

Tarea4

Alberto, Ivan, Sara, Valeria

Tarea 4

```
library(ggplot2)
library(tidyverse)
library(patchwork)
library(kableExtra)
```

1. Spiegelhalter et al. (1995) analiza la mortalidad del escarabajo del trigo en la siguiente tabla, usando BUGS.

Dosis_wi	muertos_yi	expuestos_ni
1.6907	6	59
1.7242	13	60
1.7552	18	62
1.7842	28	56
1.8113	52	63
1.8369	53	59
1.8610	61	62
1.8839	60	60

Estos autores usaron una parametrización usual en dos parámetros de la forma $p_i \equiv P(\text{muerte}|w_i)$, pero comparan tres funciones ligas diferentes:

$$\text{logit: } p_i = \frac{\exp(\alpha + \beta z_i)}{1 + \exp(\alpha + \beta z_i)}$$

$$\text{probit: } p_i = \Phi(\alpha + \beta z_i)$$

$$\text{complementario log-log: } p_i = 1 - \exp[-\exp(\alpha + \beta z_i)]$$

en donde se usa la covariada centrada $z_i = w_i - \bar{w}$ para reducir la correlación entre la ordenada α y la pendiente β . En OpenBUGS el código para implementar este modelo es el que sigue:

```
modelo1 = "model{
for (i in 1:k){
y[i] ~ dbin(p[i],n[i])
logit(p[i]) <- alpha + beta*(w[i]-mean(w[]))
# probit(p[i]) <- alpha + beta*(w[i]-mean(w[]))
# cloglog(p[i]) <- alpha + beta*(w[i]-mean(w[]))
} #fin del loop i
alpha ~ dnorm(0.0,1.0e-3)
dbeta ~ dnorm(0.0,1.0e-3)
} #fin del código"
```

Lo que sigue al símbolo `#` es un comentario, así que esta versión corresponde al modelo logit. También `dbin` denota la distribución binomial y `dnorm` denota la distribución normal, donde el segundo argumento denota la precisión, no la varianza (entonces las iniciales normales para α y β tienen precisión 0.001, que son aproximadamente iniciales planas (no informativas). Hacer el análisis en OpenBUGS.

2. Consideren las siguientes dos distribuciones condicionales completas, analizadas en el artículo de Casella y George (1992) que les incluí como lectura:

$$f(x|y) \propto ye^{-yx}, 0 < x < B < \infty$$

$$f(y|x) \propto xe^{-xy}, 0 < y < B < \infty$$

- Obtener un estimado de la distribución marginal de X cuando $B = 10$ usando el Gibbs sampler.
- Ahora supongan que $B = \infty$ así que las distribuciones condicionales completas son ahora las ordinarias distribuciones exponenciales no truncadas. Mostrar analíticamente que $f_x(t) = 1/t$ es una solución a la ecuación integral en este caso:

$$f_x(x) = \int \left[\int f_{x|y}(x|y) f_{y|t}(y|t) dy \right] f_x(t) dt$$

¿El Gibbs sampler convergerá a esta solución?

3. Supongan que una variable aleatoria y se distribuye de acuerdo a la densidad poli-Cauchy:

$$g(y) = \prod_{i=1}^n \frac{1}{\pi(1 + (y - a_i)^2)}$$

donde $a = (a_1, \dots, a_n)$ es un vector de parámetros. Supongan que $n = 6$ y $a = (1, 2, 2, 6, 7, 8)$.

- Escriban una función que calcule la log-densidad de y .
- Escriban una función que tome una muestra de tamaño 10,000 de la densidad de y , usando Metropolis-Hastings con función propuesta una caminata aleatoria con desviación estandar C . Investiguen el efecto de la elección de C en la tasa de aceptación, y la mezcla de la cadena en la densidad.
- Usando la muestra simulada de una “buena” elección de C , aproximar la probabilidad $P(6 < Y < 8)$.

4. Supongan que el vector (X, Y) tiene función de distribución conjunta:

$$f(x, y) = \frac{x^{a+y-1} e^{-(1+b)x} b^a}{y! \Gamma(a)}, x > 0, y = 0, 1, 2, \dots$$

y deseamos simular de la densidad conjunta.

- Mostrar que la densidad condicional $f(x|y)$ es una Gamma e identificar los parámetros.

Primero obtenemos $f(y)$

$$\begin{aligned} f(y) &= \int_0^\infty \frac{x^{a+y-1} e^{-(1+b)x} b^a}{y! \Gamma(a)} dx \\ &= \frac{b^a}{y! \Gamma(a)} \int_0^\infty x^{a+y-1} e^{-(1+b)x} \\ &= \frac{b^a}{y! \Gamma(a)} \left(\frac{\Gamma(a+y)}{(1+b)^{a+y}} \right) \end{aligned}$$

Luego aplicamos la definición de densidad condicional:

$$\begin{aligned} f(x|y) &= \frac{f(x, y)}{f(y)} \\ &= \frac{\frac{x^{a+y-1} e^{-(1+b)x} b^a}{y! \Gamma(a)}}{\frac{b^a}{y! \Gamma(a)} \left(\frac{\Gamma(a+y)}{(1+b)^{a+y}} \right)} \\ &= \frac{x^{a+y-1} e^{-(1+b)x}}{\Gamma(a+y)} (1+b)^{a+y} \end{aligned}$$

Con lo cual podemos ver que se trata de una $\mathcal{G}(a+y, 1+b)$

- Mostrar que la densidad condicional $f(y|x)$ es Poisson.

De nuevo obtenemos como primer paso la marginal $f(x)$

$$\begin{aligned}
 f(x) &= \sum_{y=0}^{\infty} \frac{x^{a+y-1} e^{-(1+b)x} b^a}{y! \Gamma(a)} \\
 &= \frac{e^{-(1+b)x} b^a}{\Gamma(a)} \sum_{y=0}^{\infty} \frac{x^{a+y-1}}{y!} \\
 &= \frac{e^{-(1+b)x} b^a}{\Gamma(a)} \sum_{y=0}^{\infty} \frac{x^y x^{a-1}}{y!} \\
 &= \frac{e^{-(1+b)x} b^a x^{a-1}}{\Gamma(a)} \sum_{y=0}^{\infty} \frac{x^y}{y!} \\
 &= \frac{e^{-(1+b)x} b^a x^{a-1}}{\Gamma(a)} e^x
 \end{aligned}$$

Y usamos la definición de densidad condicional:

$$\begin{aligned}
 f(x|y) &= \frac{f(x, y)}{f(y)} \\
 &= \frac{\frac{x^{a+y-1} e^{-(1+b)x} b^a}{y! \Gamma(a)}}{\frac{e^{-(1+b)x} b^a x^{a-1}}{\Gamma(a)} e^x} \\
 &= \frac{x^y e^x}{y!}
 \end{aligned}$$

Con lo cual vemos que es $\mathcal{P}(\lambda = x, k = y)$

- Escriban una función para implementar el Gibbs sampler cuando las constantes son dadas con valores $a = 1$ y $b = 1$.

```

a <- 1
b <- 1
nsim <- 1000

X_dado_Y <- Y_dado_X <- array(0, dim = c(nsim, 1))

X_dado_Y[1] <- rgamma(1, shape = a, rate = 1 + b)
Y_dado_X[1] <- rpois(1, lambda = X_dado_Y[1])

ejercicio_4 <- function(n, X_dado_Y, Y_dado_X) {
  for (i in 2:n) {
    X_dado_Y[i] <- rgamma(1, shape = a + Y_dado_X[i - 1], rate = 1 + b)
  }
}

```

```

    Y_dado_X[i] <- rpois(1, lambda = X_dado_Y[i])
  }

  return(list(X_dado_Y = X_dado_Y, Y_dado_X = Y_dado_X))
}

```

- Con su función, escriban 1000 ciclos del Gibbs sampler y de la salida, hacer los histogramas y estimar $E(Y)$.

```

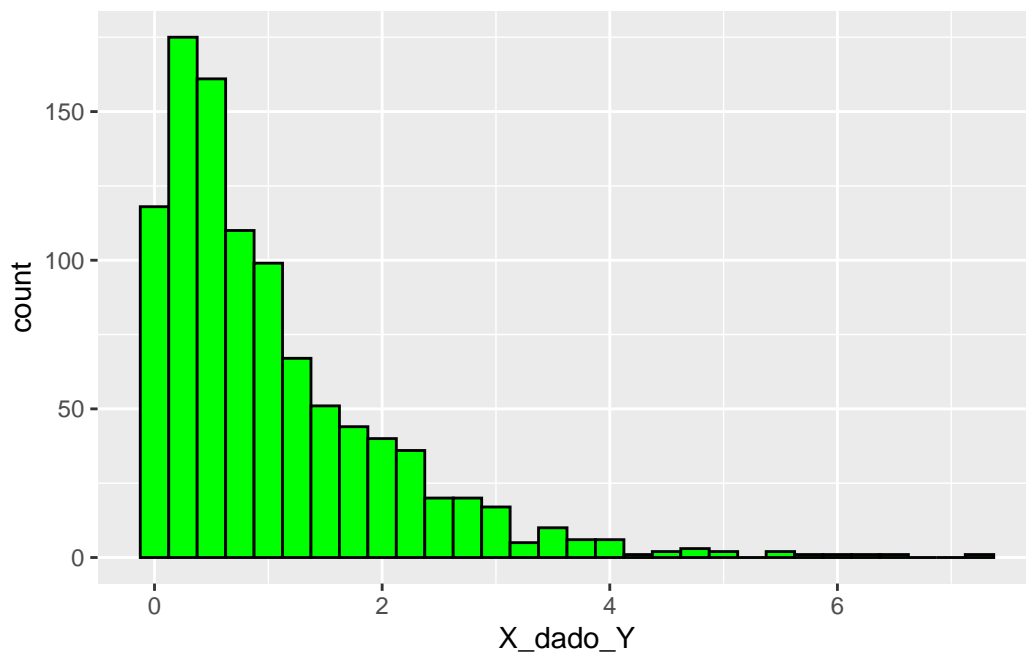
result <- ejercicio_4(nsim, X_dado_Y, Y_dado_X)
# X_dado_Y <- result$X_dado_Y
# Y_dado_X <- result$Y_dado_X

result<- data.frame(result)

ggplot(result, aes(x=X_dado_Y)) +
  geom_histogram(color="black", fill="green")

```

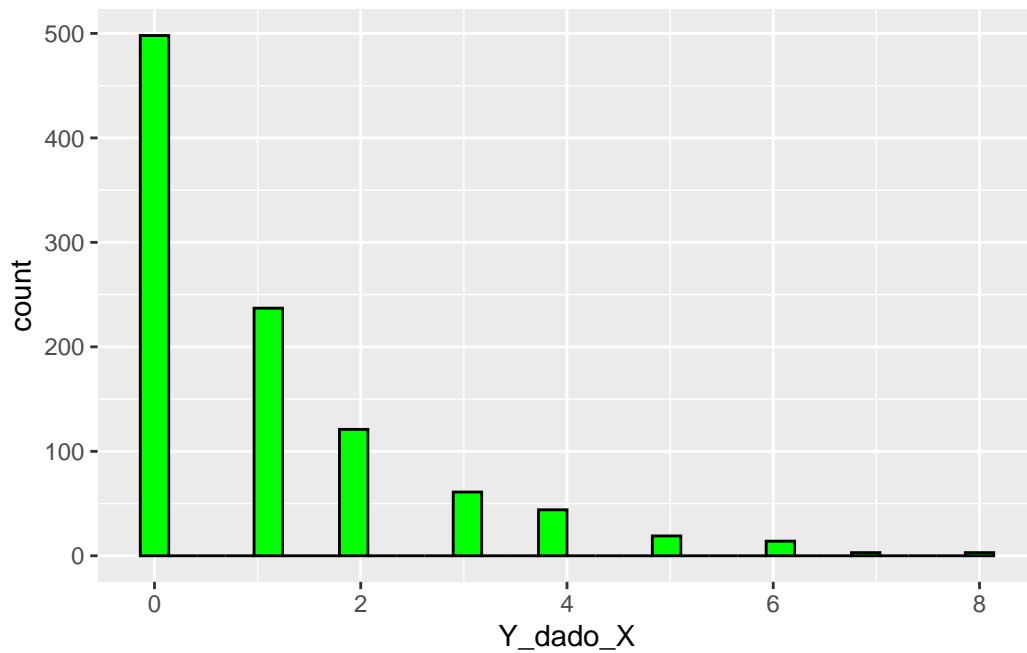
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



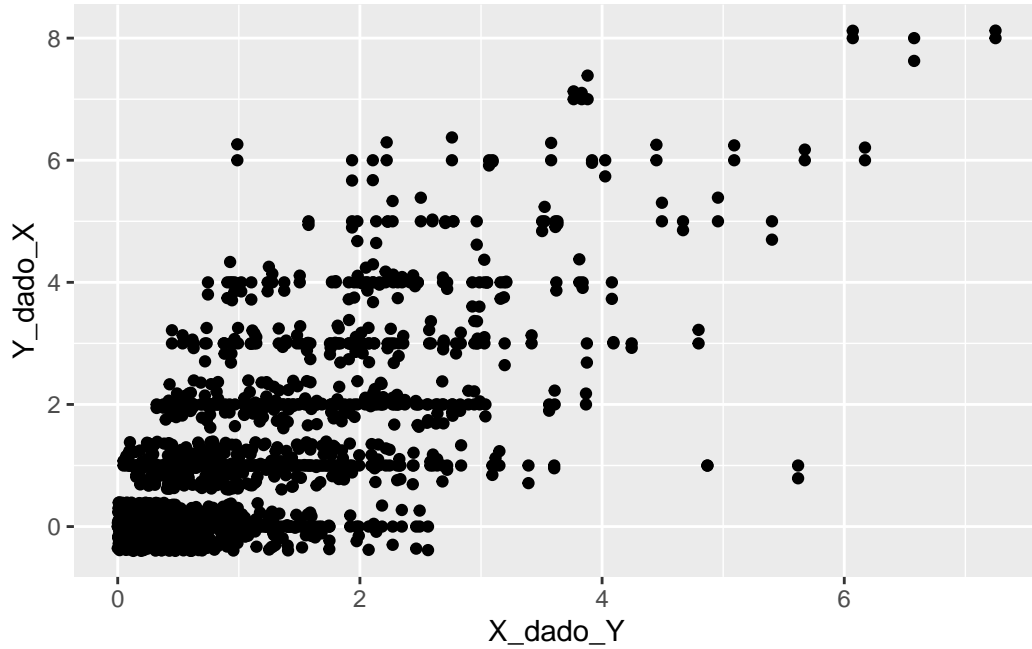
```
#hist(Y_dado_X, probability = T, main = "Y", ylab = "densidad")
```

```
ggplot(result, aes(x=Y_dado_X)) +  
  geom_histogram(color="black", fill="green")
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.



```
ggplot(result, aes(X_dado_Y, Y_dado_X))+ geom_point()+ geom_jitter()
```



5. Supongan que se tiene una matriz de 4×4 de variables aleatorias Bernoulli, y la denotamos por $[X_{ij}]$, y sea $N(X)$ el número total de éxitos en X (la suma de X) y $D(X)$ es el total de vecinos de dichos éxitos (horizontales o verticales) que difieren. Por ejemplo,

X	$N(X)$	$D(X)$
$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	1	2
$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	3	5

Noten que si se escriben los elementos de X en un vector V , entonces existe una matriz M de 24×16 tal que $D(X)$ es la suma de los valores absolutos de MV . El 24 surge porque hay 24 pares de vecinos a revisar. Supongan que $\pi(X)$, la distribución de X , es proporcional a

$$\pi(X) \propto p^{N(X)}(1-p)^{16-N(X)} \exp(-\lambda D(X))$$

Si $\lambda = 0$, las variables son iid Bernoulli (p).

Usar el método de Metropolis-Hastings usando los siguientes kernels de transición. Hay 2^{16} posibles estados, uno por cada posible valor de X .

- a) Sea q_1 tal que cada transición es igualmente plausible con probabilidad $1/2^{16}$. Esto es, el siguiente estado candidato para X es simplemente un vector de 16 iid Bernoulli (p).
- b) Sea q_2 tal que se elige una de las 16 entradas en X con probabilidad $1/16$, y luego se determina el valor de la celda a ser 0 o 1 con probabilidad 0.5 en cada caso. Entonces sólo un elemento de X puede cambiar en cada transición a lo más.

Ambas q 's son simétricas, irreducibles y tienen diagonales positivas. La primera se mueve más rápido que la segunda.

Estamos interesados en la probabilidad de que todos los elementos de la diagonal sean 1. Usen las dos q 's para estimar la probabilidad de 1's en la diagonal para $\lambda = 0, 1, 3$ y $p = 0.5, 0.8$. Esto se puede hacer calculando la fracción acumulada de veces que la cadena tiene 1's sobre la diagonal conforme se muestrea de la cadena. Comparar los resultados de las 2 cadenas. Tienen el mismo valor asintótico, pero ¿una cadena llega más rápido que la otra? ¿Alguna cadena muestra más autocorrelación que la otra (por ejemplo, estimados de la probabilidad en 100 pasos sucesivos muestran más correlación para una cadena que para la otra?).

En este problema, también determinen cuántas simulaciones hay que hacer, para desechar el periodo de calentamiento.

6. Considera los siguientes números:

0.4, 0.01, 0.2, 0.1, 2.1, 0.1, 0.9, 2.4, 0.1, 0.2

Usen la distribución exponencial $Y_i \sim \exp(\theta)$ para modelar estos datos y asignen una inicial sobre $\log \theta$.

- a) Definan los datos en WinBUGS (u OpenBUGS). Usen $\theta = 1$ como valor inicial.
- b) Escriban el código para el modelo.
- c) Compilen el modelo y obtengan una muestra de 1000 iteraciones después de descartar las 500 iteraciones iniciales como burnin.
- d) Monitoreen la convergencia gráficamente usando gráficas 'trace' y de autocorrelación.
- e) Obtengan estadísticas sumarias posteriores y densidades para θ , $1/\theta$ y $\log \theta$
- f) Con los datos el primer inciso, ahora usen las distribuciones (i) gamma (ii) log-normal para modelar los datos, así como (iii) normal para los logaritmos de los valores originales. En todos los modelos hagan un ejercicio similar al de los numerales previos, y comparen los resultados obtenidos bajo todos los modelos. Hagan todos los supuestos que tengan que hacer.