



**Universidad Autónoma de Guerrero**

Facultad de Matemáticas

---

Maestría en Matemáticas Aplicadas

**Modelación Bayesiana de una  
variable ordinal aleatorizada**

**T E S I S**

PARA OBTENER EL GRADO EN:

**Maestría en Matemáticas Aplicadas**

PRESENTA:

**Lic. Deisy Lozano Salado**

DIRECTORES DE TESIS:

**Dr. Flaviano Godínez Jaimes**

**Dr. Ramón Reyes Carreto**

Junio de 2018.

Dedicada a cada integrante de mi familia.

---

## Agradecimientos

---

Al Coordinador de la Maestría en Matemáticas Aplicadas y a cada uno de los profesores de la institución, por el conocimiento aportado en cada una de las clases que recibí.

A los directores del presente trabajo, por la dedicación, la paciencia y el apoyo brindado durante todo el tiempo que fue necesario para culminarlo.

A cada uno de los profesores que dedicaron su tiempo para la revisión del presente trabajo.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo económico durante el periodo de estudios de posgrado.

---

## Resumen

---

Los temas sensitivos se refieren a acciones delictivas o comportamientos no aceptados por la sociedad. Ejemplos de estos temas son el abuso en el consumo de drogas, preferencias sexuales, enfermedades sexuales, infidelidad, abortos ilegales, evasión de impuestos, fraudes, corrupción, plagio en exámenes, etc. En encuestas de temas sensitivos, es común que el encuestado se niegue a responder la pregunta o responda de manera no veraz. Esta forma de responder provoca que las inferencias sean incorrectas debido a la presencia de sesgo de no respuesta o sesgo a respuestas no veraces. Warner (1965) fue el primero en proponer un método de respuesta aleatoria que protege la privacidad del encuestado y disminuye el sesgo de no respuesta.

Los diseños de respuesta aleatoria se han utilizado para estimar la proporción, el total o la media de características sensitivas en la población, mediante variables aleatorizadas binarias, multinomiales o cuantitativas. También se ha modelado el efecto variables independientes en una variable binaria aleatorizada, como en Blair *et al.* (2015) donde se usó regresión logística. No se encontraron trabajos sobre modelación Bayesiana de variables aleatorizadas ordinales, por lo que en este trabajo se usa el enfoque Bayesiano, para modelar el efecto de variables independientes en una variable aleatorizada ordinal, usando regresión probit ordinal. Se consideraron cuatro distribuciones iniciales; la Normal, la Doble Exponencial, la *t* de Student y la Cauchy. Las distribuciones posteriores se obtuvieron mediante MCMC. Los estimadores se compararon en función del Sesgo, Error Cuadrático Medio, Longitud y Cubrimiento de intervalos creíbles. El mejor estimador Bayesiano fue el que usa la distribución inicial Doble Exponencial.

---

## ABSTRACT

---

Sensitive issues refer to criminal actions or behaviors not accepted by society. Examples of these issues are abuse in drug use, sexual preferences, sexual diseases, infidelity, illegal abortions, tax evasion, fraud, corruption, plagiarism in examinations, etc. In a survey on sensitive issues, it is common for the respondent to refuse to answer the question or answer in an untruthful way. This way of responding causes inferences to be incorrect due to the presence of non-response bias or bias to untruthful responses. Warner (1965) was the first to propose a randomized response method that protects the privacy of the respondent and decreases the non-response bias.

Randomized response designs have been used to estimate the proportion, total or average of sensitive characteristics in the population, using binary, multinomial or quantitative randomized variables. The effect of independent variables on a randomized binary variable has also been modeled, as in Blair *et al.* (2015) where logistic regression was used. We did not find works on Bayesian or frequentist modeling of ordinal randomized variables, so in this work the Bayesian approach is used to model the effect of independent variables in an ordinal randomized variable using ordinal probit regression. Four prior distributions were considered; Normal, Double Exponential, Student's t and Cauchy. Posterior distributions were obtained by MCMC. The estimators were compared according to the Bias, Mean Squared Error, Length and coverage of credible intervals. The best Bayesian estimator was the one that uses the prior Double Exponential distribution.

---

# Índice general

---

<b>1. Introducción</b>	<b>2</b>
1.1. Algunos diseños de respuesta aleatoria . . . . .	8
1.1.1. Modelo Warner . . . . .	8
1.1.2. Diseño de pregunta no relacionada . . . . .	10
1.1.3. Diseño de respuesta forzada . . . . .	12
1.1.4. Diseño de respuesta disfrazada . . . . .	13
1.1.5. Diseño de Mangat y Singh (1990) . . . . .	15
1.1.6. Diseño de Mangat (1994) . . . . .	17
<b>2. Teoría Bayesiana</b>	<b>18</b>
2.1. Inferencia Bayesiana . . . . .	19
2.2. Cadenas de Markov Monte Carlo . . . . .	20
2.2.1. Metrópolis-Hastings . . . . .	21
2.2.2. Muestreador de Gibbs . . . . .	22
2.2.3. Diagnósticos de convergencia de la cadena . . . . .	23
2.2.4. Implementación de métodos MCMC en JAGS . . . . .	26
2.3. Estimación de parámetros . . . . .	28
2.3.1. Estimadores puntuales . . . . .	28
2.3.2. Intervalos creíbles . . . . .	31
2.3.3. Criterios para evaluar estimadores . . . . .	32
<b>3. Metodología</b>	<b>34</b>
3.1. Diseño de respuesta forzada para variables con más de dos categorías . . . . .	34
3.2. Modelo de regresión probit ordinal . . . . .	37

3.2.1. Modelo de regresión probit ordinal aleatorizado . . . . .	38
3.3. Estudio de simulación . . . . .	44
<b>4. Resultados</b>	<b>48</b>
<b>5. Conclusiones</b>	<b>64</b>
<b>6. Referencias</b>	<b>66</b>
<b>Anexos</b>	<b>72</b>

---

## Índice de figuras

---

1.1.	Mecanismos aleatorios utilizados en el modelo Warner. . . . .	9
1.2.	Diagrama del diseño de Warner. . . . .	9
1.3.	Diagrama del diseño de pregunta no relacionada. . . . .	11
1.4.	Diagrama del diseño de respuesta forzada. . . . .	13
1.5.	Diagrama del diseño de respuesta disfrazada. . . . .	15
1.6.	Diagrama del diseño de Mangat y Singh (1990). . . . .	16
1.7.	Diagrama del diseño de Mangat (1994). . . . .	17
3.1.	Diagrama del mecanismo aleatorio para una pregunta con tres categorías. . . . .	36
4.1.	1000 estimaciones de $\beta_1$ cuando se modela $Y$ con 3 categorías y se usan tamaños de muestras de $n = 125$ (arriba) y $n = 375$ (abajo). . . . .	53
4.2.	1000 estimaciones de $\beta_1$ cuando se modela $Y$ con 4 categorías y se usan tamaños de muestras de $n = 150$ (arriba) y $n = 450$ (abajo). . . . .	54
4.3.	1000 estimaciones de $\beta_1$ cuando $Y$ tiene 3 categorías y se usan tamaños de muestras de $n = 150$ (arriba) y $n = 450$ (abajo). . . . .	60
4.4.	1000 estimaciones de $\beta_2$ cuando $Y$ tiene 3 categorías y se usan tamaños de muestras de $n = 150$ (arriba) y $n = 450$ (abajo). . . . .	61
4.5.	1000 estimaciones de $\beta_1$ cuando $Y$ tiene 4 categorías y se usan tamaños de muestras de $n = 175$ (arriba) y $n = 525$ (abajo). . . . .	62
4.6.	1000 estimaciones de $\beta_2$ cuando $Y$ tiene 4 categorías y se usan tamaños de muestras de $n = 175$ (arriba) y $n = 525$ (abajo). . . . .	63

---

## Indice de Cuadros

---

3.1. Escenarios de simulación . . . . .	45
4.1. Criterios estimados de los estimadores de $\beta_1$ al modelar $Y$ con una variable independiente. . . . .	52
4.2. Criterios estimados de los estimadores de $\beta_1$ y $\beta_2$ al modelar $Y$ con tres categorías y dos variables independientes. . . . .	58
4.3. Criterios estimados de los estimadores de $\beta_1$ y $\beta_2$ al modelar $Y$ con cuatro categorías y dos variables independientes. . . . .	59

## Introducción

---

Se conocen como temas sensitivos aquellos que se consideran amenazantes a la privacidad de una persona, que reflejan opiniones sobre acciones delictivas o comportamientos no aceptados por la sociedad. Por ejemplo, temas como abortos ilegales, consumo de drogas, infidelidad, homosexualidad, evasión de impuestos, corrupción, fraudes etc. En las encuestas cuando se preguntan temas sensitivos de forma directa, comúnmente el encuestado se niega a responder o responde deshonestamente para no dar una imagen negativa ante la sociedad. Estas situaciones provocan que los resultados de las inferencias sean incorrectos y no confiables, debido a respuestas deshonestas o al sesgo de no respuesta; y con ello, las conclusiones carecen de validez porque no aportan información real del individuo en la población. Debido a estos problemas, los investigadores han buscado estrategias que minimicen la resistencia del encuestado a responder honestamente una pregunta sobre el tema sensitivo. Una estrategia que se ha utilizado, es la técnica de respuesta aleatoria, que reduce el sesgo de no respuesta y permite que las inferencias tengan más precisión, confiabilidad y validez.

Warner (1965) fue el primero en proponer un método de respuesta aleatoria para preguntas con respuesta binaria (Sí o No). El método consiste en utilizar un mecanismo aleatorio que selecciona una de dos preguntas complementarias que debe responder el encuestado.

Otros autores hicieron modificaciones al diseño de Warner, cambiando el mecanismo aleatorio o las preguntas que debe responder el encuestado. Entre algunas de las versiones que existen son el diseño de pregunta no relacionada propuesto por Greenberg *et al.* (1969); el diseño de Moors (1971) que optimiza el de la pregunta no relacionada; el diseño de respuesta forzada propuesto por Boruch (1971); el diseño de respuesta disfrazada que propuso Kuk (1990); la implementación de la técnica de Mangat y Singh (1990) que consiste en dos etapas con dos mecanismos aleatorios; la técnica de Mangat (1994) un poco más simple que la de Mangat y Singh (1990); el mecanismo aleatorio alternativo que proponen Singh y Joarder (1997), que consiste en repetir la prueba con el mecanismo aleatorio de Warner si un encuestado que pertenece al grupo sensible no selecciona la pregunta sobre el tema sensible en el primer ensayo; el método de Chua y Tsui (2000), que requiere un valor numérico neutral a diferencia del diseño de Warner, que exige una respuesta directa *Si* o *No*; la propuesta de Odumade y Singh (2009) de utilizar como mecanismo aleatorio dos mazos de cartas en vez de uno.

Los diseños de respuesta aleatoria se han utilizado tanto en el enfoque clásico como en el enfoque Bayesiano para distintos propósitos.

1. Bajo el enfoque clásico, algunos autores han utilizado los diseños de respuesta aleatoria para:
  - a) Estimación de la proporción de una característica sensible binaria; Warner (1965) obtuvo estimaciones insesgadas de máxima verosimilitud; Greenberg *et al.* (1969) también estimaron la proporción de la característica inocua no relacionada; Abernathy *et al.* (1970) investigaron sobre la inducción del aborto ilegal en Carolina del Norte, al utilizar el diseño de la pregunta no relacionada; Chi *et al.* (1972) utilizaron la técnica de respuesta aleatoria para estimar la incidencia de aborto en Taiwan; Kuk (1990) menciona que su método puede ser aplicado a las características cualitativas y cuantitativas;

Mangat y Singh (1990) utilizaron dos mecanismos aleatorios, el primer mecanismo para determinar si el encuestado debe responder la pregunta sensitiva o utilizar el segundo mecanismo que es exactamente el de Warner (1965); Mangat (1994) hace mas simple la técnica de Mangat y Singh (1990), al utilizar el mecanismo aleatorio como en Warner (1965) sólo si el encuestado no pertenece al grupo sensitivo; Chua y Tsui (2000) utilizaron un método que consiste en dos etapas, en la primera etapa cada encuestado selecciona un número aleatorio sin reemplazo de un conjunto de naturales entre 1 y  $m$ , sin decir que número toma, en la segunda etapa genera los conjuntos  $c_1$  y  $c_2$  de  $m$  números, (reorganizándolos ascendentemente) de distribuciones conocidas  $f_1$  y  $f_2$  respectivamente, y se le indica que mencione el número mayor del conjunto  $c_1$  si posee la característica sensitiva, o el número mayor del conjunto  $c_2$  si no la posee; Chaudhuri (2001) mostró como se puede desarrollar una estimación si la muestra de individuos se extrae con probabilidades de selección desiguales con o sin reemplazo; Chang *et al.* (2004) también estimaron la probabilidad de que los individuos que pertenecen al grupo sensitivo informen la verdad; Odumade y Singh (2009) estudiaron la eficiencia relativa del estimador que propusieron, con respecto a otros estimadores; Arnab *et al.* (2012) proporcionaron una generalización de la estrategia de Odumade y Singh (2009); Arnab y Shangodoyin (2015), mediante el método de máxima varosimilitud, mejoraron el estimador de la proporción obtenido por Arnab *et al.* (2012); Alavi y Tajodini (2016) aplicaron la técnica de respuesta aleatoria de manera repetida, con el fin de incrementar la protección de la privacidad de los encuestados y la eficiencia de la estimación de la proporción; Singh y Tarray (2016) sugirieron una clase de estimadores insesgados para la proporción, al utilizar un modelo mixto de respuesta aleatoria.

- b) Estimación de la proporción de una característica sensitiva multinomial; Abul-Ela *et al.* (1967) consideraron el caso cuando al menos una y no más de  $k - 1$  categorías son sensitivas; Mukerjee (1981) sugirió pruebas de muestras grandes para la independencia de los caracteres confidenciales; Kim

y Warde (2005) utilizaron el mecanismo aleatorio de Hopkins (botella con bolas de dos colores diferentes), desarrollado por Liu y Chow (1976), para estimar la proporción bajo el supuesto que el encuestado responde con la verdad y también cuando no lo hacen, compararon proporciones de las categorías sensitivas bajo el supuesto de que el encuestado responde con la verdad, y midieron la relación lineal de dos variables sensitivas mediante la correlación del estimador de momento del producto de Pearson.

c) Estimación de la media y del total de individuos con características sensitivas en una población; Soberanis y Miranda (2011) en un muestreo sin reemplazo y asistido por un modelo de regresión logística, utilizaron la técnica de Morton para estimar el total. La técnica de Morton consiste en la selección aleatoria de preguntas no relacionadas en tres opciones:

- 1) La pregunta sensitiva.
- 2) Una opción no sensitiva, tal que la respuesta siempre es  $S_i$ , o simplemente se le indica al encuestado que diga  $S_i$ .
- 3) Una opción que es el complemento de la segunda.

cada una de las tres opciones respectivamente con probabilidades  $p_1$ ,  $p_2$  y  $p_3$ . Kalucha *et al.* (2015) propusieron un estimador de razón, de la media de una población finita al utilizar un modelo de respuesta aleatoria opcional propuesto por Gupta *et al.* (2010).

d) Modelación de variables binarias aleatorizadas.

Scheers y Dayton (1988) fueron los primeros en utilizar regresión logística con variables respuesta aleatorizadas, aplicado al modelo Warner (1965) y al modelo de Greenberg *et al.* (1969); van der Heijden y van Gils (1996) ajustaron un modelo de regresión logística con los diseños de respuesta forzada y de pregunta no relacionada. van den Hout *et al.* (2007) estudiaron un modelo de regresión logística univariante para variables respuesta aleatorizadas binarias, como un modelo lineal generalizado; también estudiaron un

modelo de regresión logística multivariante, en el que es posible la relación entre varias variables respuesta aleatorizadas y un conjunto de covariables conjuntamente; como ejemplo tomaron los datos de un estudio sobre el fraude de beneficios sociales. van der Heijden *et al.* (2011) modelaron mediante regresión logística una variable respuesta binaria aleatorizada y mediante regresión ordinal, una variable  $W$  (suma de  $m$  preguntas sobre características sensitivas binarias con respuestas  $S_i$ ) con variables independientes. Blair *et al.* (2015) utilizaron la variable binaria aleatorizada en el modelo de regresión logística de dos maneras, como variable respuesta o como variable independiente; estimaron los parámetros del modelo mediante el algoritmo EM (Esperanza-Maximización).

e) Modelación de variables multinomiales.

Cruyff *et al.* (2006) estudiaron un modelo Poisson cero inflado, para una variable aleatorizada con múltiples categorías de respuesta ordenada, para tener en cuenta el sesgo de respuesta de autoprotección; el estudio fue bajo el supuesto de que las categorías de la característica sensitiva siguen una distribución Poisson. Proporcionaron un ejemplo de una variable de una encuesta Holandesa sobre el fraude a la seguridad social, utilizaron el diseño de respuesta forzada.

f) Comparaciones entre diseños de respuesta aleatoria; Moors (1971) mejoró una de las sugerencias sobre la elección de los parámetros del modelo de Greenberg *et al.* (1969) y probó la superioridad de la optimización del modelo de pregunta no relacionada comparado con el modelo Warner. Tracy y Fox (1981) compararon la validez de la técnica de respuesta aleatoria al realizar la entrevista con respuesta aleatoria y con pregunta directa, mostraron que el método de respuesta aleatoria superó al método directo al revisar el error cuadrático medio en cada caso. Chaudhuri y Mukherjee (1987) revisaron varios diseños de respuesta aleatoria aplicados a características cuantitativas y cualitativas, algunos aplicados a poblaciones infinitas. Lensvelt-Mulders

*et al.* (2005) compararon la eficiencia de 6 diseños de respuesta aleatoria; en sus resultados mostraron que el diseño de respuesta forzada y una forma especial del diseño de pregunta no relacionada fueron más eficientes. De Schrijver (2012) analizó los diseños de respuesta forzada y de pregunta no relacionada para estudiar si los encuestados entienden y confían en los diseños, sus resultados mostraron que la comprensión fue significativamente mayor en el diseño de respuesta forzada y la confianza fue baja en ambos diseños.

2. Otro de los enfoques de la inferencia en el que se han utilizado los diseños de respuesta aleatoria es en el Bayesiano. Bajo este enfoque, Kim *et al.* (2005) propusieron una versión Bayesiana del modelo de Mangat (1994), para estimar la proporción de una característica sensitiva binaria. Kim y Heo (2011) desarrollaron un modelo de respuesta aleatoria multinomial bayesiano, al considerar el modelo de respuesta aleatoria multinomial de Kim y Warde (2005), para estimar la proporción de una característica sensitiva multinomial.

En la literatura revisada no se encontró antecedentes sobre modelación Bayesiana de variables respuesta aleatorizadas ordinales, por lo que este es el objetivo en el presente trabajo, modelar bajo el enfoque Bayesiano, variables respuesta aleatorizadas ordinales, con un conjunto de variables independientes que miden características sensitivas del individuo.

Se utiliza el modelo probit ordinal y se proponen cuatro distribuciones iniciales, la distribución Normal, la Doble Exponencial, la *t* de Student y la Cauchy. Las distribuciones posteriores se aproximan, al utilizar el programa JAGS en el software estadístico R y para mejorar la eficiencia del proceso MCMC se consideran estandarizadas las variables independientes. Las estimaciones que se obtienen con cada una de las distribuciones iniciales, se evalúan para compararlas.

En el resto de este primer capítulo se da una breve descripción de algunos diseños de respuesta aleatoria: el diseño de Mangat y Singh (1990), el diseño de Mangat (1994) y de los cuatro diseños que usan Blair *et al.* (2015), el diseño de Warner (1965), el diseño de pregunta no relacionada, el diseño de respuesta forzada y el diseño de respuesta disfrazada; aunque en el presente trabajo se utilizó un mecanismo aleatorio usado para el diseño de respuesta forzada generalizado a variables multinomiales. En el segundo capítulo se presenta la teoría sobre el enfoque Bayesiano, los métodos numéricos y los criterios de evaluación de estimadores. En el tercer capítulo se presenta la metodología utilizada y el modelo propuesto. En el cuarto capítulo se presentan los resultados obtenidos, enseguida las Conclusiones, las Referencias y finalmente los Anexos.

## 1.1. Algunos diseños de respuesta aleatoria

### 1.1.1. Modelo Warner

El modelo Warner fue el primer diseño de respuesta aleatoria propuesto para disminuir el sesgo de no respuesta y proteger la privacidad del encuestado cuando se preguntan temas sensitivos. Este diseño nombrado así en honor a Stanley Warner (1965), fue propuesto para preguntas con respuesta binaria. El método consiste en utilizar un mecanismo aleatorio, por ejemplo un juego de cartas, una ruleta, el lanzamiento de dados, una urna con bolas de colores, etc. El diseño considera a la población dividida en dos grupos: el grupo de individuos que tienen la característica sensitiva,  $A$ , con probabilidad  $\pi_A$  y el grupo de individuos que no tienen la característica sensitiva,  $A^C$ , con probabilidad  $1 - \pi_A$ . El mecanismo aleatorio determina si se debe responder la pregunta ¿Pertenece a  $A$ ? con probabilidad  $p$  o la pregunta complementaria ¿Pertenece a  $A^C$ ? con probabilidad  $1 - p$ . Se pide al encuestado que responda honestamente con un *Si* o un *No* la pregunta seleccionada, sin mencionar al entrevistador que pregunta selecciona, para proteger su privacidad.

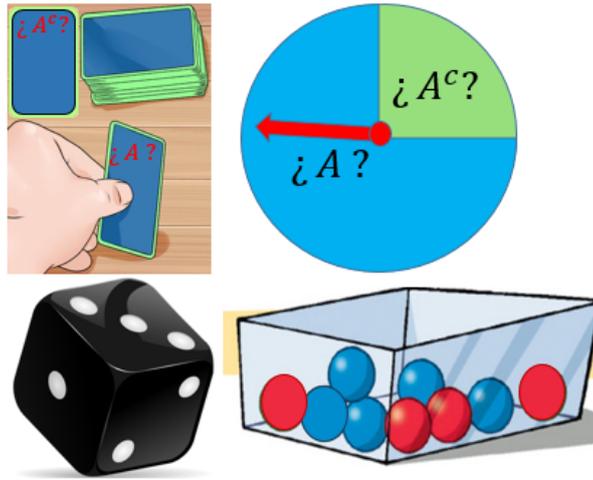


Figura 1.1: Mecanismos aleatorios utilizados en el modelo Warner.

Bajo el modelo Warner la probabilidad de pertenecer al grupo con la característica sensitiva,  $\pi_A$ , se puede obtener al considerar una variable binaria latente  $Z$  y una variable binaria observada  $Y$  tal que:

$$Y_i = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo responde si} \\ 0 & \text{si el } i\text{-ésimo individuo responde no} \end{cases}$$

$$Z_i = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo pertenece al grupo } A \\ 0 & \text{si el } i\text{-ésimo individuo pertenece al grupo } A^C \end{cases}$$

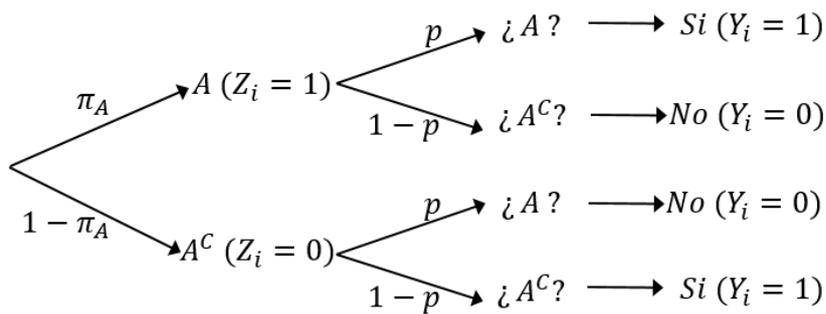


Figura 1.2: Diagrama del diseño de Warner.

Mediante el mecanismo aleatorio la probabilidad de que el  $i$ -ésimo encuestado,  $i = 1, \dots, n$ , pertenezca al grupo con la característica sensitiva,  $\pi_A = P(Z_i = 1)$ , está

relacionada con la probabilidad de respuesta  $S_i$ ,  $\pi = P(Y_i = 1)$ , mediante la ecuación:

$$\begin{aligned}\pi &= P(Z_i = 1)P(A^?|Z_i = 1)P(Y_i = 1|Z_i = 1, A^?) \\ &\quad + P(Z_i = 0)P(A^C^?|Z_i = 0)P(Y_i = 1|Z_i = 0, A^C^?) \\ &= \pi_A p + (1 - \pi_A)(1 - p) \\ &= (2p - 1)\pi_A + (1 - p)\end{aligned}$$

Así la probabilidad de que cada encuestado pertenezca al grupo con la característica sensitiva  $\pi_A$  está dada por:

$$\pi_A = \frac{\pi - (1 - p)}{2p - 1}; \quad p \neq \frac{1}{2}$$

Se debe tener especial cuidado al seleccionar el valor de  $p$ , el cual debe ser de forma que el encuestado esté convencido de que su privacidad se mantendrá oculta,  $p = 1/2$  sería ideal pero con este valor el estimador no estará definido y si  $p = 1$  sería equivalente a realizar la pregunta directa, justo lo que el encuestado se niega a responder, por lo que preferentemente debe seleccionarse  $\frac{1}{2} < p < 1$ .

### 1.1.2. Diseño de pregunta no relacionada

El diseño de pregunta no relacionada fue desarrollado por Greenberg *et al.* (1969) para dos casos, uno con y el otro sin la probabilidad conocida de la característica inocua no relacionada. Al igual que en el modelo Warner, también se consideran dos preguntas, una de las cuales es la pregunta sobre el tema sensitivo y la otra a diferencia de la pregunta complementaria, es cualquier pregunta que no esté relacionada con el tema sensitivo, por ejemplo la pregunta sobre el tema sensitivo podría ser ¿Pertenece al grupo de individuos que consumen droga? y la pregunta no relacionada con el tema sensitivo podría ser ¿Pertenece al grupo de individuos que les gusta la música clásica?

El diseño desarrollado para el caso cuando se conoce la probabilidad de la característica inocua no relacionada, se describe a continuación:

Al igual que en el modelo Warner la población se divide en dos grupos, el mecanismo

aleatorio determina si se responde la pregunta sobre el tema sensitivo, denotada por  $A$ , con probabilidad  $p$  de ser seleccionada o la pregunta no relacionada, denotada por  $B$ , con probabilidad  $1 - p$  de ser seleccionada y con probabilidad conocida  $q$  de respuesta *Si*. Igualmente que en el modelo Warner, se pide al encuestado responder con un *Si* o un *No* de manera honesta, a una de las dos preguntas seleccionadas, sin mencionar el resultado del mecanismo aleatorio para proteger su privacidad.

En la Figura 1.3, se presenta un diagrama del diseño de pregunta no relacionada, para el caso cuando se conoce la probabilidad de la característica inocua no relacionada, donde  $Z$  y  $Y$  son variables binarias descritas como en el modelo Warner.

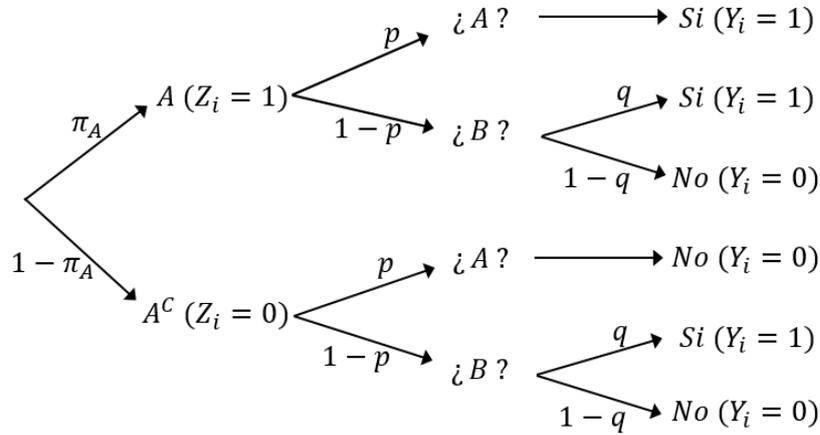


Figura 1.3: Diagrama del diseño de pregunta no relacionada.

Para este caso la probabilidad de respuesta *Si* se obtiene mediante:

$$\begin{aligned}
 \pi &= P(Z_i = 1)P(A?|Z_i = 1)P(Y_i = 1|Z_i = 1, A?) \\
 &\quad + P(Z_i = 1)P(B?|Z_i = 1)P(Y_i = 1|Z_i = 1, B?) \\
 &\quad + P(Z_i = 0)P(B?|Z_i = 0)P(Y_i = 1|Z_i = 0, B?) \\
 &= \pi_A p + \pi_A(1 - p)q + (1 - \pi_A)(1 - p)q \\
 &= p\pi_A + q(1 - p)
 \end{aligned}$$

Y la probabilidad de pertenecer al grupo con la característica sensitiva  $\pi_A$  es:

$$\pi_A = \frac{\pi - (1 - p)q}{p}, \quad p \neq 0$$

Desde el punto de vista práctico, el valor de  $p$  debe ser distinto de cero, porque con este valor es equivalente a sólo realizar la pregunta no relacionada y no la pregunta sobre el tema sensitivo, por lo que, no se obtendría la información de interés; distinto de uno porque con este valor es equivalente a realizar la pregunta sensitiva de manera directa, justo lo que el encuestado no quiere responder.

### 1.1.3. Diseño de respuesta forzada

El diseño de respuesta forzada también fue desarrollado para preguntas con respuesta binaria. Al igual que en el modelo Warner, la población también se considera dividida en dos grupos. El mecanismo aleatorio en este diseño, determina si el encuestado responde con honestidad a la pregunta sobre el tema sensitivo, con probabilidad  $p$  o responde con un *Si* o un *No* de manera forzada, con probabilidades  $p_1$  y  $p_0$  respectivamente, independientemente si tiene o no la característica sensitiva. Por ejemplo, Blair *et al.* (2015) usaron como mecanismo aleatorio el lanzamiento de un dado, en el que el encuestado responde *No* con probabilidad  $p_0$  si el resultado del lanzamiento del dado es 1, o responde con honestidad la pregunta sobre el tema sensitivo con probabilidad  $p$ , si el resultado del lanzamiento es 2, 3, 4 o 5, o responde *Si* con probabilidad  $p_1$  si el resultado es 6, donde  $p_0 + p + p_1 = 1$ . Por otro lado, van den Hout *et al.* (2007) utilizaron el lanzamiento de dos dados, en el cual consideraron la suma de los valores resultantes del lanzamiento.

En la Figura 1.4 se presenta un diagrama sobre el diseño de respuesta forzada, donde las variables binarias  $Z$  y  $Y$  se consideran como se describieron en el modelo Warner.

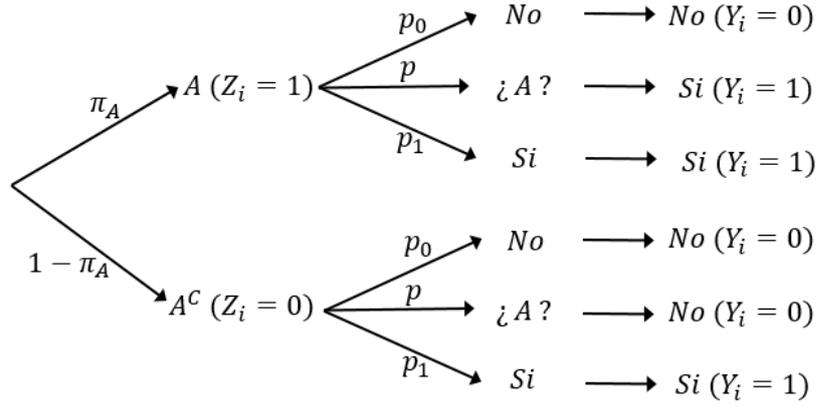


Figura 1.4: Diagrama del diseño de respuesta forzada.

En este diseño la probabilidad de respuesta  $Si$ ,  $\pi$ , para cada individuo,  $i = 1, \dots, n$ , y la probabilidad de pertenecer al grupo con la característica sensible,  $\pi_A$ , están relacionadas por:

$$\begin{aligned}
 \pi &= P(Z_i = 1)P(A?|Z_i = 1)P(Y_i = 1|Z_i = 1, A?) \\
 &\quad + P(Z_i = 1)P(Si|Z_i = 1)P(Y_i = 1|Z_i = 1, Si) \\
 &\quad + P(Z_i = 0)P(Si|Z_i = 0)P(Y_i = 1|Z_i = 0, Si) \\
 &= \pi_A p + \pi_A p_1 + (1 - \pi_A)p_1 \\
 &= p\pi_A + p_1 \\
 \pi_A &= \frac{\pi - p_1}{p} \quad p \neq 0
 \end{aligned}$$

El valor de  $p$  debe ser distinto de cero, para que tenga sentido realizar la pregunta sobre el tema sensible y se obtenga la información de interés, también debe ser distinto de uno para evitar realizar la pregunta sensible de manera directa.

#### 1.1.4. Diseño de respuesta disfrazada

Otro diseño similar al de Warner, es el de respuesta disfrazada propuesto por Kuk (1990), el cual no requiere la respuesta directa del encuestado. Igual que en el modelo Warner también se utilizan dos preguntas, la pregunta sensible con probabilidad  $p$  de ser seleccionada y la pregunta complementaria con probabilidad  $1 - p$ . La modificación

en este diseño es que el encuestado en vez de responder *Si* o *No*, sus respuestas son sustituidas por palabras inocuas, es decir, palabras que no afectan la sensibilidad de las respuestas, para evitar la incomodidad de proporcionar respuestas directas. Un ejemplo de este diseño, es cuando el mecanismo aleatorio, consiste en que al encuestado se le presentan dos cajas con cartas, una caja con cartas a la izquierda, que representa la respuesta *Si*, la cual tiene una proporción  $p$  de cartas rojas y  $1 - p$  de cartas negras, y la otra caja con cartas a la derecha, que representa la respuesta *No*, la cual contiene proporciones complementarias de las cartas, es decir, una proporción  $1 - p$  de cartas rojas y una proporción  $p$  de cartas negras. Bajo el diseño se le explica al encuestado que tome una carta de cada caja, de igual manera sólo él debe conocer el resultado del mecanismo, sin olvidar la respuesta que representan las cajas, cuando el entrevistador realiza la pregunta directa sobre el tema sensitivo, se le indica que proporcione como respuesta, el color de la carta seleccionada de la caja que representa su respuesta honesta. De esta manera su privacidad está protegida, porque no da una respuesta directa afirmativa o negativa, sino que es sustituida por el color de la carta seleccionada.

En la Figura 1.5, se presenta el diagrama sobre el diseño de respuesta disfrazada para este ejemplo, donde  $Z$  y  $Y$  son variables binarias, tal que para cada individuo, se describen como:

$$Y_i = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo responde roja} \\ 0 & \text{si el } i\text{-ésimo individuo responde negra} \end{cases}$$

$$Z_i = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo pertenece al grupo } A \\ 0 & \text{si el } i\text{-ésimo individuo pertenece al grupo } A^C \end{cases}$$

Bajo este diseño y mediante el mecanismo aleatorio, si responder *Roja* es equivalente a una respuesta *Si*, entonces la probabilidad de que el  $i$ -ésimo encuestado, responda *Roja*, es igual a la probabilidad de responder *Si* en el modelo Warner.

$$\begin{aligned} \pi &= P(Z_i = 1)P(Roja|Z_i = 1)P(Y_i = 1|Z_i = 1, Roja) \\ &\quad + P(Z_i = 0)P(Roja|Z_i = 0)P(Y_i = 1|Z_i = 0, Roja) \\ &= \pi_A p + (1 - \pi_A)(1 - p) = (2p - 1)\pi_A + (1 - p) \end{aligned}$$

Al igual que en el modelo Warner, la probabilidad de pertenecer al grupo con la característica sensitiva,  $\pi_A$ , se determina mediante:

$$\pi_A = \frac{\pi - (1 - p)}{2p - 1}$$

El valor de  $p$  debe ser distinto de  $1/2$  para que el estimador de  $\pi_A$  esté definido y distinto de uno para evitar realizar la pregunta sensitiva de manera directa, es decir,  $p \neq \frac{1}{2}, 1$ .

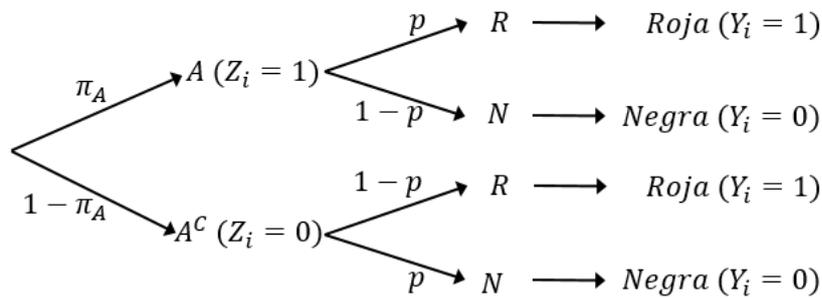


Figura 1.5: Diagrama del diseño de respuesta disfrazada.

### 1.1.5. Diseño de Mangat y Singh (1990)

El método que propusieron Mangat y Singh (1990) también considera la población dividida en dos grupos, similar al modelo Warner, consiste en que cada encuestado utiliza dos mecanismos aleatorios; con el primer mecanismo aleatorio se selecciona con probabilidad  $t$  responder la pregunta sensitiva o con probabilidad  $1 - t$  utilizar el segundo mecanismo aleatorio (M2), que es exactamente como en el modelo Warner, es decir, que con probabilidad  $p$  se selecciona la pregunta sensitiva y con probabilidad  $1 - p$  se selecciona la pregunta complementaria, para responder con un *Si* o con un *No* de manera honesta.

En la Figura 1.6, se presenta el diagrama del diseño de Mangat y Singh (1990), donde  $Z$  y  $Y$  son variables binarias como se describieron en el modelo Warner. Mediante este diseño, la probabilidad de respuesta *Si*,  $\pi$ , está relacionada con la probabilidad de pertenecer al grupo con la característica sensitiva,  $\pi_A$ , mediante la ecuación:

$$\begin{aligned}
\pi &= P(Z_i = 1)P(A?|Z_i = 1)P(Y_i = 1|Z_i = 1, A?) \\
&\quad + P(Z_i = 1)P(M2|Z_i = 1)P(A?|Z_i = 1, M2)P(Y_i = 1|Z_i = 1, M2, A?) \\
&\quad + P(Z_i = 0)P(M2|Z_i = 0)P(A^C?|Z_i = 0, M2)P(Y_i = 1|Z_i = 0, M2, A^C?) \\
&= \pi_A t + \pi_A(1-t)p + (1-\pi_A)(1-t)(1-p) \\
&= t\pi_A + (1-t)[p\pi_A + (1-p)(1-\pi_A)]
\end{aligned}$$

Entonces  $\pi_A$  es:

$$\pi_A = \frac{\pi_i - (1-p)(1-t)}{2p - 1 + 2t(1-p)}$$

El valor de  $t$  debe ser distinto de cero, porque con este valor el método se reduce exactamente al modelo Warner y distinto de uno porque es equivalente a realizar la pregunta sensitiva de forma directa, justo lo que el encuestado se niega a responder. También el valor de  $p$  debe ser distinto de cero y de uno, distinto de cero porque con este valor no tiene sentido un segundo mecanismo, se reduce a utilizar el modelo Warner, y distinto de uno porque se reduce a realizar la pregunta de forma directa.

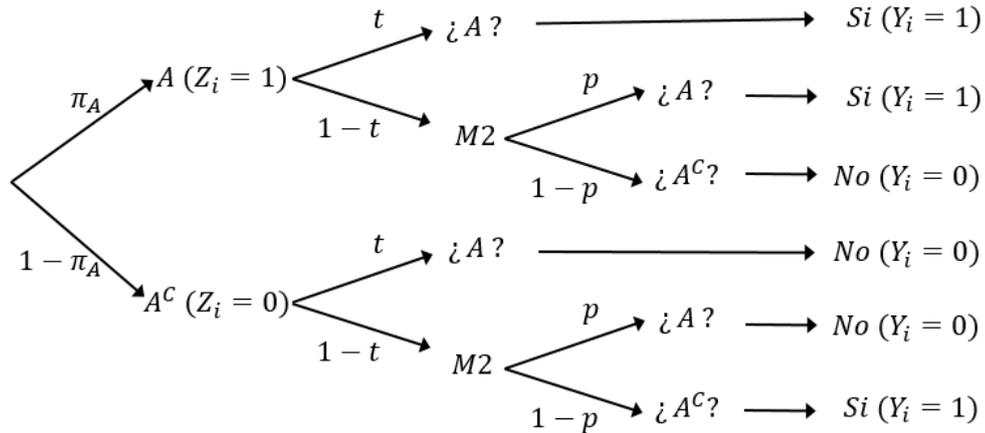


Figura 1.6: Diagrama del diseño de Mangat y Singh (1990).

### 1.1.6. Diseño de Mangat (1994)

El diseño de Mangat (1994) es un poco más simple que el diseño que desarrollaron Mangat y Singh (1990). Este método también considera a la población dividida en dos grupos complementarios. Cada uno de los encuestados se selecciona por muestreo con reemplazo con igual probabilidad; con la instrucción de responder  $S_i$ , si tiene la característica sensitiva, en caso contrario, utiliza el modelo Warner, es decir, mediante un mecanismo aleatorio se determina si responde la pregunta sensitiva o la pregunta complementaria.

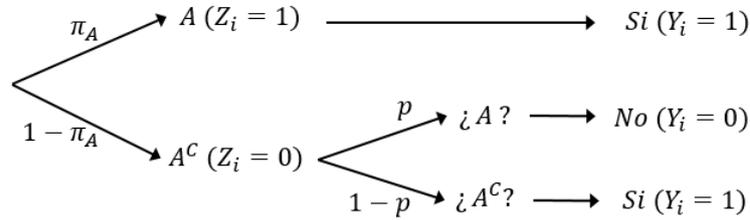


Figura 1.7: Diagrama del diseño de Mangat (1994).

Con este método, la probabilidad de respuesta  $S_i$ , al considerar las variables binarias  $Z$  y  $Y$  como en el modelo Warner, se determina como:

$$\begin{aligned}
 \pi &= P(Z_i = 1)P(Y_i = 1|Z_i = 1) + P(Z_i = 0)P(A^{C?}|Z_i = 0)P(Y_i = 1|Z_i = 0, A^{C?}) \\
 &= \pi_A + (1 - \pi_A)(1 - p) \\
 &= (1 - p) + p\pi_A
 \end{aligned}$$

Así para cada encuestado, la probabilidad de pertenecer al grupo con la característica sobre el tema sensitivo es:

$$\pi_A = \frac{\pi - (1 - p)}{p}, \quad p \neq 0$$

Desde el punto de vista práctico, el valor de  $p$  debe ser distinto de cero, porque si es cero, equivale a no utilizar el mecanismo aleatorio y las respuestas serían de forma directa de pertenecer o no al grupo con la característica sensitiva. También debe ser distinto de uno, porque eso llevaría a realizar la pregunta de manera directa, justo lo que el encuestado se negaría a responder.

---

### Teoría Bayesiana

---

La inferencia Bayesiana es una alternativa a la inferencia clásica, para estudiar problemas a través de la estimación de parámetros o contrastes de hipótesis. El término *bayesiano*, se debe a que en el proceso de inferencia se usa el famoso *teorema de Bayes*, publicado en 1763, del matemático reverendo Thomas Bayes (1702-1761).

En la inferencia Bayesiana, a diferencia de la inferencia clásica, el parámetro de interés en el modelo de probabilidad, es considerado como una variable aleatoria, donde el conocimiento que se tiene de este, tiene cierto grado de incertidumbre. Mediante el enfoque subjetivo, se puede incorporar información inicial que exprese el conocimiento o ignorancia sobre el parámetro y mediante *el teorema de Bayes* esa información se actualiza al observar la evidencia de los datos, para tomar decisiones basada en la nueva información. La inferencia Bayesiana, se basa en la función de densidad posterior del parámetro, después de analizar los datos.

## 2.1. Inferencia Bayesiana

La función de masa o de densidad de probabilidad conjunta, por definición es el producto de la distribución de los datos  $\pi(\mathbf{y}|\theta)$  y la distribución inicial  $\pi(\theta)$ :

$$\pi(\mathbf{y}, \theta) = \pi(\mathbf{y}|\theta)\pi(\theta)$$

Entonces por el teorema de Bayes, la distribución posterior de  $\theta$  dado  $\mathbf{y}$  es :

$$\pi(\theta|\mathbf{y}) = \frac{\pi(\mathbf{y}, \theta)}{\pi(\mathbf{y})} = \frac{\pi(\mathbf{y}|\theta)\pi(\theta)}{\pi(\mathbf{y})}$$

donde la distribución marginal de  $\mathbf{y}$  es:

$$\pi(\mathbf{y}) = \begin{cases} \int \pi(\mathbf{y}|\theta)\pi(\theta)d\theta, & \text{si } \theta \text{ es continua} \\ \sum \pi(\mathbf{y}|\theta)\pi(\theta), & \text{si } \theta \text{ es discreta} \end{cases}$$

la cual no depende del parámetro  $\theta$ . Dado el vector  $\mathbf{y}$  de observaciones, la distribución marginal, es una constante de normalización en la distribución condicional  $\pi(\theta|\mathbf{y})$ , por lo que puede omitirse para obtener la forma equivalente, la distribución posterior no normalizada (Gelman et al., 2014, pág. 6).

$$\pi(\theta|\mathbf{y}) \propto \pi(\mathbf{y}|\theta)\pi(\theta)$$

La distribución  $\pi(\theta)$  aporta información del parámetro  $\theta$  antes de conocer los datos; cuando no existe conocimiento sobre el parámetro o el conocimiento existente es difícil de resumir en una información previa, esa ignorancia se expresa como distribuciones iniciales no informativas. Algunas distribuciones no informativas son la distribución Uniforme entre  $-\infty$  y  $+\infty$  o entre 0 y  $+\infty$  o la inicial de Jeffreys, algunas distribuciones iniciales no informativas pueden ser impropias (cuando su integral es diferente de uno sobre el rango de valores posibles) (Congdon, 2006, pág, 4).

Las distribuciones posteriores a veces se pueden obtener de forma sencilla al realizar cálculos analíticos o numéricos, cuando la distribución inicial es conjugada. Conjugación es la propiedad de que la distribución posterior sigue la misma forma paramétrica que la distribución inicial, entonces la distribución inicial es conjugada para la función

de verosimilitud (Kruschke, 2014, pág. 126). Cuando se tiene un vector de parámetros de alta dimensión, a veces no es posible calcular la distribución posterior mediante cálculos analíticos, debido a que se deben resolver múltiples integrales complejas, por lo que es conveniente utilizar métodos computacionales, que implican un muestreo repetido que convergen al muestreo de la distribución posterior del parámetro. Estos métodos computacionales son algoritmos iterativos basados en Cadenas de Markov Monte Carlo (MCMC, por sus siglas en inglés). Dos de los algoritmos conocidos, son el Metrópolis-Hastings y el muestreador de Gibbs.

## 2.2. Cadenas de Markov Monte Carlo

Una cadena de Markov  $\mathbf{X}=\{X^{(t)}\}$ , es una secuencia de variables aleatorias dependientes indexada por el tiempo  $t \geq 0$ . La secuencia  $X^{(0)}, X^{(1)}, \dots, X^{(t)}, \dots$ , que toma valores en un espacio de estado discreto, es una cadena de Markov si cumple con la propiedad de Markov, es decir, si la distribución condicional de cualquier variable en un determinado estado dadas todas las anteriores, sólo depende de la inmediatamente anterior,  $P(X^{(t)}|X^{(0)}, X^{(1)}, \dots, X^{(t-1)}) = P(X^{(t)}|X^{(t-1)})$ . A la distribución  $P(X^{(t)}|X^{(t-1)})$  se le denomina kernel de transición o kernel de Markov. A medida que la cadena de Markov va avanzando, va convergiendo a una distribución estacionaria o invariante que se aproxima a la distribución posterior del parámetro de interés. Para que la cadena de Markov converja a una distribución estacionaria, se deben cumplir las siguientes condiciones:

1. Que sea homogénea en el tiempo, es decir, que la distribución de probabilidad de transición entre dos estados no cambie en el tiempo. En términos de probabilidad, esto significa que la probabilidad de ir del estado  $i$  al estado  $j$  en un paso no dependa del tiempo en el que se encuentra la cadena:

$$P(X_n = j|X_{n-1} = i) = P(X_1 = j|X_0 = i)$$

2. Que sea irreducible, es decir, que todos los estados de la cadena, se comuniquen con todos los demás estados, con probabilidad positiva en tiempo finito.

3. Que sea recurrente positiva, es decir, si el tiempo promedio de recurrencia de un estado recurrente  $i$  de la cadena, es finito. Un estado  $i$  es recurrente, si la cadena regresa a  $i$ , con probabilidad 1.

La mayoría de las cadenas involucradas en los algoritmos MCMC son recurrentes y la distribución límite es la distribución estacionaria. En la simulación, si un kernel dado produce una cadena de Markov ergódica con distribución estacionaria  $f$ , entonces al generar una cadena de este kernel, producirá simulaciones de la distribución  $f$ . Por lo que el objetivo de generar muestras vía Cadenas de Markov, es para estimar cualquier función integrable  $h$  de interés, a través de promedios, es decir, de las simulaciones  $X^{(0)}, X^{(1)}, \dots, X^{(t)}, \dots$ , se puede estimar  $E_f(h(X))$  con:

$$\bar{h}_t = \frac{1}{T} \sum_{t=0}^T h(X^{(t)})$$

donde  $\bar{h}_t$  converge a  $E_f(h(X))$  cuando  $t \rightarrow \infty$ , (Robert y Casella, 2010, pág. 169).

### 2.2.1. Metrópolis-Hastings

El algoritmo de Metropolis-Hastings, es un método de simulación de Cadenas de Markov, para aproximar una distribución posterior arbitraria  $\pi(\theta|\mathbf{y})$  que es difícil de muestrear directamente. El algoritmo construye un kernel de Markov con distribución estacionaria  $\pi(\theta|\mathbf{y})$ , luego se genera una cadena de Markov al usar este kernel, de manera que la distribución límite de la cadena sea  $\pi(\theta|\mathbf{y})$ . Este algoritmo se basa en una condición de aceptación o rechazo, que garantiza que el método genere muestras de la distribución estacionaria que converge a la distribución posterior de interés. Para el algoritmo se propone una función de densidad condicional  $q(\cdot|x)$  para generar valores candidatos, el valor generado, es evaluado para ser aceptado o rechazado con cierta probabilidad  $\alpha(x, \theta^*)$ . El algoritmo Metropolis-Hastings produce una cadena de Markov a través de la siguiente transición (Robert y Casella, 2010, pág. 170-171).

1. Se inicia la cadena con un valor arbitrario  $x^{(0)}$ .
2. Se repite para la iteración  $j = 1, 2, \dots, N$ .

3. Se genera un valor  $\theta^*$  de la distribución condicional  $q(\cdot|x^{(j)})$  y un valor  $u$  de la distribución  $U(0, 1)$ .

4. Se calcula la probabilidad de aceptación:

$$\alpha(x^{(j)}, \theta^*) = \min \left( 1, \frac{\pi(\theta^*|\mathbf{y})q(x^{(j)}|\theta^*)}{\pi(x^{(j)}|\mathbf{y})q(\theta^*|x^{(j)})} \right)$$

5. Si  $u \leq \alpha(x^{(j)}, \theta^*)$  entonces el valor  $\theta^*$  es aceptado y pasa a ser el siguiente estado de la cadena, es decir,  $x^{(j+1)} = \theta^*$ .

6. En caso contrario el siguiente estado de la cadena sigue siendo el valor del estado  $j$ , es decir,  $x^{(j+1)} = x^{(j)}$ .

7. El proceso se repite, tal que la distribución de la cadena de Markov  $\{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$ , converja a la distribución objetivo.

### 2.2.2. Muestreador de Gibbs

Es otro de los algoritmos de simulación utilizado para generar Cadenas de Markov Monte Carlo, cuya distribución estacionaria converge a la distribución posterior del parámetro de interés.

El algoritmo para producir una muestra de Gibbs para un vector de  $k$  parámetros  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ , está dada por la siguiente transición de  $\theta^{(j)}$  a  $\theta^{(j+1)}$  (Robert y Casella, 2010, pág. 206):

1. Se inicia la muestra al fijar valores iniciales para todos los parámetros  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_k^{(0)})$ .
2. En la transición  $j + 1$ , para cada parámetro, se genera un valor de la distribución condicional del parámetro  $\theta_k$ , dado los valores generados en la transición inmediatamente anterior de los parámetros restantes.

$$\begin{aligned}
\theta_1^{(j+1)} &\sim \pi_1(\theta_1|\theta_2^{(j)}, \theta_3^{(j)}, \theta_4^{(j)}, \dots, \theta_k^{(j)}) \\
\theta_2^{(j+1)} &\sim \pi_2(\theta_2|\theta_1^{(j+1)}, \theta_3^{(j)}, \theta_4^{(j)}, \dots, \theta_k^{(j)}) \\
\theta_3^{(j+1)} &\sim \pi_3(\theta_3|\theta_1^{(j+1)}, \theta_2^{(j+1)}, \theta_4^{(j)}, \dots, \theta_k^{(j)}) \\
&\vdots \\
\theta_k^{(j+1)} &\sim \pi_k(\theta_k|\theta_1^{(j+1)}, \theta_2^{(j+1)}, \dots, \theta_{k-1}^{(j+1)})
\end{aligned}$$

Las distribuciones  $\pi_1, \pi_2, \dots, \pi_k$  son llamadas distribuciones completas.

3. El proceso de muestreo, es repetido hasta que el límite de la cadena converja a la distribución posterior del parámetro.

### 2.2.3. Diagnósticos de convergencia de la cadena

Como se mencionó anteriormente, para que una cadena converja a una distribución estacionaria, esta debe ser irreducible, homogénea y positiva recurrente. Algunas formas de verificar la convergencia de la cadena es utilizando algunos diagnósticos, ya sean gráficos o pruebas estadísticas que den indicios de que las cadenas convergen. Entre algunas de las pruebas estadísticas, están el diagnóstico de Geweke (1992), el de Raftery y Lewis (1992), el de Heidelberger y Welch (1993) y el de Gelman y Rubin (1992) que están implementados en el paquete CODA del software estadístico R (Core Team, 2016). En este trabajo se describe el diagnóstico de Gelman y Rubin (1992), el cual será utilizado, además de un diagnóstico gráfico que es la traza de una cadena de Markov.

#### Traza de la Cadena de Markov

Una forma visual de diagnosticar la convergencia de las cadenas de Markov, es mediante el gráfico de la traza de al menos dos secuencias paralelas del parámetro de interés. En este gráfico se relaciona el número de iteraciones contra los valores de las estimaciones del parámetro en cada iteración, una forma de decir que las cadenas convergen a la distribución estacionaria, es cuando los valores de las estimaciones del

parámetro en cada cadena están en un corto rango, en el transcurso de las iteraciones, es decir, que las cadenas estén superpuestas o entrelazadas (mezcladas); en caso contrario probablemente se necesite aumentar el número de iteraciones para llegar a la convergencia o simplemente no hay convergencia en las cadenas.

## Diagnóstico de Gelman y Rubin

El diagnóstico de Gelman y Rubin (factor de reducción de escala potencial, R), es una prueba estadística para diagnosticar la convergencia de más de dos cadenas de Markov, la prueba consiste en medir la varianza entre las cadenas y dentro de las cadenas, este diagnóstico es calculado con la función `gelman.diag` del paquete CODA en el software estadístico R.

El factor de reducción de escala potencial, R, es estimado mediante el siguiente procedimiento:

1. Se generan  $m \geq 2$  cadenas, cada una con  $2n$  iteraciones y con puntos iniciales dispersos.
2. Para disminuir el efecto de las primeras iteraciones en las cadenas (periodo de calentamiento), se eliminan las primeras  $n$  iteraciones y se trabaja con las restantes.
3. Sea  $\psi$  un estadístico que estima algún parámetro de la distribución objetivo. De las  $m$  cadenas generadas con las restantes  $n$  iteraciones  $\{X_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$ , se calcula  $\psi_{nj} = \psi(X_{1j}, \dots, X_{nj})$  para cada cadena y se calculan las estimaciones de la varianza entre y dentro de las cadenas para estimar la varianza de  $\psi$ .

- Varianza entre cadenas (varianza de los promedios dentro de las cadenas):

$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\psi}_{.j} - \bar{\psi}_{..})^2$$

donde

$$\bar{\psi}_{.j} = \frac{1}{n} \sum_{i=1}^n \psi_{ij} \quad \bar{\psi}_{..} = \frac{1}{m} \sum_{j=1}^m \bar{\psi}_{.j}$$

La varianza entre cadenas,  $B$ , contiene un factor de  $n$  porque se basa en la varianza de los promedios dentro de la cadena,  $\bar{\psi}_{.j}$ , cada uno de los cuales es un promedio de  $n$  valores  $\psi_{ij}$  (Gelman et al., 2014, pág. 284).

- Varianza dentro de la cadena (promedio de las varianzas dentro de las cadenas):

$$W = \frac{1}{m} \sum_{j=1}^m s_j^2 \quad s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\psi_{ij} - \bar{\psi}_{.j})^2$$

4. Las estimaciones de la varianza entre y dentro de las cadenas se combinan para estimar un límite superior para la varianza de  $\psi$ :

$$\widehat{V}(\psi) = \frac{n-1}{n}W + \frac{1}{n}B$$

Si las cadenas son muestras aleatorias de la distribución objetivo,  $\widehat{V}(\psi)$  es un estimador insesgado de  $V(\psi)$ . Gelman y Rubin (1992) comparan  $\widehat{V}(\psi)$  y  $W$  los cuales son asintóticamente equivalentes. Si los valores iniciales de la cadena están sobredispersos pero convergen a  $V(\psi)$  cuando  $n \rightarrow \infty$ , entonces  $\widehat{V}(\psi)$  sobreestima la  $V(\psi)$ ; si las cadenas no han llegado a la convergencia en la iteración  $n$ , es decir, si aún no se han mezclado bien en todo el conjunto del soporte de la distribución objetivo, entonces la varianza  $W$  de la muestra subestima la  $V(\psi)$ . Cuando  $n \rightarrow \infty$ , el valor esperado de  $\widehat{V}(\psi)$  converge a  $V(\psi)$  por la derecha y  $W$  converge a  $V(\psi)$  por la izquierda. Si  $\widehat{V}(\psi)$  es grande en relación con  $W$ , esto sugiere que la cadena aún no converge a la distribución objetivo en la iteración  $n$  (Rizzo, 2008, pág. 267).

5. Finalmente se estima el factor de reducción de escala potencial,  $R$ :

$$\sqrt{\widehat{R}} = \sqrt{\frac{\widehat{V}(\psi)}{W}}$$

que puede ser interpretado como la medición del factor por el cual la desviación estándar de  $\psi$  podría reducirse al extender la cadena. El factor  $\sqrt{\widehat{R}}$  tiende a 1 cuando la cadena tiende a infinito, entonces  $\sqrt{\widehat{R}}$  debería estar cerca de 1 si las cadenas convergen aproximadamente a la distribución objetivo. El valor de  $\widehat{R}$  debería ser menor que 1.1 o 1.2, sugerencia de Gelman (1996) citado por Rizzo (2008).

## 2.2.4. Implementación de métodos MCMC en JAGS

JAGS (Just Another Gibbs Sampler) es un programa de gran utilidad para implementar modelos Bayesianos mediante aproximación por Cadenas de Markov Monte Carlo. JAGS fue desarrollado para trabajar estrechamente con el software estadístico R, al utilizar los paquetes `rjags`, `R2jags` y `runjags`. JAGS fue escrito bajo los siguientes tres objetivos:

- Tener un motor para el lenguaje BUGS que sea ejecutable en Unix, Mac y Windows.
- Ser extensible, es decir, permitir a los usuarios escribir sus propias funciones, distribuciones y muestreadores.
- Ser una plataforma para la experimentación de modelos Bayesianos.

Ejecutar un modelo en el programa JAGS, se refiere a generar muestras de la distribución posterior de los parámetros involucrados en el modelo. La ejecución de un modelo se realiza en cinco pasos (Plummer, 2013):

1. Definición del modelo.
2. Compilación.
3. Inicialización.
4. Adaptación y burn-in.
5. Monitoreo.

1. Definición del modelo.

Se realiza en dos partes:

- Descripción del modelo. El modelo se define en un archivo de texto al usar un dialecto del lenguaje BUGS. Se define con la palabra *model* seguido de llaves {}, dentro de las llaves se escriben una serie de relaciones denominados nodos, que están en términos de otros nodos. Las relaciones pueden

ser estocásticas o deterministas; el símbolo ( $\sim$ ) define nodos estocásticos y representan una variable aleatoria, el símbolo ( $< -$ ) define un nodo determinista, es decir, que el valor está determinado por ciertos valores específicos.

- Definición de los datos. Los datos se definen en un archivo separado de la descripción del modelo. Los valores de datos se pueden suministrar para nodos estocásticos y para constantes (incluidos los valores constantes utilizados dentro de los bucles for). Se genera un error cuando se proporciona un valor de datos a un nodo determinista.

## 2. Compilación.

Al compilar un modelo, se crea un gráfico en la memoria de la computadora que lo representa. La compilación del modelo puede fallar cuando:

- El gráfico contiene un ciclo dirigido, debido a que estos están prohibidos en JAGS.
- Un parámetro de nivel superior no está definido. Cualquier nodo que se usa en el lado derecho de una relación, pero no está definida en el lado izquierdo de ninguna relación, es un nodo constante y su valor debe ser proporcionado en el archivo de datos.
- El modelo usa una función o distribución que no se haya definido.

En el proceso de compilación, se debe especificar el número de cadenas paralelas que ejecutará JAGS, dado que por defecto sólo ejecuta una cadena.

## 3. Inicialización.

Antes de que se ejecute un modelo, debe inicializarse y para esto hay tres pasos:

- a) Se establecen los valores iniciales de los parámetros del modelo.
- b) Se elige un generador de números aleatorios (RNG, por sus siglas en inglés), para cada cadena paralela y se establece una semilla.
- c) Los Samplers se eligen automáticamente para todos los parámetros del modelo. Un Sampler es un objeto que actúa sobre un conjunto de parámetros y los actualiza de una iteración a la siguiente.

#### 4. Adaptación y burn-in.

En teoría, la muestra generada por MCMC converge a la distribución objetivo en el límite, cuando el número de iteraciones tiende a infinito, es decir, la muestra converge a la distribución posterior de los parámetros del modelo; pero en la práctica todas las ejecuciones de MCMC son finitas. Por convención, se genera una muestra por MCMC con un número de iteraciones suficientemente grande y se divide en dos partes: una primera parte de período inicial (burn-in) que se descarta, y el resto de la ejecución en la que se considera que la cadena converge (suficientemente cerca) a la distribución objetivo. La muestra de la segunda parte de la ejecución, se utiliza para estimar estadísticos que resumen la distribución objetivo.

#### 5. Monitoreo.

En JAGS un monitor es un objeto que registra valores muestreados. El monitor más simple es una traza, que almacena el valor muestreado de un nodo en cada iteración. JAGS no puede monitorear un nodo a menos que se haya definido en el archivo del modelo. Para nodos con valores vectoriales o de matriz, cada elemento debe estar definido.

## 2.3. Estimación de parámetros

### 2.3.1. Estimadores puntuales

Una vez que se obtiene la distribución posterior del parámetro de interés, dado los datos, ya sea por cálculos analíticos o por aproximación MCMC; esta distribución se utiliza para realizar inferencias acerca del parámetro. Un estimador puntual del parámetro que resuma toda la información de la distribución posterior, se puede obtener al considerar una función de pérdida  $\ell(T, \theta)$ , la cual es una función no negativa que indica la pérdida que se tiene cuando se considera un estimador  $T$  en lugar del verdadero valor del parámetro  $\theta$ . Algunas de las funciones de pérdida más utilizadas son:

- Pérdida cuadrática  $\ell(T, \theta) = (T - \theta)^2$

- Pérdida absoluta  $\ell(T, \theta) = |T - \theta|$
- Pérdida (todo o nada), donde  $\ell(T, \theta) = 0$  si  $(T - \theta) \leq \epsilon$  o  $\ell(T, \theta) = 1$  si  $(T - \theta) > \epsilon$ , para un  $\epsilon > 0$ .

Cada una de estas funciones de pérdidas conducen a diferentes estimadores puntuales Bayesianos, cuando se minimiza la pérdida esperada  $E_\theta[\ell(T, \theta)]$ :

$$\begin{aligned} \int_{\Theta} E_\theta[\ell(T, \theta)]\pi(\theta)d\theta &= \int_{\Theta} \left[ \int_Y \ell(T, \theta)\pi(\mathbf{y}|\theta) \prod_{i=1}^n dy_i \right] \pi(\theta)d\theta \\ &= \int_Y \left[ \int_{\Theta} \ell(T, \theta)\pi(\theta|\mathbf{y})d\theta \right] \pi(\mathbf{y}) \prod_{i=1}^n dy_i \end{aligned}$$

Minimizar la doble integral es equivalente a minimizar la pérdida esperada posterior,  $E_\theta[\ell(T, \theta)|\mathbf{y}]$ , (Mood *et al.*, 1964, pág. 343 - 346):

$$E_\theta[\ell(T, \theta)|\mathbf{y}] = \int_{\Theta} \ell(T, \theta)\pi(\theta|\mathbf{y})d\theta$$

En general, el estimador Bayesiano de  $\theta$  es el valor de  $T = t \in \Theta$  que minimiza la pérdida esperada posterior:

$$t = \min_{T \in \Theta} E_\theta[\ell(T, \theta)|\mathbf{y}]$$

Al utilizar la función de pérdida cuadrática, el estimador Bayesiano es la media posterior (Gómez y Delicado, 2006 pág. 80):

$$\begin{aligned} E[\ell(T, \theta)|\mathbf{y}] &= E[(T - \theta)^2|\mathbf{y}] = E[(T^2 - 2\theta + \theta^2)|\mathbf{y}] \\ &= E[T^2|\mathbf{y}] - 2E[T\theta|\mathbf{y}] + E[\theta^2|\mathbf{y}] \\ &= T^2 - 2TE[\theta|\mathbf{y}] + E[\theta^2|\mathbf{y}] \end{aligned}$$

Al derivar respecto a  $T$  e igualar a cero se obtiene:

$$2T - 2E[\theta|\mathbf{y}] + 0 = 0$$

Entonces  $T = E[\theta|\mathbf{y}] = t$  el cual es mínimo porque la segunda derivada es  $2 > 0$ .

Al minimizar el valor esperado de la función de pérdida absoluta, el estimador Bayesiano es la mediana posterior (Nájera, 2015, pág. 14):

$$E[\ell(T, \theta)|\mathbf{y}] = \int_{\Theta} \ell(T, \theta)\pi(\theta|\mathbf{y})d\theta = \int_{\Theta} |T - \theta|\pi(\theta|\mathbf{y})d\theta$$

donde

$$|T - \theta| = \begin{cases} T - \theta & \text{si } T > \theta \\ -(T - \theta) & \text{si } T \leq \theta \end{cases}$$

Luego

$$\begin{aligned} \int_{-\infty}^{\infty} |T - \theta|\pi(\theta|\mathbf{y})d\theta &= \int_{-\infty}^T (T - \theta)\pi(\theta|\mathbf{y})d\theta + \int_T^{\infty} -(T - \theta)\pi(\theta|\mathbf{y})d\theta \\ &= \int_{-\infty}^T T\pi(\theta|\mathbf{y})d\theta - \int_{-\infty}^T \theta\pi(\theta|\mathbf{y})d\theta \\ &\quad - \int_T^{\infty} T\pi(\theta|\mathbf{y})d\theta + \int_T^{\infty} \theta\pi(\theta|\mathbf{y})d\theta \\ &= T \int_{-\infty}^T \pi(\theta|\mathbf{y})d\theta - \int_{-\infty}^T \theta\pi(\theta|\mathbf{y})d\theta \\ &\quad - T \int_T^{\infty} \pi(\theta|\mathbf{y})d\theta + \int_T^{\infty} \theta\pi(\theta|\mathbf{y})d\theta \end{aligned}$$

Al derivar

$$\begin{aligned} \frac{d}{dT} \left( \int_{-\infty}^{\infty} |T - \theta|\pi(\theta|\mathbf{y})d\theta \right) &= \int_{-\infty}^T \pi(\theta|\mathbf{y})d\theta + T\pi(T|\mathbf{y}) - T\pi(T|\mathbf{y}) - T\pi(T|\mathbf{y}) \\ &\quad - \int_T^{\infty} \pi(\theta|\mathbf{y})d\theta + T\pi(T|\mathbf{y}) \\ &= \int_{-\infty}^T \pi(\theta|\mathbf{y})d\theta - \int_T^{\infty} \pi(\theta|\mathbf{y})d\theta \end{aligned}$$

Si

$$\frac{d}{dT} \left( \int_{-\infty}^{\infty} |T - \theta|\pi(\theta|\mathbf{y})d\theta \right) = 0$$

Entonces

$$\int_{-\infty}^T \pi(\theta|\mathbf{y})d\theta = \int_T^{\infty} \pi(\theta|\mathbf{y})d\theta$$

de donde

$$\int_{-\infty}^T \pi(\theta|\mathbf{y})d\theta = \frac{1}{2}$$

que es mínimo porque

$$\begin{aligned} \frac{d^2}{dT^2} \left( \int_{-\infty}^{\infty} |T - \theta| \pi(\theta|\mathbf{y}) d\theta \right) &= \frac{d}{dT} \left( \int_{-\infty}^T \pi(\theta|\mathbf{y}) d\theta - \int_T^{\infty} \pi(\theta|\mathbf{y}) d\theta \right) \\ &= \pi(T|\mathbf{y}) + \pi(T|\mathbf{y}) = 2\pi(T|\mathbf{y}) > 0 \end{aligned}$$

Por tanto el valor  $t$  que minimiza la pérdida absoluta es la mediana posterior.

Y al minimizar el valor esperado de la función de pérdida todo o nada, el estimador Bayesiano es la moda posterior.

### 2.3.2. Intervalos creíbles

Un conjunto creíble para un parámetro  $\theta$ , es el conjunto de valores, tal que  $\theta$  se encuentra dentro de este conjunto, con probabilidad  $1 - \alpha$ . La densidad posterior se usa para cuantificar la probabilidad de que  $\theta$  se encuentre en un conjunto creíble, es decir, un intervalo creíble del  $100(1 - \alpha)\%$  para  $\theta$  es un subconjunto  $C \subset \Theta$  tal que (Fox, 2010 pág. 58-59):

$$P(\theta \in C|\mathbf{y}) = \int_C \pi(\theta|\mathbf{y}) d\theta = 1 - \alpha$$

Un defecto de un conjunto creíble es que no especifica si los valores de  $\theta$  dentro del conjunto son más probables que los valores fuera del conjunto. Por lo que es preferible elegir un conjunto creíble de valores con la densidad posterior más alta. Una extensión del conjunto creíble es el intervalo de densidad posterior más alta (*HPD*, por sus siglas en inglés) que cubre al menos una probabilidad de  $1 - \alpha$  y contiene los valores más probables de  $\theta$ ,  $P(\theta \in C|\mathbf{y}) \geq 1 - \alpha$ , y para un  $\theta_1 \in C$  y  $\theta_2 \notin C$  se sigue que  $p(\theta_1|\mathbf{y}) \geq p(\theta_2|\mathbf{y})$ . La densidad posterior de cada punto dentro del *HPD* es mayor que la probabilidad de cada punto fuera del *HPD* (Fox, 2010 pág. 58-59).

El intervalo creíble con probabilidad  $1 - \alpha$ , en una distribución posterior unimodal simétrica, es el intervalo de densidad posterior más alta, que se define al tomar los cuantiles  $\alpha/2$  y  $1 - \alpha/2$  de la distribución posterior. Para un vector multidimensional de parámetros, el conjunto creíble se define como una región creíble en lugar de un intervalo creíble.

### 2.3.3. Criterios para evaluar estimadores

Cuando se estudian varios estimadores, es conveniente analizar algunos criterios de evaluación, para compararlos y determinar que estimador cumple con suficientes expectativas para considerarse como el mejor. Es conocido en estadística frecuentista, que un buen estimador es el que es insesgado y de mínima varianza, el Error Cuadrático Medio, es la medida que considera estas dos propiedades del estimador.

En el presente trabajo, se consideran algunos criterios de evaluación, para las estimaciones obtenidas en la simulación con cada distribución inicial. Los criterios que se consideran son: la Esperanza, el Sesgo, el Error Cuadrático Medio, la Longitud y el Cubrimiento de los intervalos creíbles.

Los criterios de evaluación se obtienen a partir de la simulación de 1000 repeticiones del proceso MCMC. En el presente trabajo,  $\tilde{\beta}_k$  es la estimación del parámetro  $\beta_k$ , donde  $k = 1, \dots, p$  y  $p$  es el número de variables independientes usadas en el modelo. Dado que se obtienen 1000 estimaciones para cada parámetro, entonces  $\tilde{\beta}_k^{(r)}$  es la estimación del parámetro  $\beta_k$  en la  $r$ -ésima repetición,  $r = 1, \dots, 1000$ .

Los criterios de evaluación se definen a continuación.

- La Esperanza estimada es el valor promedio sobre toda la muestra de estimaciones  $\tilde{\beta}_k$  del parámetro  $\beta_k$ .

$$\widehat{E}(\tilde{\beta}_k) = \frac{1}{1000} \sum_{r=1}^{1000} \tilde{\beta}_k^{(r)}$$

- El Sesgo estimado es el valor promedio de la diferencia de la estimación y el verdadero valor del parámetro.

$$\widehat{S}(\tilde{\beta}_k) = \frac{1}{1000} \sum_{r=1}^{1000} (\tilde{\beta}_k^{(r)} - \beta_k) = E(\tilde{\beta}_k^{(r)}) - \beta_k$$

- El Error Cuadrático Medio estimado es el valor promedio del cuadrado de la diferencia de la estimación y el verdadero valor del parámetro.

$$\widehat{ECM}(\tilde{\beta}_k) = \frac{1}{1000} \sum_{r=1}^{1000} (\tilde{\beta}_k^{(r)} - \beta_k)^2 = \widehat{Var}(\tilde{\beta}_k^{(r)}) + \widehat{S}(\tilde{\beta}_k)^2$$

- La Longitud estimada del intervalo creíble ( $\widehat{L}$ ) es el promedio de la diferencia del valor del límite superior (LS) y el valor del límite inferior (LI), del intervalo creíble de la estimación.

$$\widehat{L}(\widetilde{\beta}_k) = \frac{1}{1000} \sum_{r=1}^{1000} [LS(\widetilde{\beta}_k^{(r)}) - LI(\widetilde{\beta}_k^{(r)})]$$

- El Cubrimiento,  $C$ , es la proporción de veces que los intervalos creíbles del 95 % contienen al parámetro. El Cubrimiento estimado,  $\widehat{C}$ , es el promedio sobre toda la muestra de  $C_k^{(r)}$ , donde  $C_k^{(r)} = 1$  si en la  $r$ -ésima repetición, con probabilidad de 0.95, el parámetro  $\beta_k$  está contenido en el intervalo creíble estimado y  $C_k^{(r)} = 0$  en otro caso.

$$C_k^{(r)} = I[LI(\widetilde{\beta}_k^{(r)}) \leq \beta_k \leq LS(\widetilde{\beta}_k^{(r)})]$$

$$\widehat{C}(\widetilde{\beta}_k) = \frac{1}{1000} \sum_{r=1}^{1000} C_k^{(r)}$$

donde  $I[\cdot]$  es la función indicadora.

### Metodología

---

En el presente capítulo se describen; el mecanismo aleatorio usado para el diseño de respuesta forzada, generalizado a una pregunta sobre un tema sensitivo con respuesta ordinal; el modelo de regresión probit ordinal; la simulación de las observaciones para las variables independientes y para la variable dependiente con tres y cuatro categorías ordinales; finalmente, el estudio de simulación implementado en JAGS, para aproximar la distribución posterior de los parámetros del modelo.

#### **3.1. Diseño de respuesta forzada para variables con más de dos categorías**

Para el diseño de respuesta forzada generalizado a variables multinomiales, Liu y Chow (1976) y Kim y Warde (2005) utilizaron el mecanismo aleatorio de Hopkins. Liu y Chow (1976), lo usaron para un modelo discreto de respuesta aleatoria cuantitativo y Kim y Warde (2005) también lo usaron para variables respuesta que siguen una distribución multinomial.

En el presente trabajo también se utiliza el mecanismo aleatorio de Hopkins; el cual consiste en colocar en una botella, frasco o matraz redondo, bolas de dos colores dife-

rentes, por ejemplo en este trabajo se utilizó una proporción  $p$  de bolas verdes y una proporción  $1 - p$  de bolas azules, donde las bolas azules se marcan con proporciones  $q_j$ , no necesariamente iguales, con los números del 1 al  $J$ , dependiendo del número de categorías que tenga la variable dependiente. En este trabajo, las proporciones  $q_j$  se consideraron iguales, es decir,  $q = 1/3$  y  $q = 1/4$ , cuando la variable dependiente  $Y$  tiene  $J = 3$  o  $J = 4$  categorías ordinales respectivamente.

Al utilizar este mecanismo aleatorio, el entrevistador debe dar instrucciones al encuestado de agitar el recipiente, colocarla boca abajo y permitir que salga o se acerque una bola al cuello del recipiente sin mencionar el color de dicha bola para proteger su privacidad. Si el resultado del mecanismo es una bola color verde, el encuestado debe responder honestamente la pregunta sobre el tema sensible con el número  $j = 1, \dots, J$  que corresponda, y si el resultado es una bola color azul debe responder el número que está marcado en dicha bola. Esto produce que el mecanismo aleatorio permita obtener respuestas honestas con probabilidad  $p$  y con probabilidad  $1 - p$  obliga al encuestado a dar una respuesta forzada independientemente si pertenece o no al número de categoría que responde.

Para este mecanismo sea  $Z$  la variable latente con tres categorías ordinales y  $Y$  la variable observada tal que:

$$Y_i = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo responde 1} \\ 2 & \text{si el } i\text{-ésimo individuo responde 2} \\ 3 & \text{si el } i\text{-ésimo individuo responde 3} \end{cases}$$

$$Z_i = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo pertenece a la categoría indizada por 1} \\ 2 & \text{si el } i\text{-ésimo individuo pertenece a la categoría indizada por 2} \\ 3 & \text{si el } i\text{-ésimo individuo pertenece a la categoría indizada por 3} \end{cases}$$

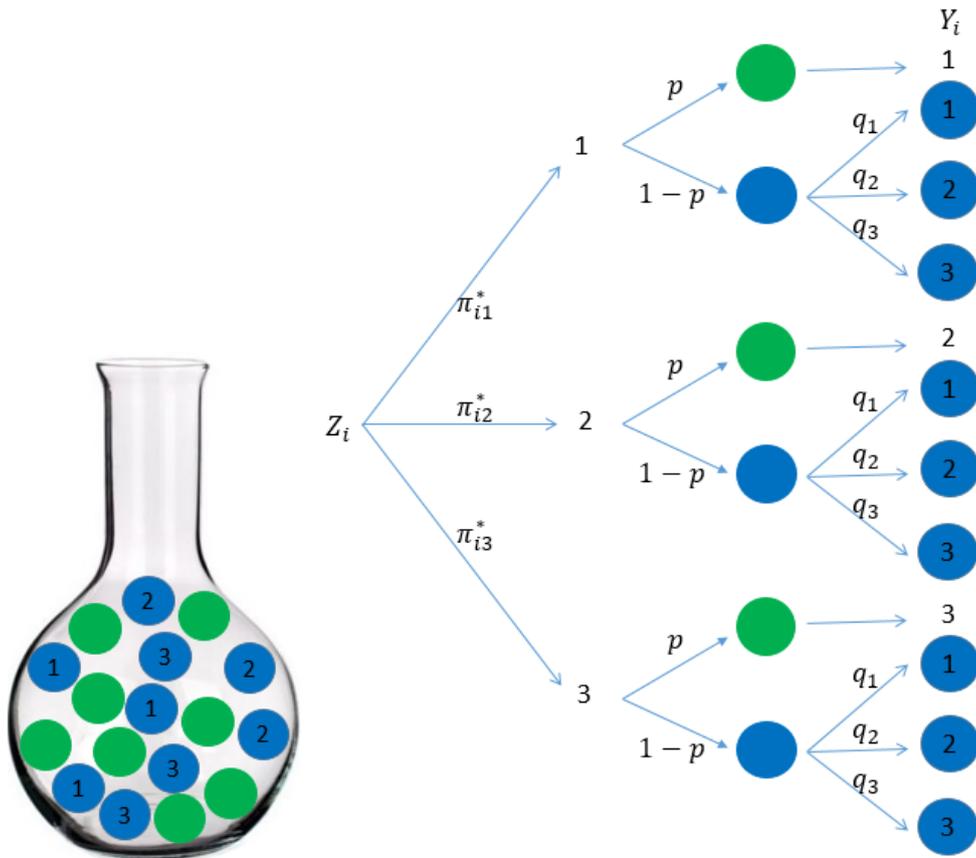


Figura 3.1: Diagrama del mecanismo aleatorio para una pregunta con tres categorías.

De manera similar si  $Y$  tiene 4 categorías ordinales:

$$Y_i = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo responde 1} \\ 2 & \text{si el } i\text{-ésimo individuo responde 2} \\ 3 & \text{si el } i\text{-ésimo individuo responde 3} \\ 4 & \text{si el } i\text{-ésimo individuo responde 4} \end{cases}$$

$$Z_i = \begin{cases} 1 & \text{si el } i\text{-ésimo individuo pertenece a la categoría indizada por 1} \\ 2 & \text{si el } i\text{-ésimo individuo pertenece a la categoría indizada por 2} \\ 3 & \text{si el } i\text{-ésimo individuo pertenece a la categoría indizada por 3} \\ 4 & \text{si el } i\text{-ésimo individuo pertenece a la categoría indizada por 4} \end{cases}$$

Mediante el mecanismo aleatorio, la probabilidad de respuesta latente para una pregunta con  $J$  categorías ordinales,  $\pi_{ij}^* = P(Z_i = j)$ ,  $i = 1, \dots, n$  y  $j = 1, \dots, J$  está

relacionada con la probabilidad de respuesta observada  $\pi_{ij} = P(Y_i = j)$  mediante:

$$\pi_{ij} = p\pi_{ij}^* + (1 - p)q \quad i = 1, \dots, n \quad j = 1, \dots, J$$

donde  $p$  es la probabilidad de responder honestamente el número de la categoría de la pregunta sobre el tema sensitivo y  $q$  la proporción de bolas azules marcadas del 1 al 3 o del 1 al 4 según el caso.

## 3.2. Modelo de regresión probit ordinal

Entre algunos de los modelos que se utilizan para explicar la relación de la variable ordinal con  $p$  variables independientes que miden características de los individuos, está el modelo probit ordinal en el que se considera que la probabilidad de un resultado ordinal en particular es el área bajo la curva normal entre dos puntos de corte en el que se encuentra. Esta probabilidad se calcula al considerar el área acumulada bajo la distribución normal hasta el punto de corte de la derecha menos el área acumulada bajo la normal al punto de corte de la izquierda (Kruschke, 2014 pág. 674).

$$P(Z = j \mid \mu, \sigma, \gamma_{j-1}, \gamma_j) = \Phi\left(\frac{\gamma_j - \mu}{\sigma}\right) - \Phi\left(\frac{\gamma_{j-1} - \mu}{\sigma}\right) \quad j = 1, \dots, J$$

donde  $\gamma_j$ ,  $\mu$  y  $\sigma$  son respectivamente los puntos de corte, la media y la desviación estándar en la distribución Normal. Se definen de manera implícita el punto de corte  $\gamma_0 = -\infty$  y  $\gamma_J = \infty$ , para que tengan sentido la probabilidad del menor y mayor resultado ordinal, las cuales están respectivamente dadas por:

$$\begin{aligned} P(Z = 1 \mid \mu, \sigma, \gamma_1) &= \Phi\left(\frac{\gamma_1 - \mu}{\sigma}\right) - \Phi\left(\frac{\gamma_0 - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{\gamma_1 - \mu}{\sigma}\right) - \Phi\left(\frac{-\infty - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{\gamma_1 - \mu}{\sigma}\right) = \Phi\left(\frac{\gamma_1 - \mu}{\sigma}\right) \end{aligned}$$

$$\begin{aligned}
P(Z = J \mid \mu, \sigma, \gamma_{J-1}) &= \Phi\left(\frac{\gamma_J - \mu}{\sigma}\right) - \Phi\left(\frac{\gamma_{J-1} - \mu}{\sigma}\right) \\
&= \Phi\left(\frac{\infty - \mu}{\sigma}\right) - \Phi\left(\frac{\gamma_{J-1} - \mu}{\sigma}\right) \\
&= 1 - \Phi\left(\frac{\gamma_{J-1} - \mu}{\sigma}\right)
\end{aligned}$$

### 3.2.1. Modelo de regresión probit ordinal aleatorizado

Para determinar el efecto de las variables independientes en la variable  $Y$  aleatorizada, dado que la respuesta es ordinal, se propone el modelo de regresión probit ordinal aleatorizado para la probabilidad  $\pi_{ij}^*$ .

Sea  $\mu = \mathbf{x}_i^T \boldsymbol{\beta}$  el predictor lineal, donde  $\mathbf{x}_i^T$  es el  $i$ -ésimo renglón de la matriz diseño  $\mathbf{X}$ , es decir,  $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$  es el vector de valores observados para el individuo  $i$  en cada una de las variables independientes  $\mathbf{x}_1, \dots, \mathbf{x}_p$ ;  $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_p)$  es el vector de parámetros asociados a las variables independientes y  $\boldsymbol{\gamma}^T = (\gamma_0, \dots, \gamma_J)$  el vector de puntos de corte. Entonces el modelo para  $P(Y_i = j)$  con  $i = 1, \dots, n$  y  $j = 1, \dots, J$  el número de categorías de  $Y$  esta dado por:

$$\begin{aligned}
\pi_{ij} = P(Y_i = j) &= p\pi_{ij}^* + (1-p)q \\
&= p \left[ \Phi\left(\frac{\gamma_j - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) - \Phi\left(\frac{\gamma_{j-1} - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \right] + (1-p)q
\end{aligned}$$

Es conocido por teoría Bayesiana que la distribución posterior conjunta de los parámetros es proporcional al producto de la función de verosimilitud y la distribución inicial conjunta. Entonces, la distribución posterior conjunta de  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$  y  $\sigma$  es:

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma \mid \mathbf{X}, \mathbf{Y}) \propto \pi(\mathbf{X}, \mathbf{Y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma) \times \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma)$$

La función de verosimilitud para variables dependientes con  $J$  categorías ordinales y  $n$  observaciones, está dada de la forma:

$$\begin{aligned}
\pi(\mathbf{X}, \mathbf{Y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma) &= \prod_{i=1}^n \sum_{j=1}^J I(Y_i = j) \pi_{ij} \\
&\prod_{i=1}^n \sum_{j=1}^J I(Y_i = j) [p\pi_{ij}^* + (1-p)q] \\
&\prod_{i=1}^n \sum_{j=1}^J I(Y_i = j) \left[ p \left[ \Phi \left( \frac{\gamma_j - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) - \Phi \left( \frac{\gamma_{j-1} - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] + (1-p)q \right]
\end{aligned}$$

La distribución inicial conjunta de los parámetros es el producto de distribuciones marginales, es decir, los parámetros se suponen independientes:  $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma) = \pi(\boldsymbol{\beta}) \times \pi(\boldsymbol{\gamma}) \times \pi(\sigma)$ . De manera similar para la distribución inicial conjunta de los  $\beta_k$ ,  $k = 1, \dots, p$ :

$$\pi(\boldsymbol{\beta}) = \pi(\beta_0) \prod_{k=1}^p \pi(\beta_k)$$

Kruschke (2014) considera una variable ordinal con  $J$  categorías, en que se fija el punto de corte  $\gamma_1$  y el  $\gamma_{J-1}$  y para los puntos de corte que están entre éstos dos se asigna la distribución inicial Normal con media  $j + 0.5$  y precisión  $\tau = \frac{1}{\sigma^2} = \frac{1}{2^2}$ , y para  $\sigma$  asigna una distribución inicial Uniforme en el intervalo  $(\frac{J}{1000}, 3J)$ , donde  $J$  es el número de categorías de la variable ordinal. Al tomar como referencia a Kruschke (2014) cuando se modela la variable ordinal con 3 categorías, los dos puntos de corte  $\gamma_1$  y  $\gamma_2$  que aparecen en el modelo quedan determinados como fijos, por lo que sólo se estimaron el vector de parámetros  $\boldsymbol{\beta}$  y  $\sigma$ . Así la distribución inicial conjunta es

$$\pi(\boldsymbol{\beta}, \sigma) = \pi(\beta_0) \prod_{k=1}^p \pi(\beta_k) \times \pi(\sigma)$$

Y cuando se modela la variable ordinal con 4 categorías se fijan los puntos de cortes  $\gamma_1$  y  $\gamma_3$  y se asigna a  $\gamma_2$  la distribución inicial Normal con media 2.5 y precisión  $1/2^2$ , así la distribución inicial conjunta es el producto

$$\pi(\boldsymbol{\beta}, \gamma_2, \sigma) = \pi(\beta_0) \prod_{k=1}^p \pi(\beta_k) \times \pi(\gamma_2) \times \pi(\sigma)$$

Las variables independientes fueron estandarizadas para reducir la autocorrelación en el muestreo MCMC y mejorar la convergencia de las cadenas. La variable dependiente  $Y$  no se estandariza porque sus valores son ordinales, es decir, sólo son categorías

ordenadas, datos que no se miden en una escala métrica, además sus probabilidades son descritas por la distribución categórica *dcat* en JAGS la cual es una distribución sobre índices enteros, por lo que sólo se estandarizan las variables independientes porque son datos medidos en una escala métrica. Estas variables estandarizadas son denotadas a lo largo del trabajo por  $x_{kz}$  (Kruschke, 2014, pág. 624 - 625).

En el modelo se usan los datos estandarizados y luego mediante JAGS se encuentran valores creíbles para los parámetros asociados a las variables estandarizadas, los cuales son  $\alpha_0$  y  $\alpha_k$ , luego se transforman los valores de los parámetros estandarizados  $\alpha_0$  y  $\alpha_k$ , a parámetros originales  $\beta_0$  y  $\beta_k$  (Kruschke, 2014, pág. 624 - 625):

$$\beta_0 = \alpha_0 - \sum_{k=1}^p \frac{\alpha_k}{s_{x_k}} \bar{x}_k \quad \beta_k = \frac{\alpha_k}{s_{x_k}} \quad k = 1, \dots, p$$

donde  $\bar{x}_k$  y  $s_{x_k}$  son la media y la desviación estándar de la  $k$ -ésima variable independiente  $x$ . Este proceso se realiza en cada paso del muestreo MCMC.

En este trabajo se estudian estimadores bajo cuatro distribuciones iniciales para los parámetros estandarizados  $\alpha_k$ ,  $k = 1, \dots, p$ ; la distribución Normal, la Doble Exponencial, la *t* de Student y la Cauchy. Para los parámetros  $\sigma$  y  $\gamma_2$  se asignan la distribuciones iniciales Uniforme y Normal respectivamente (Kruschke, 2014 pág. 677).

Las distribuciones posteriores conjuntas se obtienen mediante aproximación por muestreo MCMC en JAGS. A continuación se presentan las distribuciones iniciales y las distribuciones posteriores conjuntas, el desarrollo completo de las ecuaciones de las distribuciones posteriores conjuntas se presentan en el Anexo 1.

1. Distribución inicial Normal para  $\alpha_k$ :

$$\alpha_0 \sim N\left(\frac{1+J}{2}, \frac{1}{J^2}\right) \quad \alpha_k \sim N\left(0, \frac{1}{J^2}\right)$$

$$\gamma_j \sim N\left(j + 0.5, \frac{1}{2^2}\right) \quad \sigma \sim U\left(\frac{J}{1000}, 10J\right)$$

donde  $J=3$  o  $4$  según el número de categorías de  $Y$ .

Cuando  $Y$  tiene 3 categorías la distribución posterior es:

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &\propto \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\ &\times \exp \left\{ -\frac{(\alpha_0^2 - 4\alpha_0 + \sum_{k=1}^p \alpha_k^2)}{18} \right\} \\ &\times \sigma^{-n} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \end{aligned}$$

Cuando  $Y$  tiene 4 categorías la distribución posterior es:

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \gamma_2, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &\propto \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\ &\times \exp \left\{ -\frac{(\alpha_0^2 - 5\alpha_0 + \sum_{k=1}^p \alpha_k^2)}{32} - \frac{(\gamma_2^2 - 5\gamma_2)}{8} \right\} \\ &\times \sigma^{-n} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \end{aligned}$$

2. Distribución inicial Doble Exponencial para  $\alpha_k$ :

$$\begin{aligned} \alpha_0 &\sim N\left(\frac{1+J}{2}, \frac{1}{J^2}\right) & \alpha_k &\sim DE(0, \lambda) & \lambda &\sim U(0.001, 10) \\ \gamma_j &\sim N\left(j + 0.5, \frac{1}{2^2}\right) & \sigma &\sim U\left(\frac{J}{1000}, 10J\right) \end{aligned}$$

Cuando  $Y$  tiene 3 categorías la distribución posterior es:

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &\propto \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\ &\times \left(\frac{1}{\lambda}\right)^p \exp \left\{ -\frac{\sum_{k=1}^p |\alpha_k|}{\lambda} - \frac{(\alpha_0^2 - 4\alpha_0)}{18} \right\} \\ &\times \sigma^{-n} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \end{aligned}$$

Cuando  $Y$  tiene 4 categorías la distribución posterior es:

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \gamma_2, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &\propto \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\ &\times \left( \frac{1}{\lambda} \right)^p \exp \left\{ -\frac{\sum_{k=1}^p |\alpha_k|}{\lambda} - \frac{(\alpha_0^2 - 5\alpha_0)}{32} \right\} \\ &\times \sigma^{-n} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} - \frac{(\gamma_2^2 - 5\gamma_2)}{8} \right\} \end{aligned}$$

3. Distribución inicial t de Student para  $\alpha_k$ :

$$\begin{aligned} \alpha_0 &\sim N\left(\frac{1+J}{2}, \frac{1}{J^2}\right) & \alpha_k &\sim t(0, \sigma_B^2, n^* + 1) & n^* &\sim \text{Exp}\left(\frac{1}{29}\right) \\ \gamma_j &\sim N\left(j + 0.5, \frac{1}{2^2}\right) & \sigma &\sim U\left(\frac{J}{1000}, 10J\right) & \sigma_B &\sim U(0.001, 1000) \end{aligned}$$

Cuando  $Y$  tiene 3 categorías la distribución posterior es:

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &\propto \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\ &\times \exp \left\{ -\frac{(\alpha_0^2 - 4\alpha_0)}{18} - \frac{n^*}{29} \right\} \prod_{k=1}^p \left( 1 + \frac{\alpha_k^2}{(n^* + 1)\sigma_B^2} \right)^{-\frac{n^*+2}{2}} \\ &\times \sigma^{-n} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \\ &\times \left( \frac{\Gamma(\frac{n^*+2}{2})}{\Gamma(\frac{n^*+1}{2}) \sqrt{(n^* + 1)\sigma_B^2}} \right)^p \end{aligned}$$

Cuando  $Y$  tiene 4 categorías la distribución posterior es:

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \gamma_2, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &\propto \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\ &\times \prod_{k=1}^p \left( 1 + \frac{\alpha_k^2}{(n^* + 1)\sigma_B^2} \right)^{-\frac{n^*+2}{2}} \left( \frac{\Gamma(\frac{n^*+2}{2})}{\Gamma(\frac{n^*+1}{2}) \sqrt{(n^* + 1)\sigma_B^2}} \right)^p \\ &\times \exp \left\{ -\frac{(\alpha_0^2 - 5\alpha_0)}{32} - \frac{n^*}{29} - \frac{(\gamma_2^2 - 5\gamma_2)}{8} \right\} \\ &\times \sigma^{-n} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \end{aligned}$$

4. Distribución inicial Cauchy para  $\alpha_k$ :

$$\alpha_0 \sim N\left(\frac{1+J}{2}, \frac{1}{J^2}\right) \quad \alpha_k \sim t(0, \sigma_B^2, 1)$$

$$\gamma_j \sim N\left(j + 0.5, \frac{1}{2^2}\right) \quad \sigma \sim U\left(\frac{J}{1000}, 10J\right) \quad \sigma_B \sim U(0.001, 1000)$$

Cuando  $Y$  tiene 3 categorías la distribución posterior es:

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &\propto \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\ &\times \sigma_B^{-p} \prod_{k=1}^p \left(1 + \frac{\alpha_k^2}{\sigma_B^2}\right)^{-1} \exp\left\{-\frac{(\alpha_0^2 - 4\alpha_0)}{18}\right\} \\ &\times \sigma^{-n} \exp\left\{-\frac{\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2}\right\} \end{aligned}$$

Cuando  $Y$  tiene 4 categorías la distribución posterior es proporcional a:

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \gamma_2, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &\propto \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\ &\times \sigma_B^{-p} \prod_{k=1}^p \left(1 + \frac{\alpha_k^2}{\sigma_B^2}\right)^{-1} \exp\left\{-\frac{(\alpha_0^2 - 5\alpha_0)}{32}\right\} \\ &\times \sigma^{-n} \exp\left\{-\frac{\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} - \frac{(\gamma_2^2 - 5\gamma_2)}{8}\right\} \end{aligned}$$

Se propusieron estas 4 distribuciones iniciales porque tienen soporte igual que el espacio paramétrico de los parámetros  $\alpha_k$ . La distribución inicial Normal generalmente se ha utilizado en modelos de regresión, además se considera como referencia Kruschke (2014); la distribución inicial Doble Exponencial se utiliza generalmente para considerar valores de los parámetros cercanos a cero; la distribución inicial t de Student se utiliza porque tiene colas más altas o pesadas que la distribución Normal, y esta acomoda más datos que tienen valores atípicos que por lo regular no acomoda la distribución Normal, los grados de libertad de esta distribución varían de 1 a  $\infty$ , cuando es 1 la distribución t de Student tiene colas aún más pesadas, y cuando es aproximadamente 30, la distribución t de Student es la normal (Kruschke, 2014, pág. 458); la distribución inicial Cauchy también se utiliza por tener colas pesadas y es equivalente a una distribución t de Student con un grado de libertad.

### 3.3. Estudio de simulación

Muchos factores influyen al momento de estimar los parámetros del modelo, como el tamaño de muestra, los valores verdaderos de los parámetros, la varianza de la distribución de los datos, las distribuciones iniciales, etc. En este trabajo la simulación para la modelación de la variable dependiente ordinal referente a un tema sensitivo, se basó en considerar los siguientes puntos de interés:

- La variable dependiente  $Y$  con 3 y 4 categorías.
- Se fijaron los valores considerados como parámetros verdaderos.
- Los tamaños de muestra se asignaron según la cantidad de parámetros en el modelo (Kruschke, 2014).
- Se consideraron una y dos variables independientes.
- Los puntos de cortes se seleccionaron de forma que las proporciones en las categorías fueran semejantes en los escenarios de acuerdo al número de categorías de  $Y$ .
- Se triplicó el tamaño de muestra para realizar una comparación entre los estimadores que se obtienen en las diferentes simulaciones.
- Se fijaron el valor de  $p$  y de  $q$  que respectivamente representan la proporción de bolas verdes y la proporción de bolas azules, marcadas con los números correspondientes a las categorías de la variable  $Y$ , según el mecanismo aleatorio de Hopkins.
- Se fijó la desviación estándar de la distribución de los datos.

Kruschke (2014) presentó un ejemplo en el que se modela una variable dependiente con 7 categorías ordinales y una variable independiente; para el modelo se consideró un tamaño de muestra de 200, y se desconocen 8 parámetros, por lo que aproximadamente se puede pensar que se consideró un tamaño de muestra de 25 observaciones por cada parámetro. En la simulación del presente trabajo, también se consideró el supuesto

de tomar 25 observaciones por cada parámetro del modelo, esto bajo 8 escenarios de simulación. Un primer escenario fue modelar una variable dependiente ordinal con tres categorías con una variable independiente, donde se fijaron los dos puntos de corte como 3.7 y 6.2, la probabilidad de seleccionar una bola verde en el mecanismo aleatorio como  $p = 2/3$ , la proporción de bolas azules marcadas con el número de la categoría como  $q = 1/3$ , el vector de parámetros que se supone como verdadero fue  $\beta = (-10, 10)$ , la desviación estándar de los datos como  $\sigma = 1/2$  y un tamaño de muestra de  $n = 125$  para considerar aproximadamente 25 observaciones por parámetro. El segundo escenario se consideró con los mismos supuestos que el primero, con excepción del tamaño de muestra que se triplicó, es decir,  $n = 375$ , para realizar una comparación entre las estimaciones que se obtuvieron en el primer y segundo escenario. Los valores de los parámetros que se consideraron en el resto de los escenarios se presentan en el Cuadro 3.1.

Cuadro 3.1: Escenarios de simulación

Número de categorías de $Y$	Número de variables independientes	
	1	2
3	$n = 125, 375$	$n = 150, 450$
	$\beta_0 = -10, \beta_1 = 10$	$\beta_0 = -10, \beta_1 = 10, \beta_2 = 5$
	$\gamma_1 = 3.7, \gamma_2 = 6.2$	$\gamma_1 = 10, \gamma_2 = 13.3$
	$p = 2/3$	$p = 2/3$
	$q = 1/3$	$q = 1/3$
	$\sigma = 1/2$	$\sigma = 1/2$
4	$n = 150, 450$	$n = 175, 525$
	$\beta_0 = -10, \beta_1 = 10$	$\beta_0 = -10, \beta_1 = 10, \beta_2 = 5$
	$\gamma_1 = 2.6, \gamma_2 = 5.0, \gamma_3 = 6.7$	$\gamma_1 = 9.6, \gamma_2 = 11.8, \gamma_3 = 13.5$
	$p = 2/3$	$p = 2/3$
	$q = 1/4$	$q = 1/4$
	$\sigma = 1/2$	$\sigma = 1/2$

Para la simulación se supone que las proporciones de respuestas es ligeramente mayor en la primera categoría y con menor proporción en la tercera o cuarta según el

total de categorías de la variable dependiente. Por ejemplo si la variable dependiente con tres categorías sobre el tema sensitivo es abortos que se ha practicado una joven durante el periodo de estudios de Licenciatura y las categorías de respuestas fueran respectivamente *Ningún aborto*, *un aborto* y *más de un aborto*, se esperaría que en algunos lugares contestarán con mayor proporción *Ningún aborto* y con menor proporción la respuesta de la tercera categoría *más de un aborto*.

La simulación de los valores para la variable dependiente ordinal se realizó al considerar la recomendación de Kruschke (2014) de la siguiente forma:

Para el primer escenario de simulación se consideró un tamaño de muestra de  $n = 125$  observaciones, una variable independiente y como vector de parámetros verdaderos  $\boldsymbol{\beta} = (-10, 10)$  (Kruschke, 2014). Para las variables independientes  $x_1$  y  $x_2$  se generaron  $n$  observaciones de la distribución Uniforme en el intervalo  $(0.9, 1.9)$ . Después se formó el predictor lineal que en este caso es  $\mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_1$  y se generaron valores para una variable continua  $Z_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$  con  $\varepsilon_i \sim N(0, \sigma^2)$  y  $\sigma = 1/2$ . La variable  $Z^*$  se categorizó para generar la variable latente  $Z$  con tres categorías. Para la variable dependiente  $Y$  con tres categorías se fijaron dos puntos de corte como 3.7 y 6.2 para considerar que aproximadamente en la primera categoría hubiera un 40% de observaciones y en la segunda y tercera un 30%, además se consideró  $\gamma_0 = -\infty$  para que  $\boldsymbol{\gamma} = (-\infty, 3.7, 6.2)$ , finalmente se generaron los valores de la variable latente  $Z$  con tres categorías que se determinaron como  $Z_i = j$ ,  $j = 1, 2, 3$  si  $Z_i^* > \gamma_j$ .

Una vez que se obtuvieron los valores 1, 2 y 3 que representan las categorías de  $Z$ , se generaron los valores para la variable observada  $Y$  de la siguiente forma:

Para cada observación  $Z_i = j$ ,  $j = 1, 2, 3$  se generó un número aleatorio Bernoulli con probabilidad de éxito  $p = 2/3$ , de manera que si el número aleatorio generado correspondía a un éxito, se asignó a  $Y_i$  el valor de  $Z_i = j$ ; en caso contrario se asignó a  $Y_i$  un número generado de una distribución multinomial  $(1, 2, 3)$  con probabilidad  $q = 1/3$  que representan las proporciones de bolas azules en el mecanismo aleatorio.

Para el caso cuando la variable dependiente tiene cuatro categorías ordinales, el procedimiento de generar los valores para las variables  $Z$  y  $Y$ , se realizó de forma similar al que se acaba de describir, donde se fijaron tres puntos de corte (Cuadro 3.1), de manera que se consideraron aproximadamente en la primera categoría un 35 % de observaciones, en la segunda un 25 % y en la tercera y cuarta un 20 %, además se fijó  $q = 1/4$  que representa la proporción de bolas azules que se marcaron con los números  $j = 1, \dots, 4$ , como fue descrito en el mecanismo aleatorio de Hopkins.

Cada uno de los 8 escenarios fue simulado en el software estadístico R (Core Team, 2016). Para las estimaciones de los parámetros, se consideró la media posterior de las aproximaciones a las distribuciones posteriores, obtenidas mediante JAGS (Plummer, 2013) con las 4 distribuciones iniciales, antes mencionadas. Para aproximar cada distribución posterior, al momento de compilar el modelo en JAGS, se especificaron tres Cadenas de Markov Monte Carlo, cada cadena con 10,000 iteraciones, de las cuales se descartaron las primeras 5000 iteraciones en cada una, y de las 5000 restantes se especificó un adelgazamiento de 5 unidades. El proceso fue repetido 1000 veces, en cada repetición se calcularon los estimadores y después se obtuvieron las estimaciones de la Esperanza, ECM, Sesgo, Longitud y Cubrimiento de los intervalos creíbles. Se verificó la convergencia de las cadenas de Markov, al utilizar el diagnóstico de convergencia de Gelman y Rubin (1992), el factor de reducción de escala potencial.

### Resultados

---

Se obtuvieron las estimaciones Bayesianas de todos los parámetros involucrados en el modelo de regresión probit ordinal aleatorizado, pero sólo se presentan las estimaciones de los parámetros  $\beta_k$  asociados a las variables independientes, porque es de interés ver el efecto de tales variables en la variable ordinal aleatorizada.

En el Cuadro 4.1 se presentan las estimaciones de Esperanza, Sesgo, ECM, Longitud y Cubrimiento de los intervalos creíbles, cuando se modela la variable aleatorizada ordinal con una variable independiente.

#### **Efecto de las distribuciones iniciales:**

En general, los estimadores Bayesianos son mejores con la distribución inicial Doble Exponencial; por lo que se comparan con los de las distribuciones Normal, t de Student y Cauchy para observar los efectos en Esperanza, Sesgo, ECM, Longitud y Cubrimiento de los intervalos creíbles (Cuadro 4.1).

La Esperanza de los estimadores Bayesianos con las distribuciones iniciales Normal, t de Student y Cauchy está entre 1 % y 6 % mayor que con la distribución inicial Doble Exponencial.

El Sesgo de los estimadores Bayesianos con las distribuciones iniciales Normal, t de Student y Cauchy está entre 22 % y 99 % mayor que con la distribución inicial Doble Exponencial. El Sesgo del estimador Bayesiano con la distribución inicial Doble Exponencial, fue cero en el caso de Y con  $J = 4$  categorías y  $n = 450$ , lo que muestra la superioridad del estimador Bayesiano con esta distribución inicial.

En general, el ECM de los estimadores Bayesianos con las distribuciones iniciales Normal, t de Student y Cauchy está entre 5 % y 84 % mayor que con la distribución inicial Doble Exponencial. Pero hubo un caso, en Y con  $J = 3$  y  $n = 125$ , en que el ECM del estimador Bayesiano con la distribución inicial Normal fue menor que con la distribución inicial Doble Exponencial y otros casos, en Y con  $J = 4$  y  $n = 450$ , en que el ECM de los estimadores Bayesianos es igual en las 4 distribuciones iniciales.

En general, la Longitud de los estimadores Bayesianos con las distribuciones iniciales Normal, t de Student y Cauchy está entre 2 % y 17 % mayor que con la distribución inicial Doble Exponencial. Pero hubo un caso, en Y con  $J = 3$  y  $n = 125$ , en que la Longitud del estimador Bayesiano con la distribución inicial Normal fue menor que con la distribución Doble Exponencial y otros casos; en Y con  $J = 4$  y  $n = 450$  en que la Longitud de los estimadores Bayesianos es igual en las 4 distribuciones iniciales y en Y con  $J = 3$  y  $n = 375$ , y en Y con  $J = 4$  y  $n = 150$  en que la Longitud de los estimadores Bayesianos es igual con las distribuciones iniciales t de Student y Cauchy.

El Cubrimiento de los estimadores Bayesianos es 0.95 con las 4 distribuciones iniciales cuando el tamaño de muestra es triple. En general, el Cubrimiento de los estimadores Bayesianos con las distribuciones iniciales Normal, t de Student y Cauchy, con tamaños de muestra menor es similar o menor que con la distribución inicial Doble Exponencial.

#### **Efecto del tamaño de muestra:**

Al comparar el efecto con respecto al tamaño muestral; se mencionan como tamaños de muestra menor, los tamaños de muestra de  $n = 125, 150$  y como tamaños de muestra mayor, los tamaños de muestra triples,  $n = 375, 450$  respectivamente (Cuadro 4.1).

En general, los estimadores Bayesianos son mejores cuando se triplica el tamaño de muestra.

El efecto del tamaño de muestra es leve en la Esperanza de los estimadores Bayesianos. La Esperanza está entre 3 % y 9 % mayor en tamaños de muestra menor.

El efecto del tamaño de muestra es fuerte en el Sesgo de los estimadores Bayesianos. El Sesgo está entre 165 % y 340 % mayor cuando se usan tamaños de muestra menor.

El efecto del tamaño de muestra es muy fuerte en el ECM de los estimadores Bayesianos. El ECM está entre 358 % y 714 % mayor cuando se usan tamaños de muestra menor.

El efecto del tamaño de muestra es moderado en la Longitud de los intervalos creíbles de los estimadores Bayesianos. La Longitud de los intervalos creíbles está entre 97 % y 139 % mayor en tamaños de muestra menor.

El efecto del tamaño de muestra es leve en el Cubrimiento de los intervalos creíbles de los estimadores Bayesianos. El Cubrimiento está entre 1 % y 4 % mayor cuando se usan tamaños de muestra mayor.

#### **Efecto del número de categorías:**

El efecto del número de categorías es leve en la Esperanza de los estimadores Bayesianos. La Esperanza está entre 2 % y 9 % mayor cuando  $Y$  tiene 3 categorías.

El efecto del número de categorías es fuerte en el Sesgo de los estimadores Bayesianos. El Sesgo está entre 105 % y 280 % mayor cuando  $Y$  tiene 3 categorías. Hubo un caso en que el estimador Bayesiano con la distribución Doble Exponencial tuvo Sesgo 0, esto fue en  $Y$  con  $J = 4$  y  $n = 450$ .

El efecto del número de categorías es muy fuerte en el ECM de los estimadores

Bayesianos. El ECM está entre 280 % y 619 % mayor cuando  $Y$  tiene 3 categorías.

El efecto del número de categorías es moderado en la Longitud de los intervalos creíbles de los estimadores Bayesianos. La Longitud de los intervalos creíbles está entre 100 % y 142 % mayor cuando  $Y$  tiene 3 categorías.

El efecto del número de categorías es leve en el Cubrimiento de los intervalos creíbles de los estimadores Bayesianos. El Cubrimiento está entre 1 % y 2 % mayor cuando  $Y$  tiene 4 categorías y tamaño de muestra menor. Y con tamaño de muestra mayor, el cubrimiento es 0.95 en ambos casos. Como el cubrimiento de los estimadores Bayesianos con las 4 distribuciones iniciales es 0.95 con tamaños de muestra mayor, en general, el Cubrimiento de los estimadores Bayesianos con tamaños de muestra menor es similar o menor cuando  $Y$  tiene  $J = 3$  categorías comparado con  $J = 4$  (Cuadro 4.1).

Cuadro 4.1: Criterios estimados de los estimadores de  $\beta_1$  al modelar  $Y$  con una variable independiente.

J	$n$	Distribución inicial	$\hat{\beta}_1$	$\hat{E}$	$\hat{S}$	$\widehat{ECM}$	$\hat{L}$	$\hat{C}$
3	125	Normal	$\hat{\beta}_1$	10.82	0.82	3.66	6.91	0.93
		Doble Exponencial	$\hat{\beta}_1$	10.67	0.67	3.99	7.14	0.94
		t de Student	$\hat{\beta}_1$	11.33	1.33	7.33	8.37	0.91
		Cauchy	$\hat{\beta}_1$	11.32	1.32	7.14	8.30	0.91
	375	Normal	$\hat{\beta}_1$	10.31	0.31	0.80	3.40	0.95
		Doble Exponencial	$\hat{\beta}_1$	10.24	0.24	0.76	3.40	0.95
		t de Student	$\hat{\beta}_1$	10.38	0.38	0.90	3.50	0.95
		Cauchy	$\hat{\beta}_1$	10.38	0.38	0.90	3.50	0.95
4	150	Normal	$\hat{\beta}_1$	10.40	0.40	0.96	3.42	0.94
		Doble Exponencial	$\hat{\beta}_1$	10.30	0.30	0.85	3.35	0.94
		t de Student	$\hat{\beta}_1$	10.44	0.44	1.02	3.46	0.93
		Cauchy	$\hat{\beta}_1$	10.43	0.43	1.01	3.46	0.93
	450	Normal	$\hat{\beta}_1$	10.10	0.10	0.20	1.70	0.95
		Doble Exponencial	$\hat{\beta}_1$	10.00	0.00	0.20	1.70	0.95
		t de Student	$\hat{\beta}_1$	10.10	0.10	0.20	1.70	0.95
		Cauchy	$\hat{\beta}_1$	10.10	0.10	0.20	1.70	0.95

En las Figuras 4.1 y 4.2 se presentan las 1000 estimaciones del parámetro  $\beta_1$  cuando se modela  $Y$  con 3 y 4 categorías al usar los tamaños de muestra de  $n = 125, 375$  y  $n = 150, 450$  respectivamente.

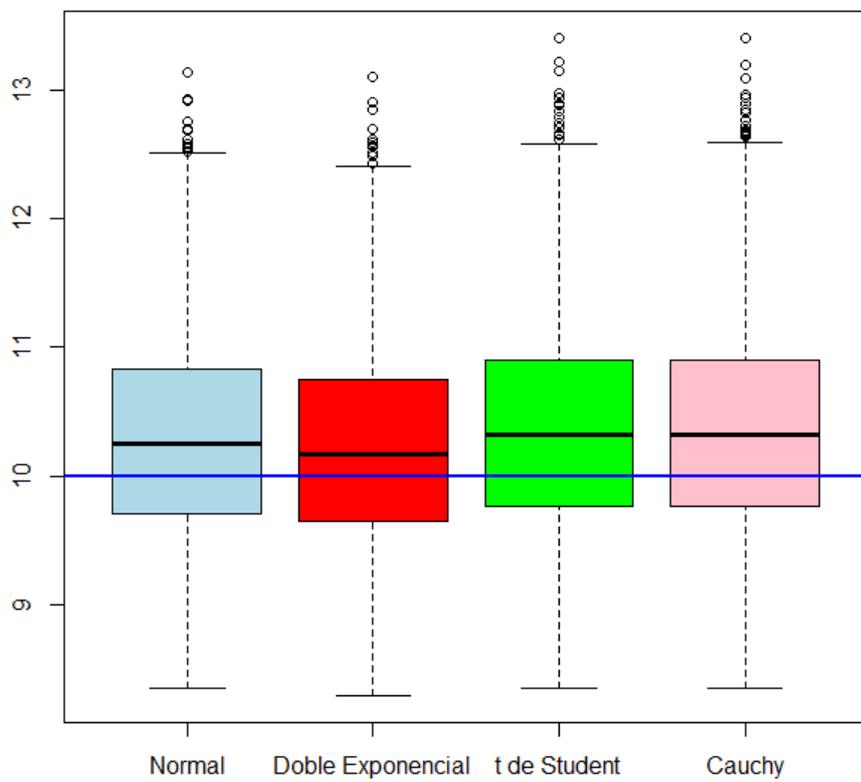
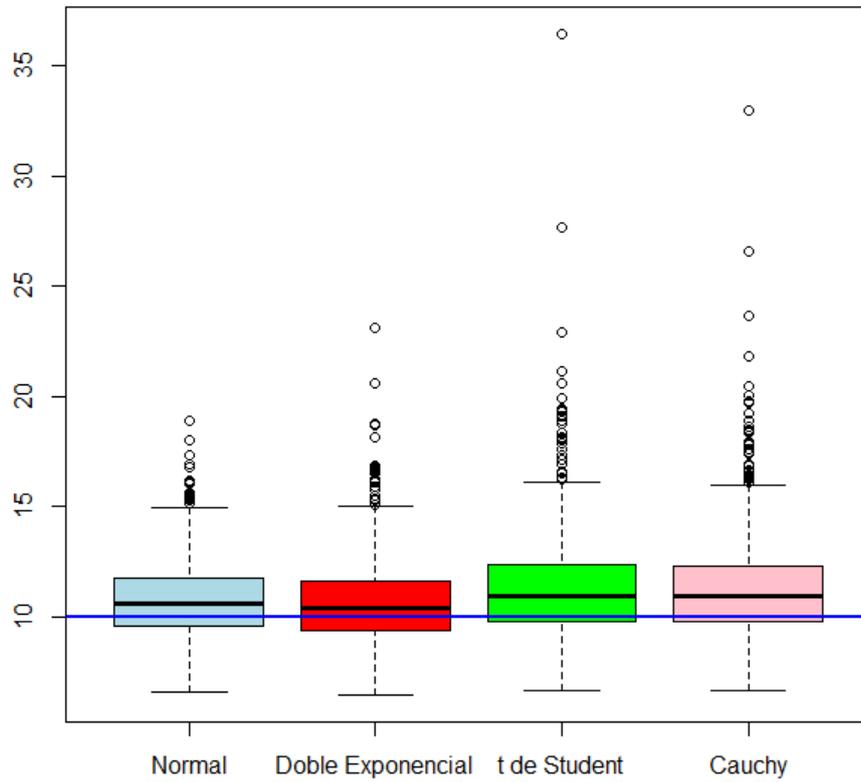


Figura 4.1: 1000 estimaciones de  $\beta_1$  cuando se modela  $Y$  con 3 categorías y se usan tamaños de muestras de  $n = 125$  (arriba) y  $n = 375$  (abajo).

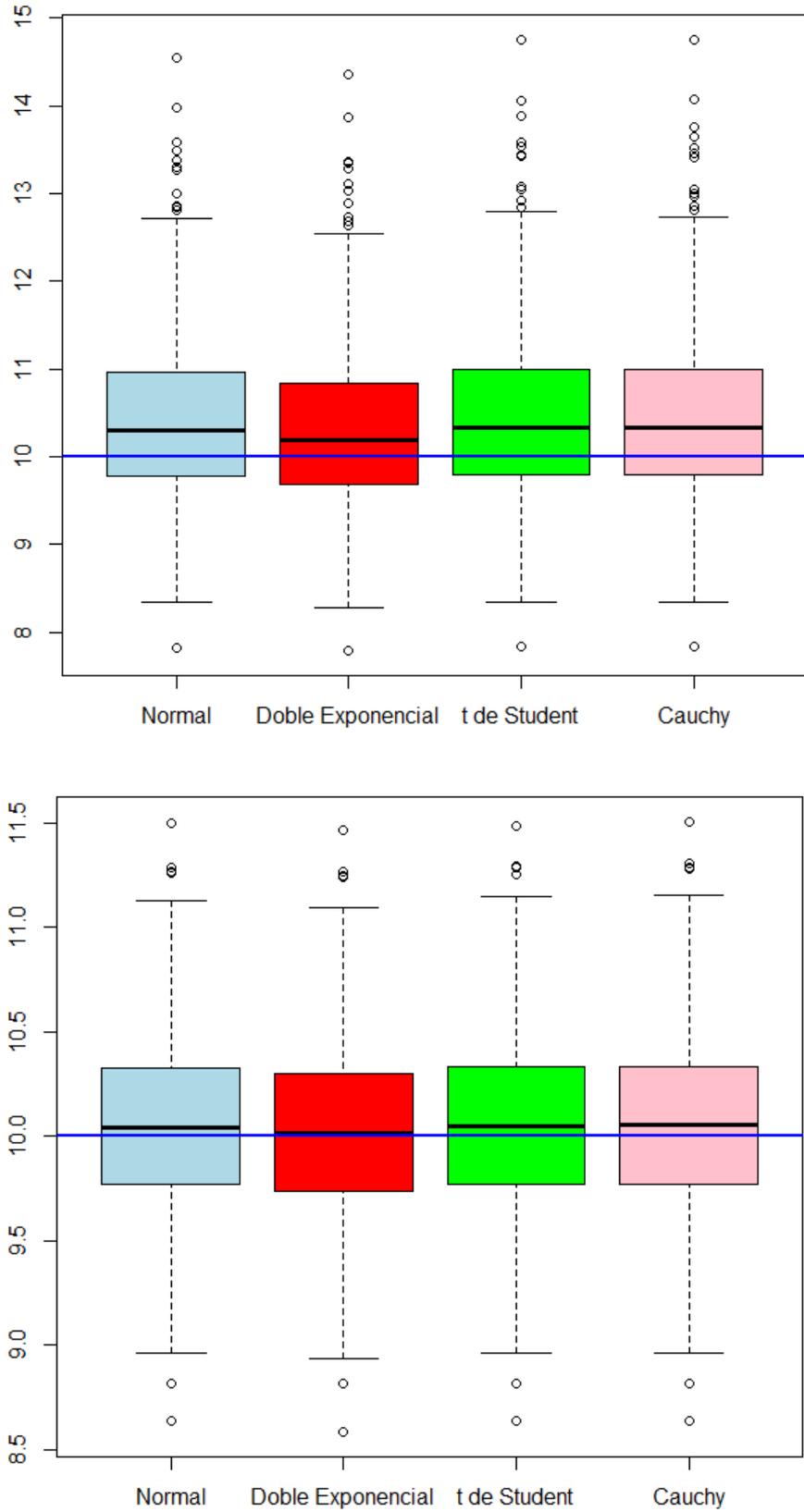


Figura 4.2: 1000 estimaciones de  $\beta_1$  cuando se modela  $Y$  con 4 categorías y se usan tamaños de muestras de  $n = 150$  (arriba) y  $n = 450$  (abajo).

En los Cuadros 4.2 y 4.3 se presentan las estimaciones de Esperanza, Sesgo, ECM, Longitud y Cubrimiento de los intervalos creíbles, cuando se modela la variable aleatorizada ordinal con  $J = 3$  y 4 categorías, con dos variables independientes respectivamente.

#### **Efecto de las distribuciones iniciales:**

En general, los estimadores Bayesianos son mejores con la distribución inicial Doble Exponencial, por lo que se comparan con los estimadores Bayesianos de las distribuciones Normal, t de Student y Cauchy para observar los efectos en Esperanza, Sesgo, ECM, Longitud y Cubrimiento de intervalos creíbles.

La Esperanza de los estimadores Bayesianos con las distribuciones iniciales Normal, t de Student y Cauchy es igual o hasta 5 % mayor que con la distribución inicial Doble Exponencial. Para  $\beta_1$  la Esperanza es igual o hasta 4 % mayor y para  $\beta_2$  es igual o hasta 5 % mayor.

En general, el Sesgo de los estimadores Bayesianos con las distribuciones iniciales Normal, t de Student y Cauchy es igual o hasta 167 % mayor que con la distribución inicial Doble Exponencial. Para  $\beta_1$  el Sesgo es igual o hasta 167 % mayor y para  $\beta_2$  es igual o hasta 70 % mayor.

En general, el ECM de los estimadores Bayesianos con las distribuciones iniciales Normal, t de Student y Cauchy es igual o hasta 85 % mayor que con la distribución inicial Doble Exponencial. Para  $\beta_1$  el ECM es entre 5 % y 85 % mayor y para  $\beta_2$  es igual o hasta 75 % mayor.

En general, la Longitud de los intervalos creíbles de los estimadores Bayesianos con las distribuciones iniciales Normal, t de Student y Cauchy es igual o hasta 18 % mayor que con la distribución inicial Doble Exponencial. Para  $\beta_1$  es igual o hasta 18 % mayor y para  $\beta_2$  es entre 1 % y 16 % mayor.

Cuando  $Y$  tiene 3 categorías, el Cubrimiento de los intervalos creíbles de los estimadores Bayesianos con las distribuciones iniciales Normal,  $t$  de Student y Cauchy es igual o hasta 3 % menor que con la distribución inicial Doble Exponencial. Cuando  $Y$  tiene 4 categorías, el Cubrimiento de los intervalos creíbles de los estimadores Bayesianos es igual en las 4 distribuciones iniciales estudiadas.

### **Efecto del tamaño de muestra:**

Al comparar el efecto con respecto al tamaño de muestra; se mencionan como tamaños de muestra menor, los tamaños de muestra de  $n = 150, 175$  y como tamaños de muestra mayor, los tamaños de muestra triple,  $n = 450, 525$  respectivamente (Cuadros 4.2 y 4.3).

El efecto del tamaño de muestra es leve en la Esperanza de los estimadores Bayesianos. En general, la Esperanza es igual o hasta 11 % mayor cuando se usan tamaños de muestra menor. Para  $\beta_1$  la Esperanza es igual o hasta 10 % mayor cuando se usan tamaños de muestra menor y para  $\beta_2$  la Esperanza es igual o hasta 11 % mayor cuando se usan tamaños de muestra menor.

El efecto del tamaño de muestra es fuerte en el Sesgo de los estimadores Bayesianos. En general, el Sesgo está entre 100 % y 580 % mayor cuando se usan tamaños de muestra menor. Para  $\beta_1$  el Sesgo está entre 125 % y 327 % mayor cuando se usan tamaños de muestra menor y para  $\beta_2$  el Sesgo está entre 100 % y 580 % mayor cuando se usan tamaños de muestra menor. Una excepción ocurrió para  $\beta_2$  en  $Y$  con  $J = 4$  y distribución inicial Doble Exponencial donde el Sesgo fue 200 % menor con tamaño de muestra menor.

El efecto del tamaño de muestra es muy fuerte en el ECM de los estimadores Bayesianos. En general, el ECM está entre 224 % y 920 % mayor cuando se usan tamaños de muestra menor. Para  $\beta_1$  el ECM está entre 224 % y 920 % mayor cuando se usan tamaños de muestra menor y para  $\beta_2$  el ECM está entre 230 % y 790 % mayor cuando se usan tamaños de muestra menor.

El efecto del tamaño de muestra es moderado en la Longitud de los intervalos creíbles de los estimadores Bayesianos. La Longitud de los intervalos creíbles está entre 83 % y 165 % mayor con tamaños de muestra menor. Para  $\beta_1$  la Longitud de los intervalos creíbles está entre 83 % y 165 % mayor cuando se usan tamaños de muestra menor y para  $\beta_2$  la Longitud de los intervalos creíbles está entre 88 % y 139 % mayor cuando se usan tamaños de muestra menor.

El efecto del tamaño de muestra es leve en el Cubrimiento de los intervalos creíbles de los estimadores Bayesianos. Para  $\beta_1$  el Cubrimiento de los intervalos creíbles es igual o hasta 4 % mayor cuando se usan tamaños de muestra mayor y para  $\beta_2$  el Cubrimiento de los intervalos creíbles es igual o hasta 3 % mayor cuando se usan tamaños de muestra mayor.

#### **Efecto del número de categorías:**

El efecto del número de categorías es leve en la Esperanza de los estimadores Bayesianos. Para  $\beta_1$  la esperanza está entre 2 % y 10 % mayor cuando Y tiene 3 categorías y para  $\beta_2$  la Esperanza está entre 1 % y 12 % mayor cuando Y tiene 3 categorías.

El efecto del número de categorías es muy fuerte en el Sesgo de los estimadores Bayesianos. Para  $\beta_1$  el Sesgo está entre 275 % y 822 % mayor cuando Y tiene 3 categorías y para  $\beta_2$  el Sesgo está entre 150 % y 4100 % mayor cuando Y tiene 3 categorías.

El efecto del número de categorías es fuerte en el ECM de los estimadores Bayesianos. Para  $\beta_1$  el ECM está entre 90 % y 571 % mayor cuando Y tiene 3 categorías y para  $\beta_2$  el ECM está entre 100 % y 409 % mayor cuando Y tiene 3 categorías.

El efecto del número de categorías es moderado en la Longitud de los intervalos creíbles de los estimadores Bayesianos. Para  $\beta_1$  la Longitud de los intervalos creíbles está entre 44 % y 105 % mayor cuando Y tiene 3 categorías y para  $\beta_2$  la Longitud de los intervalos creíbles está entre 39 % y 85 % mayor cuando Y tiene 3 categorías.

El efecto del número de categorías es leve en el Cubrimiento de los intervalos creíbles de los estimadores Bayesianos. Para  $\beta_1$  el Cubrimiento es igual o hasta 4 % mayor cuando  $Y$  tiene 4 categorías y para  $\beta_2$  el Cubrimiento es igual o hasta 3 % mayor cuando  $Y$  tiene 4 categorías. Cuando  $Y$  tiene 4 categorías, el Cubrimiento de los intervalos creíbles de los estimadores Bayesianos es igual a 0.95 con las 4 distribuciones iniciales.

Cuadro 4.2: Criterios estimados de los estimadores de  $\beta_1$  y  $\beta_2$  al modelar  $Y$  con tres categorías y dos variables independientes.

J	n	Distribución inicial	$\hat{\beta}$	$\hat{E}$	$\hat{S}$	$\widehat{ECM}$	$\hat{L}$	$\hat{C}$
3	150	Normal	$\hat{\beta}_1$	11.03	1.03	2.98	5.65	0.91
		Doble Exponencial	$\hat{\beta}_1$	10.83	0.83	2.75	5.60	0.93
		t de Student	$\hat{\beta}_1$	11.27	1.27	4.96	6.57	0.90
		Cauchy	$\hat{\beta}_1$	11.28	1.28	5.10	6.63	0.90
		Normal	$\hat{\beta}_2$	5.55	0.55	1.14	3.78	0.94
		Doble Exponencial	$\hat{\beta}_2$	5.40	0.40	1.02	3.72	0.95
		t de Student	$\hat{\beta}_2$	5.68	0.68	1.73	4.26	0.93
		Cauchy	$\hat{\beta}_2$	5.68	0.68	1.78	4.30	0.93
450		Normal	$\hat{\beta}_1$	10.30	0.30	0.50	2.50	0.94
		Doble Exponencial	$\hat{\beta}_1$	10.20	0.20	0.40	2.50	0.94
		t de Student	$\hat{\beta}_1$	10.30	0.30	0.50	2.50	0.94
		Cauchy	$\hat{\beta}_1$	10.30	0.30	0.50	2.50	0.93
		Normal	$\hat{\beta}_2$	5.10	0.10	0.20	1.80	0.95
		Doble Exponencial	$\hat{\beta}_2$	5.10	0.10	0.20	1.70	0.96
		t de Student	$\hat{\beta}_2$	5.10	0.10	0.20	1.80	0.95
		Cauchy	$\hat{\beta}_2$	5.10	0.10	0.20	1.80	0.96

Cuadro 4.3: Criterios estimados de los estimadores de  $\beta_1$  y  $\beta_2$  al modelar  $Y$  con cuatro categorías y dos variables independientes.

J	$n$	Distribución inicial	$\hat{\beta}$	$\hat{E}$	$\hat{S}$	$\widehat{ECM}$	$\hat{L}$	$\hat{C}$	
4	175	Normal	$\hat{\beta}_1$	10.24	0.24	0.76	3.23	0.94	
		Doble Exponencial	$\hat{\beta}_1$	10.09	0.09	0.68	3.16	0.94	
		t de Student	$\hat{\beta}_1$	10.22	0.22	0.76	3.23	0.94	
		Cauchy	$\hat{\beta}_1$	10.22	0.22	0.76	3.23	0.94	
	525	175	Normal	$\hat{\beta}_2$	5.10	0.10	0.35	2.31	0.96
			Doble Exponencial	$\hat{\beta}_2$	4.99	-0.01	0.33	2.29	0.96
			t de Student	$\hat{\beta}_2$	5.09	0.09	0.35	2.32	0.96
			Cauchy	$\hat{\beta}_2$	5.08	0.08	0.35	2.32	0.96
4	525	Normal	$\hat{\beta}_1$	10.08	0.08	0.22	1.74	0.94	
		Doble Exponencial	$\hat{\beta}_1$	10.04	0.04	0.21	1.73	0.94	
		t de Student	$\hat{\beta}_1$	10.08	0.08	0.22	1.74	0.94	
		Cauchy	$\hat{\beta}_1$	10.08	0.08	0.22	1.74	0.94	
	525	525	Normal	$\hat{\beta}_2$	5.04	0.04	0.10	1.23	0.95
			Doble Exponencial	$\hat{\beta}_2$	5.01	0.01	0.10	1.22	0.95
			t de Student	$\hat{\beta}_2$	5.04	0.04	0.10	1.23	0.95
			Cauchy	$\hat{\beta}_2$	5.04	0.04	0.10	1.23	0.95

En las Figuras 4.3 a 4.6 se presentan las 1000 estimaciones de los parámetros  $\beta_1$  y  $\beta_2$ , cuando se modela  $Y$  con 3 y 4 categorías ordinales al usar diferentes tamaños de muestra; ligeramente se observa que con la distribución inicial Doble Exponencial, la mediana en las estimaciones se encuentra más cercana al verdadero valor del parámetro, que con las otras distribuciones iniciales.

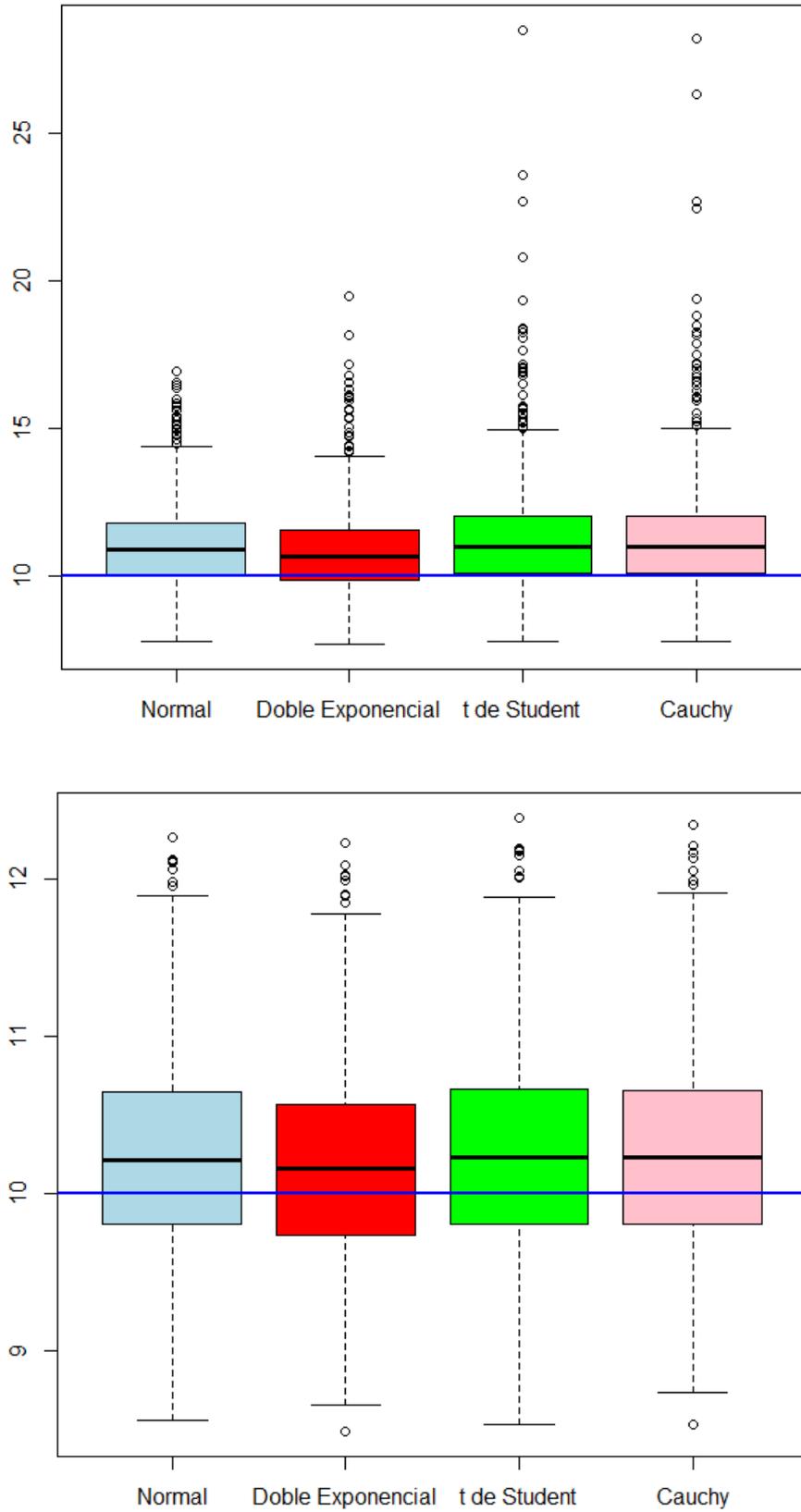


Figura 4.3: 1000 estimaciones de  $\beta_1$  cuando  $Y$  tiene 3 categorías y se usan tamaños de muestras de  $n = 150$  (arriba) y  $n = 450$  (abajo).

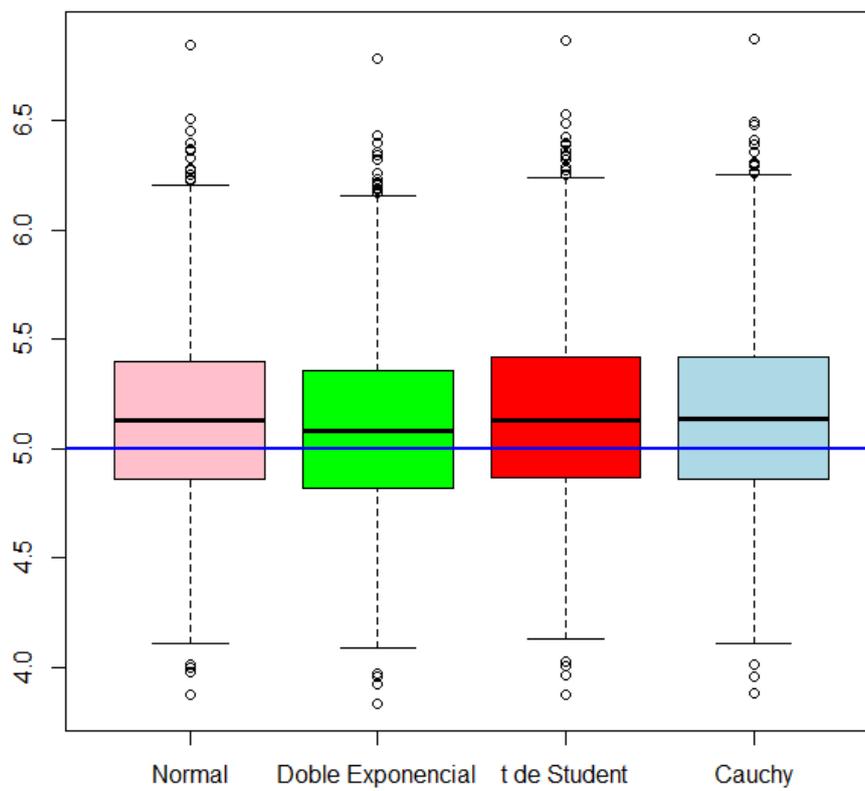
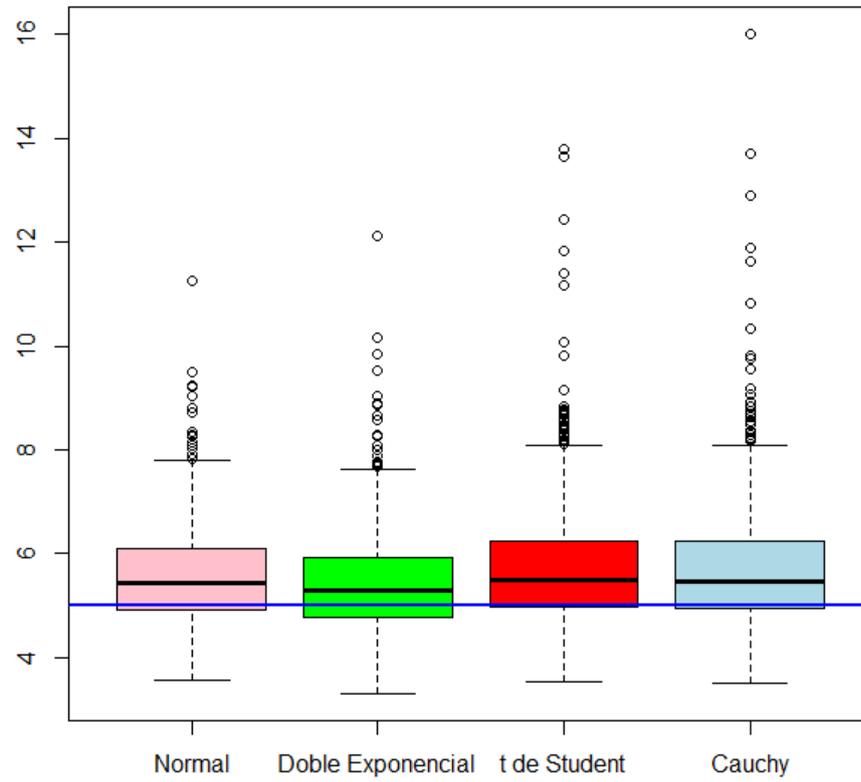


Figura 4.4: 1000 estimaciones de  $\beta_2$  cuando  $Y$  tiene 3 categorías y se usan tamaños de muestras de  $n = 150$  (arriba) y  $n = 450$  (abajo).

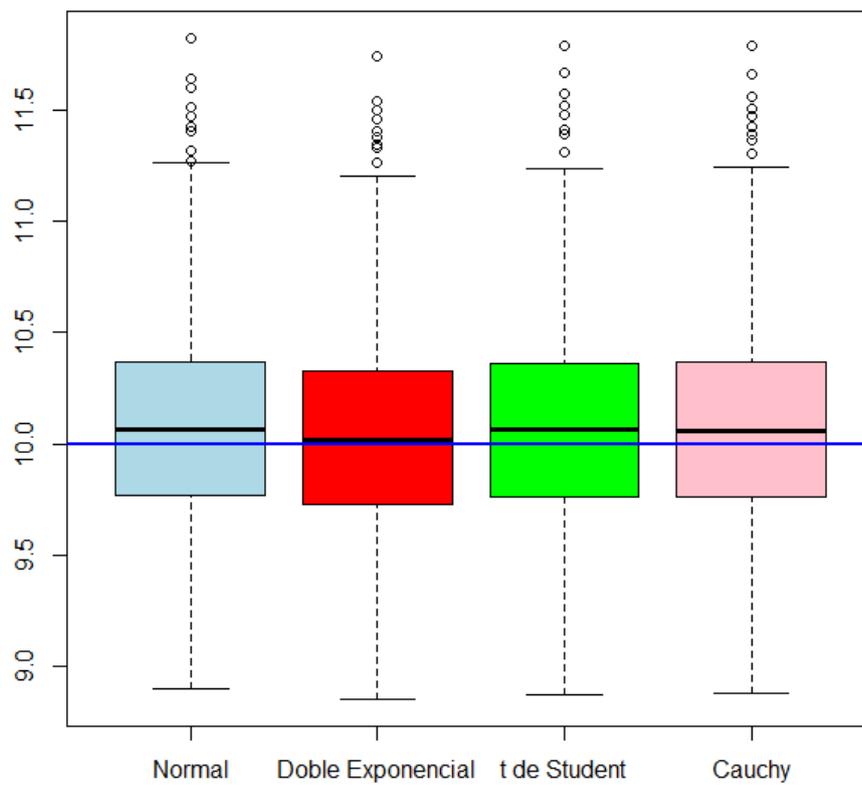
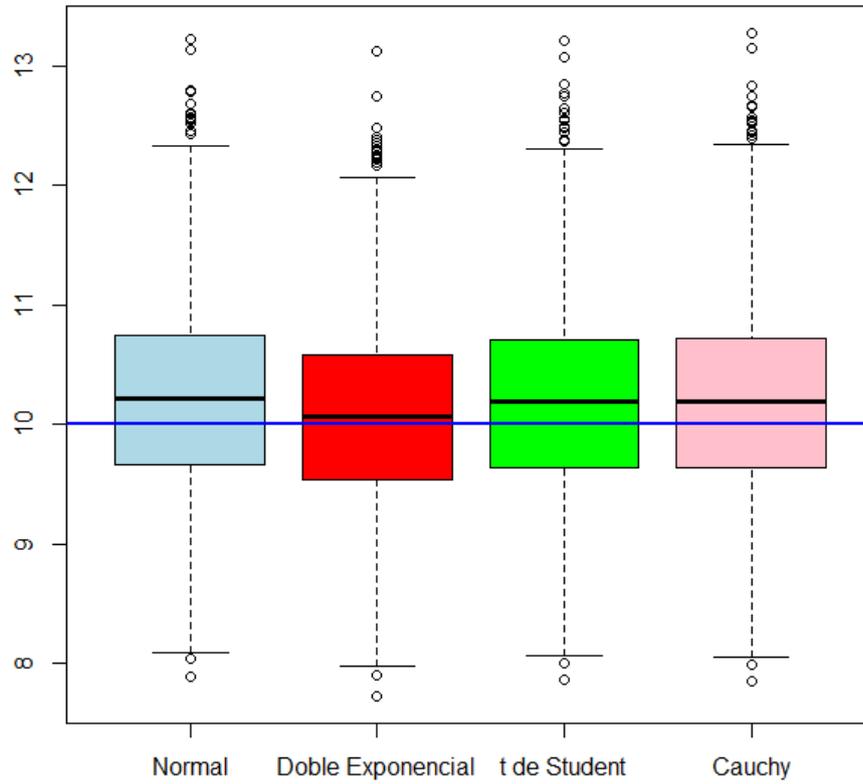


Figura 4.5: 1000 estimaciones de  $\beta_1$  cuando  $Y$  tiene 4 categorías y se usan tamaños de muestras de  $n = 175$  (arriba) y  $n = 525$  (abajo).

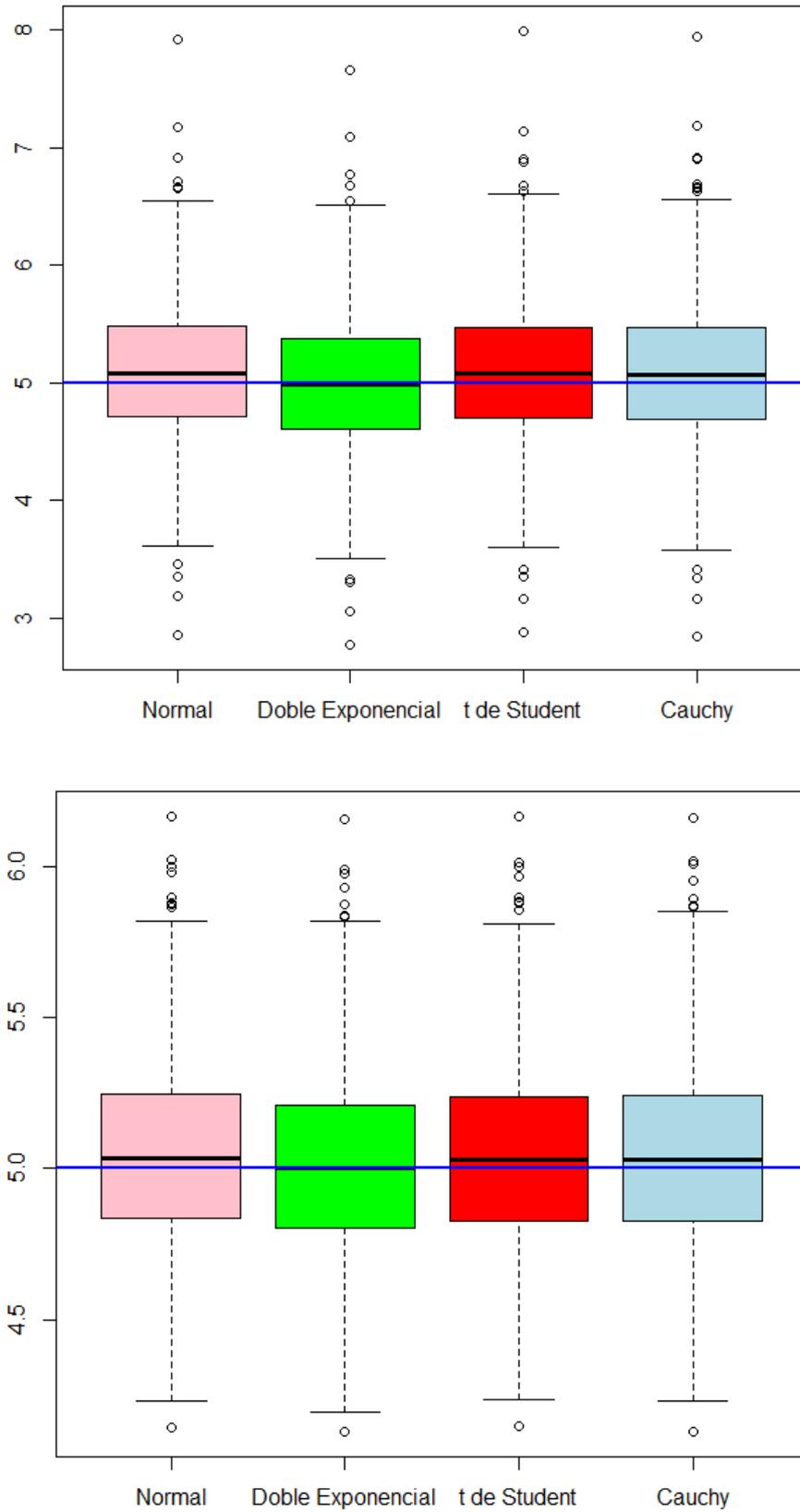


Figura 4.6: 1000 estimaciones de  $\beta_2$  cuando  $Y$  tiene 4 categorías y se usan tamaños de muestras de  $n = 175$  (arriba) y  $n = 525$  (abajo).

---

### Conclusiones

---

En la literatura no se encontraron trabajos previos sobre la modelación Bayesiana del efecto de variables independientes en una variable respuesta aleatorizada ordinal.

Los trabajos de Cruyff *et al.* (2006) y van der Heijden *et al.* (2011) son los más parecidos al tema desarrollado en este trabajo. Cruyff *et al.* (2006) modelaron una variable aleatorizada con categorías de respuesta ordenadas, pero con un modelo Poisson cero inflado y bajo el enfoque clásico. Por otro lado, van der Heijden *et al.* (2011) modelaron la suma de  $m$  preguntas sobre características sensitivas binarias con respuestas *Si* usando el modelo de regresión ordinal.

Los resultados obtenidos en este trabajo se basa en un diseño de respuesta aleatoria específico, que es el diseño de respuesta forzada. Si se usara otro mecanismo aleatorio se esperaría que los resultados fueran similares en cuanto al mejor estimador Bayesiano obtenido.

En la mayoría de los escenarios simulados, los mejores estimadores Bayesianos fueron con la distribución inicial Doble Exponencial, porque las estimaciones del Sesgo, Error Cuadrático Medio y Longitud de los intervalos creíbles fueron menores y el Cubrimiento fue mayor. A excepción de un sólo caso, donde se modeló  $Y$  con tres categorías,

con una variable independiente y  $n = 125$ , que ligeramente fueron menores las estimaciones del Error Cuadrático Medio y la Longitud de los intervalos creíbles, con la distribución inicial Normal.

En general, al triplicar el tamaño de muestra, la Esperanza estimada fue más próxima al verdadero valor del parámetro y coincidió en un caso, cuando se modeló  $Y$  con 4 categorías, con una variable independiente y con la distribución inicial Doble Exponencial. Además fueron menores las estimaciones del Sesgo, Error Cuadrático Medio y Longitud de los intervalos creíbles y el Cubrimiento estimado en la mayoría de las veces aumentó hasta un 4%.

### Referencias

---

Abernathy, J. R., B. G. Greenberg, and D. G. Horvitz. 1970. Estimates of induced abortion in urban North Carolina. *Demography* 7: 19–29. 3

Abul-Ela A., B. G. Greenberg, and D. G. Horvitz. 1967. A multiproportions randomized response model. *Journal of the American Statistical Association*. 62:990–1008. 4

Alavi, S. M. R., and M. Tajodini. 2016. Maximum likelihood estimation of sensitive proportion using repeated randomized response techniques. *Journal of Applied Statistics*. 43:3, 563-571, DOI: 10.1080/02664763.2015.1070811. 4

Arnab R., and D. K. Shangodoyin. 2015. Randomized Response Techniques Using Maximum Likelihood Estimator. *Communications in Statistics - Theory and Methods*. 44: 16, 3340-3349,10.1080/03610926.2013.841935. 4

Arnab R., S. Singh, and D. North. 2012. Use of two deck of cards in randomized response techniques for complex survey designs. *Commun. Statisti. Theor. Meth.* 41:3198–3210. 4

Blair G., K. Imai, and Y.-Y. Zhou. 2015. Design and Analysis of the Randomized Response Technique. *Journal of the American Statistical Association*. 110: 511, 1304-1319, DOI:10.1080/01621459.2015.1050028. I, II, 6, 8, 12

- Boruch, R. F. 1971. Assuring confidentiality of responses in social research: a note on strategies. *The American Sociologist*. 6, 308-311. 3
- Casella G., and L. R. Berger. 2002. *Statistical Inference*. Second Edition. Wadsworth Group. pág. 330.
- Chang, H.-J., C.-L. Wang, and K.-C. Huang. 2004. Using Randomized Response to Estimate the Proportion and Truthful Reporting Probability in a Dichotomous Finite Population, *Journal of Applied Statistics*. 31:5, 565-573, DOI: 10.1080/02664760410001681819. 4
- Chaudhuri A. 2001. Using randomized response from a complex survey to estimate a sensitive proportion in a dichotomous finite population. *Journal of Statistical Planning and Inference*. 94, 37-42. 4
- Chaudhuri A., and R. Mukherjee. 1987. Randomized response techniques: a review. *Statistica Neerlandica*. 41, pp. 27-44. 6
- Chi, I.-C., L. P. Chow, and R. V. Rider. 1972. The Randomized Response Technique as Used in the Taiwan Outcome of Pregnancy Study. *Studies in Family Planning*. Vol. 3, 265-269. 3
- Chua, T. C., and A. K. Tsui. 2000. Procuring honest responses indirectly. *Journal of Statistical Planning and Inference*. 90, pp. 107-116. 3, 4
- Congdon, Peter. 2006. *Bayesian Statistical Modelling*. Second Edition. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England. 19
- Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>. 23, 47
- Cruyff, M.J.L.F., A. van den Hout, P.G.M. van der Heijden, and U. Böckenholt. 2006. A Log-Linear Randomized-Response Model to Account for Cheating. In: *Proceedings of the 21st international workshop on statistical modelling*. 6, 64

- De Schrijver A. 2012. Sample survey on sensitive topics: Investigating respondents understanding and trust in alternative versions of the randomized response technique. *Journal of Research Practice*. 8(1). 7
- Fox, J.-P. 2010. *Bayesian Item Response Modeling: Theory and Applications*. Statistics for Social and Behavioral Sciences. Springer Science + Business Media, LLC. pp. 58-59. 31
- Gelman A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian data analysis*. Taylor and Francis. Third Edition. pp: 6-7, 275-286. 19, 25
- Gelman A. and D. B. Rubin. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*. Vol. 7: 457-511. 23, 25, 47
- Gómez G., y P. Delicado. 2006. *Curso de Inferencia y Decisión*. <http://www-eio.upc.edu/delicado/docencia/IyDapuntes.pdf>. Pág. 80. 29
- Greenberg, B.G., E. L. A. Abul-Ela, W.R. Simmons, and D.G. Horvitz. 1969. The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association*. 64, 520-539. 3, 5, 6, 10
- Kalucha G., S. Gupta, and B. K. Dass. 2015 Ratio Estimation of Finite Population Mean Using Optional Randomized Response Models. *Journal of Statistical Theory and Practice*. 9:3, 633-645, DOI: 10.1080/15598608.2014.977982. 5
- Kim, J.-M., and T.-Y. Heo. 2011. A Bayesian Analysis of the Multinomial Randomized Response Model using Dirichlet Prior Distribution. In *Proceedings for the Spring Conference*. pp. 239-244. 7
- Kim, J.-M., J. Tebbs, and S.-W. An. 2005. Extensions of Mangat's randomized-response model. *Journal of Statistical Planning and Inference*, In press. 5, 7
- Kim, J.-M., and W. D. Warde. 2005. Some new results on the multinomial randomized response model. *Communications in Statistics: Theory and Methods*. 34(4), 847-856. 7, 34

- Kruschke, J. K. 2014. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Library of Congress Cataloging in Publication Data. Second Edition. pp: 671-702. 39, 43, 44, 46
- Kuk, A.Y.C. 1990. Asking sensitive question indirectly. *Biometrika* 77, 436-438. 3, 13
- Lensvelt-Mulders G.J.L.M., J.J. Hox. P.G.M. van der Heijden, C.J.M. Maas. 2005. Meta-analysis of randomized response research: thirty-five years of validation. *Soc. Methods Res.* 33, 319–348. 7
- Liu, P. T., and L. P. Chow. 1976. A new discrete quantitative randomized response model. *Journal of the American Statistical Association.* 71:353, 72–73. 5, 34
- Mangat, N.S. 1994. An Improved Randomized Response Strategy. *Journal of the Royal Statistical Society. Series B*, 56, 93–95. III, 3, 4, 7, 8, 17
- Mangat, N.S., and R. Singh. 1990. An alternative randomized response procedure. *Biometrika* 77, 439-442. III, 3, 4, 8, 15, 17
- Mood, A. M., F. A. Graybill, and D. C. Boes. 1964. Introduction to the theory of statistics. McGraw-Hill, Inc. pág. 343-346. 29
- Moors, J. J. A. 1971. Optimization of the Unrelated Question Randomized Response Model. *Journal of the American Statistical Association.* 66:335, 627-629. 3, 6
- Mukerjee R. 1981. Inference on confidential characters from survey data. *Calcutta Statistical Association Bull.* 30, 77-88. 4
- Nájera, E. R. 2015. ¿Qué es la Estadística Bayesiana?. *Journal of Basic Sciences.* 30
- Odumade O., and S. Singh. 2009. Efficient Use of Two Decks of Cards in Randomized Response Sampling. *Communications in Statistics - Theory and Methods.* 38:4, 439-446, DOI: 10.1080/03610920802225065. 3, 4

- Plummer M. 2013. JAGS Version 3.4.0 User Manual. 26, 47
- Rizzo, M. L. 2008. Statistical Computing with R. Taylor & Francis Group, LLC. pág. 267.25
- Robert, C. P., and G. Casella. 2010. Introducing Monte Carlo Methods with R. Springer Science + Business Media, LLC. pág. 169, 206. 21, 22
- Scheers N., and C.M. Dayton. 1988. Covariate randomized response models. Journal of the American Statistical Association. 83, 969–974. 5
- Singh, S., and A. H. Joarder. 1997. Unknown repeated trials in randomized response sampling. JISAS 50, 70-74. 3
- Singh, H. P., and T. A. Tarray. 2016. Role of weights in improving the efficiency of Kim and Warde's mixed randomized response model. Communications in Statistics - Theory and Methods. 45:4, 1014-1030, DOI: 10.1080/03610926.2013.853795. 4
- Soberanis, V. H., y V. Miranda. 2011. The Generalized Logistic Regression Estimator in a Finite Population Sampling without Replacement Setting with Randomized Response. Revista Colombiana de Estadística. 34, 3, 451–460. 5
- Tracy, P. E., and J. A. Fox. 1981. The validity of randomized response for sensitive measurements. American Sociological Association. 46(2), 187-200. 6
- van den Hout A., P. G. van der Heijden, and R. Gilchrist. 2007. The logistic regression model with response variables subject to randomized response. Computational Statistics and Data Analysis. 51, 6060-6069. 5, 12
- van der Heijden, P. G., and G. van Gils. 1996. Some Logistic Regression Models for Randomized Response Data. In Proceedings of the 11th International Workshop on Statistical Modeling. pp. 15–19. 5
- van der Heijden, P. G., L. Frank, M. Cruyff, and U. Böckenholt. 2011. Some regression procedures for randomized response data. Int. Statistical Inst.: Proc. 58th World Statistical Congress. pp: 1958-1966.6, 64

Warner, S. L. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association* 60: 63-69.  
I, II, 2, 3, 4, 5, 8

## Anexo 1. Distribuciones posteriores conjuntas

Modelos de la variable ordinal  $Y$  con 3 y 4 categorías con las variables independientes estandarizadas.

Por teoría bayesiana la densidad posterior conjunta de  $\boldsymbol{\alpha}$  y  $\sigma$ , al modelar  $Y$  con 3 categorías y la densidad posterior conjunta de  $\boldsymbol{\alpha}$ ,  $\sigma$  y  $\gamma_2$  al modelar  $Y$  con 4 categorías están dadas respectivamente por:

$$\pi(\boldsymbol{\alpha}, \sigma \mid \mathbf{X}, \mathbf{Y}) = C \pi(\boldsymbol{\alpha}, \sigma) \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) \pi_{ij} \right\}$$

y

$$\pi(\boldsymbol{\alpha}, \gamma_2, \sigma \mid \mathbf{X}, \mathbf{Y}) = C \pi(\boldsymbol{\alpha}, \gamma_2, \sigma) \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) \pi_{ij} \right\}$$

Pero al introducir la variable latente con distribución Normal y las distribuciones iniciales marginales entonces están dadas respectivamente por:

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \sigma, \mathbf{Z} \mid \mathbf{Y}, \mathbf{Y}) &= C \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \\ &\quad \times \pi(\sigma) \times \phi(Y_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2) \end{aligned}$$

y

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \gamma_2, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &= C \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \times \pi(\gamma_2) \\ &\quad \times \pi(\sigma) \times \phi(Z_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2) \end{aligned}$$

**Las distribuciones posteriores son:**

1.  $\alpha_k$  con distribución inicial Normal.

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \sigma, \mathbf{Z} \mid \mathbf{Y}, \mathbf{Y}) &= C \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \\ &\quad \times \pi(\sigma) \times \phi(Y_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2) \end{aligned}$$

$$\begin{aligned}
&= C \prod_{i=1}^n \sum_{j=1}^3 I(Y_i = j) \left[ p \left[ \Phi \left( \frac{\gamma_j - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) - \Phi \left( \frac{\gamma_{j-1} - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) \right] + (1-p)q \right] \\
&\quad \times \left[ \left( \frac{1}{18\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 2)^2}{18} \right\} \right] \left[ \left( \frac{1}{18\pi} \right)^{\frac{p}{2}} \exp \left\{ \frac{-\sum_{k=1}^p \alpha_k^2}{18} \right\} \right] \\
&\quad \times \left( \frac{1}{30 - 0.003} \right) I_{(0.003, 30)}(\sigma^2) \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \\
&= C \prod_{i=1}^n \left[ \sum_{j=1}^3 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right] \\
&\quad \times \left[ \left( \frac{1}{18\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0^2 - 4\alpha_0 + 4)}{18} \right\} \right] \left[ \left( \frac{1}{18\pi} \right)^{\frac{p}{2}} \exp \left\{ \frac{-\sum_{k=1}^p \alpha_k^2}{18} \right\} \right] \\
&\quad \times \left( \frac{1}{30 - 0.003} \right) I_{(0.003, 30)}(\sigma^2) \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right]
\end{aligned}$$

$$\begin{aligned}
\pi(\boldsymbol{\alpha}, \gamma_2, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &= C \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \times \pi(\gamma_2) \\
&\quad \times \pi(\sigma) \times \phi(Z_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2)
\end{aligned}$$

$$\begin{aligned}
&= C \prod_{i=1}^n \sum_{j=1}^4 I(Y_i = j) \left[ p \left[ \Phi \left( \frac{\gamma_j - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) - \Phi \left( \frac{\gamma_{j-1} - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) \right] + (1-p)q \right] \\
&\quad \times \left[ \left( \frac{1}{32\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 2.5)^2}{32} \right\} \right] \left[ \left( \frac{1}{32\pi} \right)^{\frac{p}{2}} \exp \left\{ \frac{-\sum_{k=1}^p \alpha_k^2}{32} \right\} \right] \\
&\quad \times \left[ \left( \frac{1}{8\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\gamma_2 - 2.5)^2}{8} \right\} \right] \left( \frac{1}{40 - 0.004} \right) I_{(0.004, 40)}(\sigma^2) \\
&\quad \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \\
&= C \prod_{i=1}^n \left[ \sum_{j=1}^4 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right] \\
&\quad \times \left[ \left( \frac{1}{32\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0^2 - 5\alpha_0 + 6.25)}{32} \right\} \right] \left[ \left( \frac{1}{32\pi} \right)^{\frac{p}{2}} \exp \left\{ \frac{-\sum_{k=1}^p \alpha_k^2}{32} \right\} \right]
\end{aligned}$$

$$\begin{aligned} & \times \left[ \left( \frac{1}{8\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\gamma_2^2 - 5\gamma_2 + 6.25)}{8} \right\} \right] \left( \frac{1}{40 - 0.004} \right) I_{(0.004, 40)}(\sigma^2) \\ & \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \end{aligned}$$

## 2. $\alpha_k$ con distribución inicial Doble Exponencial.

$$\begin{aligned} \pi(\boldsymbol{\alpha}, \sigma, \mathbf{Z} \mid \mathbf{Y}, \mathbf{Y}) &= C \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \\ & \quad \times \pi(\sigma) \times \phi(Y_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2) \\ &= C \prod_{i=1}^n \sum_{j=1}^3 I(Y_i = j) \left[ p \left[ \Phi \left( \frac{\gamma_j - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) - \Phi \left( \frac{\gamma_{j-1} - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) \right] + (1-p)q \right] \\ & \quad \times \left[ \left( \frac{1}{18\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 2)^2}{18} \right\} \right] \times \left( \frac{1}{2\lambda} \right)^p \exp \left( -\frac{\sum_{k=1}^p |\alpha_k|}{\lambda} \right) \\ & \quad \times \left( \frac{1}{10 - 0.001} \right) I_{(0.001, 10)}(\lambda) \left( \frac{1}{30 - 0.003} \right) I_{(0.003, 30)}(\sigma^2) \\ & \quad \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \\ &= C \prod_{i=1}^n \left[ \sum_{j=1}^3 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right] \\ & \quad \times \left[ \left( \frac{1}{18\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 4\alpha_0 + 4)}{18} \right\} \right] \times \left( \frac{1}{2\lambda} \right)^p \exp \left( -\frac{\sum_{k=1}^p |\alpha_k|}{\lambda} \right) \\ & \quad \times \left( \frac{1}{10 - 0.001} \right) I_{(0.001, 10)}(\lambda) \left( \frac{1}{30 - 0.003} \right) I_{(0.003, 30)}(\sigma^2) \\ & \quad \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \\ \pi(\boldsymbol{\alpha}, \gamma_2, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &= C \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \times \pi(\gamma_2) \\ & \quad \times \pi(\sigma) \times \phi(Z_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2) \end{aligned}$$

$$\begin{aligned}
&= C \prod_{i=1}^n \sum_{j=1}^4 I(Y_i = j) \left[ p \left[ \Phi \left( \frac{\gamma_j - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) - \Phi \left( \frac{\gamma_{j-1} - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) \right] + (1-p)q \right] \\
&\quad \times \left[ \left( \frac{1}{32\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 2.5)^2}{32} \right\} \right] \times \left( \frac{1}{2\lambda} \right)^p \exp \left( -\frac{\sum_{k=1}^p |\alpha_k|}{\lambda} \right) \\
&\quad \times \left( \frac{1}{10 - 0.001} \right) I_{(0.001,10)}(\lambda) \left( \frac{1}{40 - 0.004} \right) I_{(0.004,40)}(\sigma^2) \\
&\quad \times \left[ \left( \frac{1}{8\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\gamma_2 - 2.5)^2}{8} \right\} \right] \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \\
&= C \prod_{i=1}^n \left[ \sum_{j=1}^4 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right] \\
&\quad \times \left[ \left( \frac{1}{32\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 5\alpha_0 + 6.25)}{32} \right\} \right] \times \left( \frac{1}{2\lambda} \right)^p \exp \left( -\frac{\sum_{k=1}^p |\alpha_k|}{\lambda} \right) \\
&\quad \times \left( \frac{1}{10 - 0.001} \right) I_{(0.001,10)}(\lambda) \left( \frac{1}{40 - 0.004} \right) I_{(0.004,40)}(\sigma^2) \\
&\quad \times \left[ \left( \frac{1}{8\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\gamma_2^2 - 5\gamma_2 + 6.25)}{8} \right\} \right] \\
&\quad \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right]
\end{aligned}$$

### 3. $\alpha_k$ con distribución inicial t de Student

$$\begin{aligned}
\pi(\boldsymbol{\alpha}, \sigma, \mathbf{Z} \mid \mathbf{Y}, \mathbf{Y}) &= C \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \\
&\quad \times \pi(\sigma) \times \phi(Y_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2) \\
&= C \prod_{i=1}^n \sum_{j=1}^3 I(Y_i = j) \left[ p \left[ \Phi \left( \frac{\gamma_j - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) - \Phi \left( \frac{\gamma_{j-1} - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) \right] + (1-p)q \right] \\
&\quad \times \left[ \left( \frac{1}{18\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 2)^2}{18} \right\} \right] \left( \frac{\Gamma(\frac{(n^*+1)+1}{2})}{\Gamma(\frac{n^*+1}{2}) \sqrt{\pi(n^*+1)\sigma_B^2}} \right)^p \\
&\quad \times \prod_{k=1}^p \left( 1 + \frac{1}{(n^*+1)\sigma_B^2} \alpha_k^2 \right)^{-\frac{(n^*+1)+1}{2}} \left[ \left( \frac{1}{29} \right) \exp \left\{ -\left( \frac{1}{29} \right) n^* \right\} \right]
\end{aligned}$$

$$\begin{aligned}
& \times \left( \frac{1}{1000 - 0.001} \right) I_{(0.001, 1000)}(\sigma_B^2) \left( \frac{1}{30 - 0.003} \right) I_{(0.003, 30)}(\sigma^2) \\
& \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \\
= & C \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\
& \times \left[ \left( \frac{1}{18\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0^2 - 4\alpha_0 + 4)}{18} \right\} \right] \left( \frac{\Gamma(\frac{(n^*+1)+1}{2})}{\Gamma(\frac{n^*+1}{2}) \sqrt{\pi(n^*+1)\sigma_B^2}} \right)^p \\
& \times \prod_{k=1}^p \left( 1 + \frac{1}{(n^*+1)\sigma_B^2} \alpha_k^2 \right)^{-\frac{(n^*+2)}{2}} \left[ \left( \frac{1}{29} \right) \exp \left\{ - \left( \frac{1}{29} \right) n^* \right\} \right] \\
& \times \left( \frac{1}{1000 - 0.001} \right) I_{(0.001, 1000)}(\sigma_B^2) \left( \frac{1}{30 - 0.003} \right) I_{(0.003, 30)}(\sigma^2) \\
& \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \\
\pi(\boldsymbol{\alpha}, \gamma_2, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) = & C \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \times \pi(\gamma_2) \\
& \times \pi(\sigma) \times \phi(Z_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2) \\
= & C \prod_{i=1}^n \sum_{j=1}^4 I(Y_i = j) \left[ p \left[ \Phi \left( \frac{\gamma_j - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) - \Phi \left( \frac{\gamma_{j-1} - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) \right] + (1-p)q \right] \\
& \times \left[ \left( \frac{1}{32\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 2.5)^2}{32} \right\} \right] \left( \frac{\Gamma(\frac{(n^*+1)+1}{2})}{\Gamma(\frac{n^*+1}{2}) \sqrt{\pi(n^*+1)\sigma_B^2}} \right)^p \\
& \times \prod_{k=1}^p \left( 1 + \frac{1}{(n^*+1)\sigma_B^2} \alpha_k^2 \right)^{-\frac{(n^*+1)+1}{2}} \left[ \left( \frac{1}{29} \right) \exp \left\{ - \left( \frac{1}{29} \right) n^* \right\} \right] \\
& \times \left( \frac{1}{1000 - 0.001} \right) I_{(0.001, 1000)}(\sigma_B^2) \left( \frac{1}{40 - 0.004} \right) I_{(0.004, 40)}(\sigma^2) \\
& \times \left[ \left( \frac{1}{8\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\gamma_2 - 2.5)^2}{8} \right\} \right] \\
& \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right]
\end{aligned}$$

$$\begin{aligned}
&= C \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right\} \\
&\quad \times \left[ \left( \frac{1}{32\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0^2 - 5\alpha_0 + 6.25)}{32} \right\} \right] \left( \frac{\Gamma(\frac{(n^*+1)+1}{2})}{\Gamma(\frac{n^*+1}{2}) \sqrt{\pi(n^*+1)\sigma_B^2}} \right)^p \\
&\quad \times \prod_{k=1}^p \left( 1 + \frac{1}{(n^*+1)\sigma_B^2} \alpha_k^2 \right)^{-\frac{(n^*+2)}{2}} \left[ \left( \frac{1}{29} \right) \exp \left\{ - \left( \frac{1}{29} \right) n^* \right\} \right] \\
&\quad \times \left( \frac{1}{1000 - 0.001} \right) I_{(0.001, 1000)}(\sigma_B^2) \left( \frac{1}{40 - 0.004} \right) I_{(0.004, 40)}(\sigma^2) \\
&\quad \times \left[ \left( \frac{1}{8\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\gamma_2^2 - 5\gamma_2 + 6.25)}{8} \right\} \right] \\
&\quad \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right]
\end{aligned}$$

#### 4. $\alpha_k$ con distribución inicial Cauchy

$$\begin{aligned}
\pi(\boldsymbol{\alpha}, \sigma, \mathbf{Z} \mid \mathbf{Y}, \mathbf{Y}) &= C \prod_{i=1}^n \left\{ \sum_{j=1}^3 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \\
&\quad \times \pi(\sigma) \times \phi(Y_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2) \\
&= C \prod_{i=1}^n \sum_{j=1}^3 I(Y_i = j) \left[ p \left[ \Phi \left( \frac{\gamma_j - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) - \Phi \left( \frac{\gamma_{j-1} - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) \right] + (1-p)q \right] \\
&\quad \times \left[ \left( \frac{1}{18\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 2)^2}{18} \right\} \right] \left( \frac{\Gamma(\frac{1+1}{2})}{\Gamma(\frac{1}{2}) \sqrt{\pi(1)\sigma_B^2}} \right)^p \\
&\quad \times \prod_{k=1}^p \left( 1 + \frac{1}{(1)\sigma_B^2} \alpha_k^2 \right)^{-\frac{1+1}{2}} \left( \frac{1}{1000 - 0.001} \right) I_{(0.001, 1000)}(\sigma_B^2) \\
&\quad \times \left( \frac{1}{30 - 0.003} \right) I_{(0.003, 30)}(\sigma^2) \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \\
&= C \prod_{i=1}^n \left[ \sum_{j=1}^3 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right]
\end{aligned}$$

$$\begin{aligned}
& \times \left[ \left( \frac{1}{18\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0^2 - 4\alpha_0 + 4)}{18} \right\} \right] \left( \frac{\Gamma(1)}{\Gamma(\frac{1}{2}) \sqrt{\pi} \sigma_B} \right)^p \\
& \times \prod_{k=1}^p \left( 1 + \frac{\alpha_k^2}{\sigma_B^2} \right)^{-1} \left( \frac{1}{1000 - 0.001} \right) I_{(0.001, 1000)}(\sigma_B^2) \\
& \times \left( \frac{1}{30 - 0.003} \right) I_{(0.003, 30)}(\sigma^2) \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right]
\end{aligned}$$

$$\begin{aligned}
\pi(\boldsymbol{\alpha}, \gamma_2, \sigma, \mathbf{Z} \mid \mathbf{X}, \mathbf{Y}) &= C \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(Y_i = j) \pi_{ij} \right\} \times \pi(\alpha_0) \prod_{k=1}^p \pi(\alpha_k) \times \pi(\gamma_2) \\
& \quad \times \pi(\sigma) \times \phi(Z_i, \mathbf{x}_{iz}^T \boldsymbol{\alpha}, \sigma^2) \\
&= C \prod_{i=1}^n \sum_{j=1}^4 I(Y_i = j) \left[ p \left[ \Phi \left( \frac{\gamma_j - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) - \Phi \left( \frac{\gamma_{j-1} - \mathbf{x}_{iz}^T \boldsymbol{\alpha}}{\sigma} \right) \right] + (1-p)q \right] \\
& \quad \times \left[ \left( \frac{1}{32\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0 - 2.5)^2}{32} \right\} \right] \left( \frac{\Gamma(\frac{1+1}{2})}{\Gamma(\frac{1}{2}) \sqrt{\pi(1)\sigma_B^2}} \right)^p \prod_{k=1}^p \left( 1 + \frac{1}{(1)} \frac{\alpha_k^2}{\sigma_B^2} \right)^{-\frac{1+1}{2}} \\
& \quad \times \left( \frac{1}{1000 - 0.001} \right) I_{(0.001, 1000)}(\sigma_B^2) \times \left( \frac{1}{40 - 0.004} \right) I_{(0.004, 40)}(\sigma^2) \\
& \quad \times \left[ \left( \frac{1}{8\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\gamma_2 - 2.5)^2}{8} \right\} \right] \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right] \\
&= C \prod_{i=1}^n \left[ \sum_{j=1}^4 I(Y_i = j) [pI(\gamma_{j-1} < Z_i < \gamma_j) + (1-p)q] \right] \\
& \quad \times \left[ \left( \frac{1}{32\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\alpha_0^2 - 5\alpha_0 + 6.25)}{32} \right\} \right] \left( \frac{1}{\Gamma(\frac{1}{2}) \sqrt{\pi} \sigma_B} \right)^p \prod_{k=1}^p \left( 1 + \frac{\alpha_k^2}{\sigma_B^2} \right)^{-1} \\
& \quad \times \left( \frac{1}{1000 - 0.001} \right) I_{(0.001, 1000)}(\sigma_B^2) \times \left( \frac{1}{40 - 0.004} \right) I_{(0.004, 40)}(\sigma^2) \\
& \quad \times \left[ \left( \frac{1}{8\pi} \right)^{\frac{1}{2}} \exp \left\{ \frac{-(\gamma_2^2 - 5\gamma_2 + 6.25)}{8} \right\} \right] \\
& \quad \times \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left\{ \frac{-\sum_{i=1}^n (Z_i - \mathbf{x}_{iz}^T \boldsymbol{\alpha})^2}{2\sigma^2} \right\} \right]
\end{aligned}$$