Seminario de Estadística 1
PROYECTO FINAL
ANALISIS BAYESIANO DE DATOS

Soriano Flores Antonio

Noviembre 2019

1. La tabla FootballLeague.csv contiene los datos sobre el desempeño de los equipos de la liga nacional de fútbol de E.U.A. durante 1976.

   a) Ajuste un modelo lineal (con todas las covariables) bajo el supuesto de normalidad de la variable respuesta y encuentre la distribución final de los parámetros, para ello utilice JAGS y como distribución inicial de los parámetros suponga que:

   $$\beta_i \sim N\left(\beta_i|0, 0.001\right) \qquad \tau \sim Gamma\left(\tau|0.01, 0.01\right)$$

   con β_i independiente de τ

   b) Suponga que se desea ajustar un modelo mas sencillo, para ello lleve a cabo pruebas individuales de los parámetros para ir eliminando en cada paso una covariable hasta que todas sean significativas. (Con intervalos de 95 % de probabilidad) (Selección de modelo Backward)

   c) Con el modelo seleccionado en el paso anterior calcule el CPO_i de cada una de las observaciones y despúes obtenga el LPML (Geisser and Eddy, 1979) (Logarithm of the Pseudomarginal Likelihood) definido como:

   $$LPML = \sum_{i=1}^{n} \log CPO_i$$

   d) Repita los incisos a), b) y c) pero ahora ajustando un modelo lineal poisson, es decir, bajo el supuesto de que la variable respuesta sigue un modelo poisson. Para este caso nuevamente utilice como distirbuciones iniciales de los parámetros:

   $$\beta_i \sim N\left(\beta_i|0, 0.001\right)$$

   e) Bajo el criterio de los CPO's, ¿Con cual modelo ajustaria estos datos?

REGRESION BETA

2. La regresión beta pertenece a la familia de los modelos lineales generalizados y supone lo siguiente:

   - $(y_1, \ldots, y_n)$ siguen una distribución $Beta(\alpha, \delta_i)$
   - $\mu_i = \mathbb{E}(y_i \mid x_i) = \frac{\alpha}{\alpha+\delta_i} = \mu_i \in (0, 1)\,, \Rightarrow \delta_i = \frac{\alpha(1-\mu_i)}{\mu_i}$
   - $\eta_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}$ la **componente sistemática** (predictor lineal)
   - $g(\mu_i) = logit\,(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$.

Obs: En este caso se modela la media μ_i de la variable y_i como

$$\mu_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}$$

Observe que en este caso α no depende del componente lineal sino que pertenece fijo. (Similar al caso de la regresión lineal normal donde $\tau = \sigma^{-2}$ es fijo.)

Note entonces que el modelo asume lo siguiente:

$$y_i \sim Beta\,(y_i|\alpha, \delta_i) = Beta\left(y_i \middle| \alpha, \frac{\alpha(1-\mu_i)}{\mu_i}\right)$$

con $logit\,(\mu_i) = \eta_i$.
Asumiendo las siguientes iniciales en los parámetros del predictor lineal y de α

$$\beta_i \sim N(\beta_i|0, 0.001); \quad i \in \{0, 1, \ldots, k\}; \qquad \alpha \sim Ga(\alpha|0.01, 0.01)$$

a) Escriba el codigo en JAGS para ajustar este modelo asumiento que X es la matriz diseño y y el vector de variables respuesta observadas de tamaño n

b) Suponga que usted trabaja en el area de riesgo de una aseguradora y pretende modelar el monto reclamado de cada poliza de seguro de autos con ayuda de las siguientes variables auxilares.

- x_1 := Edad del dueño del auto
- x_2 := Ingresos mensuales del dueño del auto
- x_3 := Experiencia de conducir (en años)
- x_4 := Numero de siniestros que ha tenido.

El modelo que se propone es el siguiente:

$$y_i = z_i * M_i$$

Donde y_i es el monto reclamado de la poilza i, $z_i \sim Beta\,(\alpha, \delta_i)$ y M_i es el valor del auto que se asegura.

Para ajustar el modelo, la aseguradora cuenta con 20 observaciones de 20 siniestros reportados con el monto pagado en cada uno de ellos. La tabla BASE_SINIESTROS.xlsx contiene la información correspondiente a los 20 siniestros.

Ajuste un modelo de regresion lineal beta para modelar a z_i en función de las covariables x_1, x_2, x_3, x_4 (Ajuste el modelo sin intercepto). Con ayuda de JAGS y evaluando la correcta convergencia de las cadenas, grafique la distribución final de los parámetros de este modelo $(\alpha, \beta_1, \beta_2, \beta_3, \beta_4)$. Y encuentre los intervalos de probabilidad al 90 % para cada parámetro.

Finalmente con su modelo ajustado, suponga que llegan 2 nuevos asegurados con los siguientes valores de covariables:

Observación	x_1 (edad)	x_2 (Ingresos)	x_3 Experiencia	x_4 Numero Siniestros	M (valor del auto)
21	37	50000	15	0	50000
22	22	10000	1	7	120000

Con el modelo ajustado, realice un programa que simule observaciones de la variable y_{21} , y_{22} y grafique el histograma correspondiente que indique el monto a reclamar

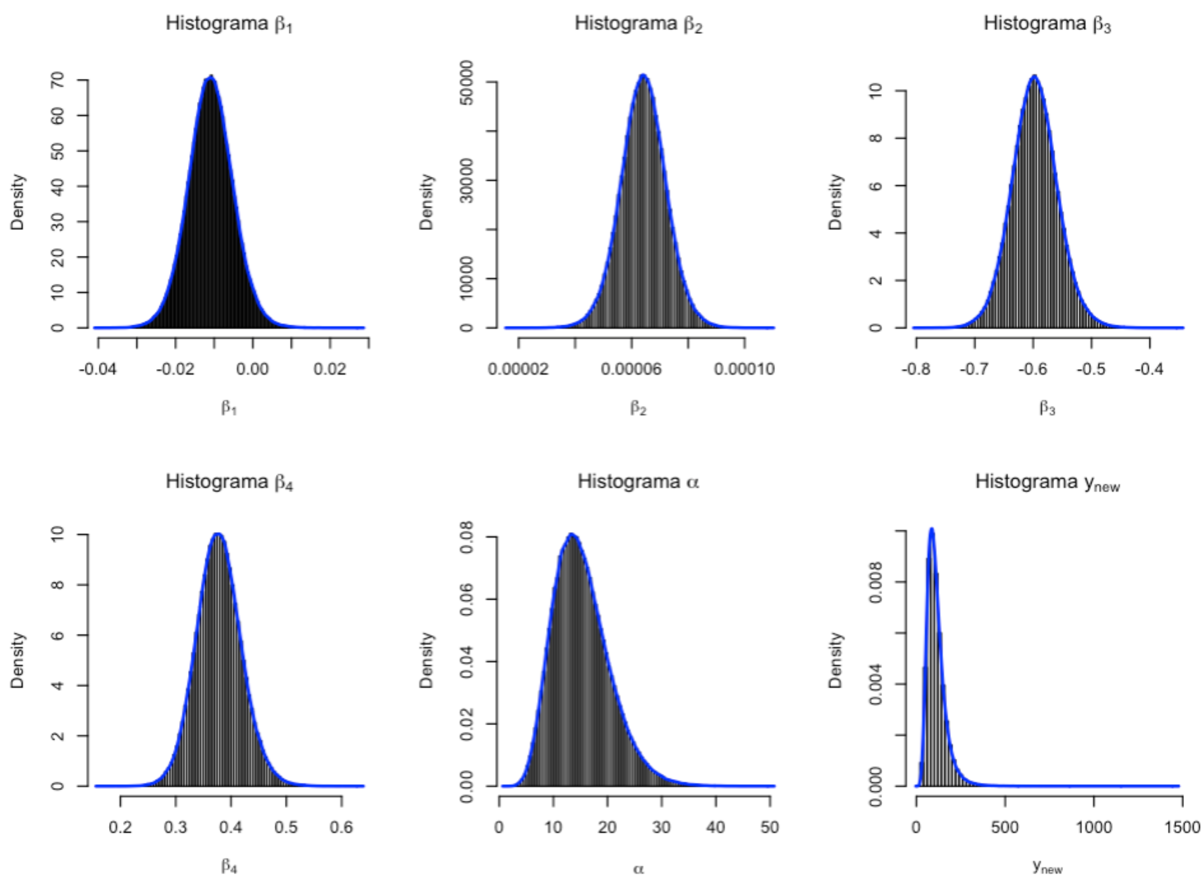

Figura 1: Simulacion de la distribución final de los parámetros y de la primera predicción