particular el Muestreador de Gibbs, se obtienen con relativa facilidad la densidad empírica posterior.

Además la implementación hecha en MATLAB demostró ser buena ya que la estimación obtenida fue la que menor riesgo tuvo lo que habla a favor del algoritmo empleado. Esta implementación nos permite también tener en cuenta este algoritmo en otros programas de computación en los que sea necesario calcular las estimaciones bayesianas de los coeficientes del modelo de regresión logística para su aplicación de forma regular en otros ejemplos prácticos.

REFERENCIAS

- BERNARDO, J. M. (2003) "Bayesian Statistics". Encyclopedia of Live Support Systems (EOLSS), UNESCO.
- CARLIN, B.P. and T.A. LOUIS (1996): **Bayes and Empirical Bayes Methods for Data Analysis**. Chapman & Hall.
- GELMAN, A.; J. CARLIN; H. STERN and D.B. RUBIN (1995): **Bayesian Data Analysis**. New York: Chapman & Hall.
- GUTIÉRREZ, E. (2003): **Métodos Computacionales en la Inferencia Bayesiana**
www.dpye.iimas.unam/eduardo/MCB/MCB.html
- HOSMER, D.W. and S. LEMESHOW (1989): **Applied Logistic Regresión**. John Wiley and Sons, Inc, New York.
- MEDEROS, M.V. y V. SISTACHS (2003): Programas en MATLAB para el muestreo de Gibbs en el modelo de regresión logística binario. Ponencia presentada en COMPUMAT 2003(Santi Spíritus 17-21 nov).
- LEE, P. M., (1997): **Bayesian Statistics an Introduction**. Second Edition. London. Edward Arnold. Pp 249-270.
- SISTACHS, V. (1996): Pronósticos Bioclimáticos a través de técnicas estadísticas con datos categóricos. Tesis de maestría. Universidad de la Habana, Cuba.
- SILVA, L.C. (1994): **Excursión a la Regresión Logística en la Ciencia de la Salud**. Ediciones DIAS DE SANTOS, S.A., ESPAÑA.
- SPIEGELHALTER, D.J.; N.G. BEST; W.R. GILKS and H. INSKIP (1996): **Hepatitis B: a case study in MCMC methods**, 21-43 in Gilks, W., Richardson, S. and Spiegelhalter, D. (ed) Markov Chain Monte Carlo in Practice. Chapman & Hall. London.

Anexo 2. Implementación en Matlab del Muestreador de Gibbs.

