



ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

INFERENCIA VARIACIONAL PARA LA ESTIMACIÓN DE PARÁMETROS EN MODELOS DE TEORÍA DE RESPUESTA AL ÍTEM

**TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO
REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERA
MATEMÁTICA**

ERIKA ALEXANDRA GALINDES HERNÁNDEZ

erika.galindes@epn.edu.ec

DIRECTOR: CARLOS ALBERTO ALMEIDA RODRÍGUEZ

carlos.almeidar@epn.edu.ec

DMQ, SEPTIEMBRE 2022

CERTIFICACIONES

Yo, ERIKA ALEXANDRA GALINDES HERNÁNDEZ, declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.



Erika Alexandra Galindes Hernández

Certifico que el presente trabajo de integración curricular fue desarrollado por Erika Alexandra Galindes Hernández, bajo mi supervisión.

Carlos Alberto Almeida Rodríguez

DIRECTOR

DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el(los) producto(s) resultante(s) del mismo, es(son) público(s) y estará(n) a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

Erika Alexandra Galindes Hernández

Carlos Alberto Almeida Rodríguez

RESUMEN

La Teoría de Respuesta al Ítem (IRT, por sus siglas en inglés) trata de medir rasgos latentes a través de modelos estadísticos. En los últimos años se han llevado a cabo numerosas aplicaciones a la hora de evaluar del rendimiento académico, la medición de aptitudes, competencias, etc. [11]. En el presente trabajo se tratará de medir el interés político de un grupo de individuos, utilizando las respuestas proporcionadas a una encuesta. Existe una gran variedad de modelos que permiten estimar la probabilidad de responder correctamente a una pregunta, dados los parámetros de habilidad, dificultad del ítem y/o parámetro de pseudo-azar, según sea el caso. Más detalles acerca de estos modelos se muestran en [16].

A la hora de estimar la distribución de la variable latente, es bastante común recurrir a los métodos de Máxima Verosimilitud, Monte Carlo vía Cadenas de Markov, entre otros. Existe otra forma de estimar esta probabilidad, la cual será empleada en el presente trabajo, esta es la Inferencia Variacional; este método aproxima densidades de probabilidad mediante optimización. La idea es transformar el problema inicial para que nuestro nuevo objetivo sea escoger al miembro de una familia de distribuciones que mejor aproxima la densidad a-posteriori de la variable latente, el interés político en este caso, minimizando una estimación del desajuste entre las distribuciones verdadera y aproximada. Para esto, se usa la divergencia de Kullback-Leibler [17].

Palabras clave: Teoría de Respuesta al Ítem, IRT, Inferencia Variacional, VI, Kullback-Leibler, Variable Latente.

ABSTRACT

Item Response Theory (IRT) attempts to measure latent traits through statistical models. In recent years, many applications have been made for evaluating academic performance, measuring skills, competences, etc. [11]. The present work attempts to measure the political interest of a group of individuals using the answers provided in survey. There are a variety of models that allow to estimate the probability of answering a question correctly, given the ability, difficulty and/or pseudo-random parameters for each item, as the case may be. More details about these models are shown in [16].

When estimating the distribution of the latent variable, it is quite common to use Maximum Likelihood, Monte Carlo via Markov Chains methods, among others. There is another way to estimate this probability, which is used in the present work, that is Variational Inference; this method approximates densities of probability by optimization. The idea is to transform the initial problem to another where our new goal is to choose the member of a distribution family that best approximates the posterior density of the latent variable, the political interest in this case, minimizing an estimate of the distance between true and approximate distributions. For this, the Kullback-Leibler divergence is used [17].

Keywords: Item Response Theory, IRT, Variational Inference, VI, Kullback-Leibler, Latent Variable.

Índice general

1. Descripción del componente desarrollado	1
1.1. Objetivo general	1
1.2. Objetivos específicos	2
1.3. Alcance	2
1.4. Marco teórico	3
1.4.1. Variable Latente	3
1.4.2. Curva Característica del Ítem	3
1.4.3. Modelos tradicionales de la Curva Característica del Ítem	4
1.4.3.1. Modelo logístico	4
1.4.3.2. Modelo de Ojiva Normal	5
1.4.3.3. Partial Credit Model (PCM)	5
1.4.3.4. Graded Response Model (GRM)	5
1.4.4. Características y supuestos IRT	6
1.4.5. Inferencia Variacional	7
1.4.5.1. Divergencia de Kullback-Leibler	7
1.4.6. Problema	8
2. Metodología	10
2.1. Especificación del problema	10

2.1.0.1. Especificación de la familia de densidades Q .	10
2.1.1. Distribuciones a priori de los parámetros	11
2.1.1.1. Densidad conjunta	11
2.1.1.2. Densidad condicional	12
2.2. Algoritmo e implementación	13
2.3. Bondad de Ajuste	28
3. Resultados, conclusiones y recomendaciones	29
3.1. Resultados	29
3.2. Conclusiones y recomendaciones	34
3.2.1. Conclusiones	34
3.2.2. Recomendaciones	34
Bibliografía	36
A. Título anexo	38
A.1. Óptimo de los parámetros Variacionales	38
A.2. Implementación CAVI en Python	39

Índice de figuras

3.1. Category Response Curves Ítem 1	30
3.2. Category Response Curves Ítem 2	31
3.3. Category Response Curves Ítem 3	32
3.4. Distribución de la variable latente	33

Capítulo 1

Descripción del componente desarrollado

Se presentan, en primer lugar, algunos conceptos generales de la Teoría de Respuesta al Ítem; además de algunos de los posibles modelos para curva característica del ítem (ICC, por sus siglas en inglés). También se planteará el problema de optimización a resolver, y para ello se hará una breve revisión de varios algoritmos de Inferencia Variacional y definiciones.

Además, se determinará las características que definen a la familia de densidades sobre la que se va a optimizar. Una vez definido el problema se procede con la elección del algoritmo a utilizar para la estimación y su implementación. Finalmente, se filtra los datos y se realiza el ajuste del modelo y el análisis de los resultados.

1.1. Objetivo general

Aplicar Inferencia Variacional para la estimación de la distribución de la variable latente, el interés político de los franceses, utilizando los datos de la ronda 10 de la Encuesta Social Europea llevada a cabo en el año 2020 [7].

1.2. Objetivos específicos

1. Investigar los supuestos de la IRT, características y estimación de parámetros.
2. Estudiar algoritmos de Inferencia Variacional para la estimación de la densidad de la variable latente.
3. Determinar las características de la familia de densidades sobre la que se va a optimizar.
4. Filtrar los datos de la Encuesta Social Europea (ESS, por sus siglas en inglés) y realizar el ajuste el modelo.
5. Analizar y presentar los resultados.

1.3. Alcance

Para cumplir con el objetivo establecido, es necesario adquirir un conocimiento básico en la Teoría de Respuesta al ítem, así que se empezará por estudiar algunos conceptos y resultados básicos, que se pueden encontrar en [8]. También se realizará un estudio de los algoritmos de Inferencia Variacional y se llevará a cabo un estudio de las herramientas computacionales disponibles para la implementación del Algoritmo solución.

Una vez estudiados los conceptos procedemos con el tratamiento de la base de datos y luego se identificará la variable latente. Como dato para la estimación del parámetro de habilidad (variable latente), buscamos el modelo de la ICC que vincule la respuesta que proporciona el individuo a cada ítem con el nivel de habilidad; para ello, se realizarán pruebas sobre los posibles valores de los parámetros de cada modelo [8].

Después de haber identificado la Curva Característica del Ítem, se considera algunos de los posibles caminos por los cuales acercarse a la función objetivo; es decir, se define la familia de densidades sobre la cual se va a optimizar, se la dota de ciertas características para así completar la especificación del problema de optimización [4].

Finalmente, se realizará la estimación y las pruebas pertinentes para evaluar el ajuste del modelo; pues, explica Leenen [10] que, aunque se admite que ningún modelo es una representación perfecta del proceso cognitivo que subyace a los datos, se considera esencial evaluar a través de pruebas estadísticas de bondad de ajuste si es justificable mantener el modelo como estimación. Así, en caso de obtener un resultado poco satisfactorio, se realizan las respectivas modificaciones hasta obtener un modelo final con un ajuste aceptable, y posterior a todo este trabajo se presentarán los resultados, análisis, y se establecerán conclusiones.

1.4. Marco teórico

Esta sección pretende dar a conocer algunos de los conceptos de la IRT e Inferencia Variacional, los cuales se requiere conocer para cumplir con el objetivo del presente trabajo: la estimación de la densidad de variable latente, que como ya se ha mencionado puede representar la capacidad, interés, actitudes, etc., de un individuo. Estos conceptos se toman de los trabajos desarrollados por [8], [5], y para los conceptos referentes a la Inferencia Variacional se refiere a [3] y [18].

1.4.1. Variable Latente

Una variable latente se define como una variable aleatoria cuyas realizaciones no se pueden observar de forma directa. Las respuestas a los ítems nos proporcionan información para medir esta variable oculta. Ésta puede aparecer como una variable continua, o en muchos casos, como una variable categórica, [13].

1.4.2. Curva Característica del Ítem

La curva característica del ítem (ICC, por sus siglas en inglés) establece una relación funcional entre la variable latente y la probabilidad de obtener la respuesta correcta a un ítem, y se la puede expresar como sigue,

$$P_i(\theta) = f(\eta, \theta)$$

donde f es función de la variable latente y los parámetros del ítem, el parámetro de discriminación del ítem por ejemplo.[2]

1.4.3. Modelos tradicionales de la Curva Característica del Ítem

Los modelos buscan medir un constructo hipotético que se encuentra presente pero no se manifiesta, y puede ser medido únicamente a través de otras variables que sí se manifiestan, [8]. A este constructo se lo llama parámetro de habilidad (variable latente), y se lo notará por θ .

A continuación se muestran algunos de los modelos para la Curva Característica del Ítem que podemos encontrar en la literatura.

1.4.3.1. Modelo logístico

Para este modelo la probabilidad de obtener una respuesta correcta está dada por:

$$P(Y_{ik} = 1 | \theta_i, a_k, b_k) = \frac{\exp(a_k \theta_i - b_k)}{(1 + \exp(a_k \theta_i - b_k))} = (1 + \exp(b_k - a_k \theta_i))^{-1} \quad (1.1)$$

donde i es el individuo con parámetro de habilidad θ_i , b_k es el parámetro de dificultad del ítem, que es el punto en la escala de habilidad para el cual la probabilidad de obtener una respuesta correcta es de 0,5; y el parámetro de discriminación a_k , que determina la capacidad que tiene el ítem para diferenciar entre niveles altos y bajos del parámetro de habilidad. Cada ICC describe la relación entre el nivel de habilidad y la probabilidad de obtener una respuesta correcta para un ítem k en específico.

Al modelo dado por la expresión (1.1) se lo conoce como modelo logístico de dos parámetros. Si se considera un parámetro adicional, c_k (de adivinación) se lo denomina modelo logístico de tres parámetros; y a aquel que contiene únicamente al parámetro a_k se lo conoce como modelo logístico de un parámetro, [8]. A continuación se presenta otra versión de este modelo en la cual se expresa la probabilidad $P(Y_{ik} = 1 | \theta_i, a_k, b_k)$ usando la distribución normal estándar.

1.4.3.2. Modelo de Ojiva Normal

En este modelo, la ICC toma la forma de una distribución normal acumulada, y la probabilidad de obtener una respuesta correcta a un ítem k está dada por:

$$P(Y_{ik} = 1 | \theta_i, a_k, b_k) = \Phi(a_k \theta_i - b_k) = \int_{-\infty}^{a_k \theta_i - b_k} \phi(z) dz \quad (1.2)$$

donde $\Phi(\cdot)$ y $\phi(\cdot)$ son la función de distribución de una normal y su densidad, respectivamente. Los modelos de ojiva normal pueden ser de uno, dos o tres parámetros, al igual que los modelos logísticos. Más detalles acerca de estos modelos se pueden encontrar en [5].

Los modelos presentados anteriormente son válidos únicamente cuando se dispone de datos dicotómico. Los modelos que se presentan a continuación son para casos más generales que los anteriores y válidos para datos politómicos.

1.4.3.3. Partial Credit Model (PCM)

Se suele utilizar este modelo a la hora de tratar con datos politómicos. La probabilidad de obtener una respuesta para el ítem k en la categoría $c = 1, \dots, C_k$ está dada por:

$$P(Y_{ik} = c | \theta_i, a_k, \kappa_k) = \frac{\exp \sum_{l=1}^c (\theta_i - \kappa_{k,l})}{\sum_{r=1}^{C_k} (\exp \sum_{l=1}^r (\theta_i - \kappa_{k,l}))} \quad (1.3)$$

donde $\kappa_{k,l}$ es el parámetro *item step difficulty* del ítem k y categoría l .

1.4.3.4. Graded Response Model (GRM)

Al igual que el PCM, a este se lo suele usar con datos politómicos, y la probabilidad de obtener una respuesta para el ítem k en la categoría

$c = 1, \dots, C_k$ para un individuo i está dada por la expresión:

$$\begin{aligned} P(Y_{ik} = c | \theta_i, a_k, \kappa_k) &= P(Y_{ik} \geq c - 1 | \theta_i, a_k, \kappa_k) - P(Y_{ik} \geq c | \theta_i, a_k, \kappa_k) \\ &= \frac{1}{1 + \exp \{ \kappa_{k,c-1} - a_k \theta_i \}} - \frac{1}{1 + \exp \{ \kappa_{k,c} - a_k \theta_i \}} \end{aligned} \quad (1.4)$$

donde la probabilidad de obtener una respuesta en la categoría C_k o mayor es igual a 0, y la probabilidad de obtener una respuesta en la categoría mayor a 1 es igual a uno, y los parámetros a_k y $\kappa_{k,c}$ son los parámetros de discriminación del ítem y *step difficulty* para el ítem k y la categoría c .

Se prestará especial atención a este último modelo, ya que es el que se va a utilizar para el caso de aplicación: medir el interés político utilizando los datos de la encuesta social europea [7].

1.4.4. Características y supuestos IRT

Características

Los modelos IRT poseen muchas características deseables, una de ellas es que las estimaciones no dependen del test; es decir, si se realiza la estimación de una variable latente específica usando conjuntos de ítems diferentes pero que midan la misma variable, los resultados obtenidos en cada estimación son comparables. Una segunda característica que es importante mencionar es que las estimaciones de las características de los ítems no dependen de la muestra de individuos seleccionada para la estimación; es decir, si se realiza la estimación de los parámetros del ítem, del parámetro de discriminación por ejemplo, utilizando las respuestas a los test de diferentes grupos de individuos de la misma población, los resultados de las estimaciones son comparables, [8].

Supuestos

El primer supuesto que hace la IRT es que la curva característica del ítem describe completamente cómo cambia la probabilidad de una respuesta correcta según los cambios que se produzcan en la variable latente. La variedad de modelos IRT se definen por la forma esta curva, la ICC.

El segundo supuesto establece que, si los ítems miden una variable

latente unidimensional, es decir, un solo constructo, las respuestas para cada par de ítems son independientes. A este supuesto se lo puede expresar como sigue. Sea y_i un vector de longitud K de posibles respuestas a un test, cuyos valores de respuesta proporcionados por un individuo i con parámetro de habilidad θ_i , son: $y_i = (y_{i1}, y_{i2}, \dots, y_{iK})$. El supuesto de independencia local, también conocido como independencia condicional, se expresa:

$$P(y_i|\theta_i) = P(y_{i1}|\theta_i)P(y_{i2}|\theta_i) \dots P(y_{iK}|\theta_i) = \prod_{k=1}^K P(y_{ik}|\theta_i) \quad (1.5)$$

1.4.5. Inferencia Variacional

Los métodos de Inferencia Variacional son una herramienta que nos permite aproximar densidades, estos métodos tratan de encontrar la mejor aproximación de la densidad posterior buscando de entre una familia de distribuciones a la que más se acerca a la distribución que se desea aproximar. La distancia se mide utilizando la divergencia de Kullback-Leibler (KL), y la idea es llevar el problema a uno de optimización, cuya solución proporcionará una aproximación de la verdadera distribución posterior de la variable latente.

$$q^*(z) = \arg \min_{q(z) \in Q} \{\text{KL}(q(z) || p(z|\mathbf{y}))\} \quad (1.6)$$

Para más detalles ver [4] y [15].

1.4.5.1. Divergencia de Kullback-Leibler

Sean $\mathbf{y} = (y_1, \dots, y_n)$ un vector de observaciones de una variable, y $z = (z_1, \dots, z_m)$ variables latentes. La divergencia de Kullback-Leibler (KL) mide la cercanía entre dos distribuciones [18], en este caso, entre las distribuciones $q(z)$ y $p(z|\mathbf{y})$, y se expresa como sigue:

$$\text{KL}(q || p) = \int q(z) \log \frac{q(z)}{p(z|\mathbf{y})} = \mathbb{E} \left[\log \frac{q(z)}{p(z|\mathbf{y})} \right] \quad (1.7)$$

Desarrollando la expresión del lado derecho se obtiene la llamada cota inferior de la evidencia (ELBO, por sus siglas en inglés) [4] se tiene que,

$$\begin{aligned}\text{KL}(q || p) &= \mathbb{E}_q [\log q(z)] - \mathbb{E}_q [\log p(z|\mathbf{y})] \\ &= \mathbb{E}_q [\log q(z)] - \mathbb{E}_q [\log p(z, \mathbf{y})] + \mathbb{E}_q [\log p(\mathbf{y})] \\ &= \mathbb{E}_q [\log p(\mathbf{y})] - \{\mathbb{E}_q [\log p(z, \mathbf{y})] - \mathbb{E}_q [\log q(z)]\}\end{aligned}$$

luego,

$$\text{KL}(q || p) = \log p(\mathbf{y}) - \mathbf{ELBO}(q) \quad (1.8)$$

Así, el problema (1.6) es equivalente a resolver:

$$q^*(z) = \arg \max_{q(z) \in Q} \{\mathbf{ELBO}(q)\} \quad (1.9)$$

donde $q^*(z)$ es una aproximación de la densidad posterior $p(z|\mathbf{y})$.

1.4.6. Problema

El objetivo del presente trabajo es estimar la distribución de la variable θ , que mide el interés político de un individuo, para ello se requiere estimar otras variables que se encuentran en el modelo. Así, se define el vector $z = (\boldsymbol{\theta}, \boldsymbol{\eta})$, donde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ serían las realizaciones de la variable θ y $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_K)$ tal que $\boldsymbol{\eta}_k = (a_k, \kappa_{k,1}, \dots, \kappa_{k,C_k-1})$ el vector de parámetros de discriminación y *step difficulty* de los ítems $k = 1, \dots, K$ del test. Notemos que existen $m = n + \sum_{j=1}^K C_j$ variables latentes:

$$\begin{aligned}\boldsymbol{\theta} &= (\theta_1, \dots, \theta_n) \\ \boldsymbol{\eta} &= (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_K) = (a_1, \kappa_{1,1}, \dots, \kappa_{1,C_1-1}, \dots, a_K, \kappa_{K,1}, \dots, \kappa_{K,C_K-1})\end{aligned} \quad (1.10)$$

Se busca de entre las distribuciones que pertenecen a una familia de densidades Q a la densidad $q^*(z)$ que minimiza la divergencia de Kullback-Leibler, y ésta es una estimación de la distribución verdadera $q(z)$. Así, el problema a resolver se formula como en (1.6). Luego, considerando que la divergencia KL puede expresarse como en (1.8), y que el término $\log p(\mathbf{y})$ es constante para el problema de maximización, el problema (1.6)

es equivalente a (1.9); es decir,

$$q^*(z) = \arg \max_{q \in Q} \{\mathbb{E}_q [\log p(z, \mathbf{y})] - \mathbb{E}_q [\log q(z)]\} \quad (1.11)$$

La especificación del problema se presenta a detalle en el capítulo siguiente.

Capítulo 2

Metodología

2.1. Especificación del problema

Para definir completamente problema:

$$q^*(z) = \arg \max_{q \in Q} \{\mathbb{E}_q [\log p(z, \mathbf{y})] - \mathbb{E}_q [\log q(z)]\} \quad (2.1)$$

se requiere definir la familia Q a la que pertenece $q^*(z)$ y sobre la cual se va a optimizar, para este caso se utilizará *Mean-field variational inference*, que sugiere que la familia Q contenga densidades $q(z)$ cuyas variables sean independientes. Así,

$$q(z) = \prod_{j=1}^m q_j(z_j) \quad (2.2)$$

con $z = (z_1, \dots, z_m) = (\boldsymbol{\theta}, \boldsymbol{\eta})$ las variables latentes como en (1.10).

2.1.0.1. Especificación de la familia de densidades \mathcal{Q}

En el trabajo guía [4] se menciona que definiendo la familia \mathcal{Q} como una *Bayesian mixture of Gaussians* se obtiene buena precisión en la aproximación de la densidad posterior. Así, las características de la familia sobre la cual se va a optimizar son que sus variables son mutuamente

independientes y se distribuyen como una mezcla de gaussianas, es decir,

$$q(z) = q(\boldsymbol{\theta}, \boldsymbol{\eta}) = \prod_{i=1}^n q(\theta_i; \mu_\theta, \sigma_\theta^2) \prod_{k=1}^K \prod_{j=0}^{C_j-1} q(\eta_{k,j}; m_{k,j}, s_{k,j}^2) \quad (2.3)$$

donde,

$$q(\eta_{k,j}; m_{k,j}, s_{k,j}^2) \sim N(m_{k,j}, s_{k,j}^2) \quad (2.4)$$

$$q(\theta_i; \mu_\theta, \sigma_\theta^2) \sim N(\mu_\theta, \sigma_\theta^2) \quad (2.5)$$

Además, la variable θ que mide el interés político se distribuye como:

$$\theta \sim N(\mu_\theta, \sigma_\theta^2)$$

Los parámetros variacionales son $m_{k,j}, s_{k,j}^2, \mu_\theta, \sigma_\theta^2$, así tenemos el problema completamente definido. También será de utilidad tomar en cuenta que el $q^*(z)$ que minimiza la divergencia de Kullback-Leibler es tal que,

$$q^*(z) \propto \exp\{\log p(z, \mathbf{y})\} \quad (2.6)$$

con $z = (\boldsymbol{\theta}, \boldsymbol{\eta})$. El detalle del proceso por el cual se obtiene la expresión anterior se muestra en el Anexo 1.

2.1.1. Distribuciones a priori de los parámetros

Dado el problema (2.1), se requiere encontrar las expresiones para $p(z, \mathbf{y})$ y $q(z)$. Así,

2.1.1.1. Densidad conjunta

La densidad conjunta $p(z, \mathbf{y}) = p(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{y})$ puede escribirse como:

$$p(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{y}) = p(\boldsymbol{\theta}, \boldsymbol{\eta})p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\eta})$$

Y, gracias a las características de los ítems en la IRT que se presentaron en el capítulo 1: las propiedades de los ítems no dependen de la muestra de individuos seleccionada y que los individuos se consideran indepen-

dientes, se tiene que:

$$\begin{aligned} p(\boldsymbol{\eta}, \boldsymbol{\theta}, \mathbf{y}) &= p(\boldsymbol{\theta})p(\boldsymbol{\eta})p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\eta}) \\ &= p(\boldsymbol{\eta}) \prod_{i=1}^n p(\theta_i)p(y_i | \theta_i, \boldsymbol{\eta}) \end{aligned} \quad (2.7)$$

donde θ_i representa el nivel de habilidad del individuo i .

2.1.1.2. Densidad condicional

Gracias al supuesto de independencia local presentado en (1.5) se tiene que,

$$p(y_i | \theta_i, \boldsymbol{\eta}) = \prod_{k=1}^K p(y_{ik} | \theta_i, \boldsymbol{\eta}_k) \quad (2.8)$$

con $y_i = (y_{i1}, \dots, y_{iK})$ las respuestas del individuo i a los K ítems. Ahora, $p(y_{ik} | \theta_i, \boldsymbol{\eta}_k)$, por (1.4) está dada por la expresión siguiente:

$$p(y_{ik} | \theta_i, \boldsymbol{\eta}_k) = \begin{cases} \left(1 - \frac{1}{1 + \exp \{ \eta_{k,y_{ik}} - \eta_{k,0}\theta_i \}} \right), & \mathbf{si} \ y_{ik} = 1 \\ \left(\frac{1}{1 + \exp \{ \eta_{k,y_{ik}-1} - \eta_{k,0}\theta_i \}} - \frac{1}{1 + \exp \{ \eta_{k,y_{ik}} - \eta_{k,0}\theta_i \}} \right), & \mathbf{si} \ y_{ik} \in \{2, \dots, C_k - 1\} \\ \left(\frac{1}{1 + \exp \{ \eta_{k,y_{ik}-1} - \eta_{k,0}\theta_i \}} \right), & \mathbf{si} \ y_{ik} = C_k \end{cases} \quad (2.9)$$

con C_k el número de categorías del ítem k .

Ahora, necesitamos definir las distribuciones a priori de los parámetros de discriminación a_k para cada ítem y los parámetros *step difficulty* $\kappa_{k,c}$ donde $c = 1, \dots, C_k$ con C_k el número de categorías del ítem k . Para las asunciones respecto a estas distribuciones se usará como guía [8], capítulo 4.

Se comienza por definir un vector con todos los parámetros que se desea estimar para cada ítem k ,

$$\boldsymbol{\eta}_k = (\alpha_k, \kappa_{k,1}, \dots, \kappa_{k,C_k-1}) \quad (2.10)$$

y las distribuciones a priori de cada uno de los parámetros son:

$$a_k \sim N(0,1) \text{ truncada en } [0, \infty) \quad (2.11)$$

$$\eta_{k,j} \sim U(-1, 1) \text{ para } j = 1, \dots, C_k - 1 \quad (2.12)$$

También se tiene que los parámetros de habilidad de los individuos θ_i se distribuyen de forma independiente y de acuerdo a una distribución normal,

$$\theta_i \sim N(\mu_\theta, \sigma_\theta^2) \quad i = 1, \dots, n \quad (2.13)$$

donde $\mu_\theta = 0$ y $\sigma_\theta^2 = 1$ para fijar el origen y la escala de habilidad.

2.2. Algoritmo e implementación

Para la implementación se usará el algoritmo CAVI (Coordinate Ascent Mean-Field Variational Inference) [4], que en cada iteración actualiza los parámetros variacionales hasta que converjan a un óptimo local mientras los demás parámetros se mantienen fijos. El algoritmo utiliza el siguiente resultado.

Si se considera la j -ésima variable latente, $j = 1, \dots, m$, la densidad condicional completa de z_j es $p(z_j | z_{-j}, \mathbf{y})$; es decir, la densidad de z_j dadas las observaciones \mathbf{y} y las variables latentes distintas de z_j . Fijados los demás parámetros variacionales, el óptimo $q_j^*(z_j)$ satisface:

$$q_j^*(z_j) \approx \exp \{E_{-j} [\log p(z_j | z_{-j}, \mathbf{y})]\} \quad (2.14)$$

donde $E_{-j} [\cdot]$ es la esperanza con respecto a la densidad variacional $\prod_{l \neq j} q_l(z_l)$.

El algoritmo tiene la siguiente estructura:

A continuación se determinará el ELBO y las actualizaciones de los parámetros variacionales. El ELBO es función de los parámetros $m, \mathbf{s}^2, \mu_\theta, \sigma_\theta^2$,

Algorithm 1: CAVI

Entrada: La matriz de Datos $Y_{n \times K}$, con n individuos y K ítems;
Salida: Densidades normales $q(\eta_{k,j}; m_{k,j}, s_{k,j}^2)$ y $q(\theta; \mu_\theta, \sigma_\theta^2)$;
Inicialización: Cálculo de las varianzas σ_θ^2 y $s_{k,j}^2$ y valores iniciales para los parámetros variacionales μ_θ^2 y $m_{k,j}$;
while el ELBO no converge **do**
 for $i \in \{1, \dots, N\}$ **do**
 | Actualizar μ_θ
 end
 for $k \in \{1, \dots, K\}$ **do**
 | Actualizar las medias
 end
 Calcular ELBO
end
return $q(\eta_{k,j}; m_{k,j}, s_{k,j}^2)$ y $q(\theta; \mu_\theta, \sigma_\theta^2)$

y se puede escribir, por (2.7), (2.8) y (2.3), como sigue:

$$\begin{aligned} \text{ELBO}(\mathbf{m}, \mathbf{s}^2, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) &= E [\log p(z, y)] - E [\log q(z)] \\ &= E \left[\log \left(\prod_{k=1}^K \prod_{j=0}^{C_k-1} p(\eta_{k,j}) \right) \right] + E \left[\log \left(\prod_{i=1}^n p(\theta_i) p(y_i | \theta_i, \boldsymbol{\eta}) \right) \right] \\ &\quad - E \left[\log \left(\prod_{k=1}^K \prod_{j=0}^{C_k-1} q(\eta_{k,j}; m_{k,j}, s_{k,j}^2) \right) \right] - E \left[\log \left(\prod_{i=1}^n q(\theta_i; \mu_\theta, \sigma_\theta^2) \right) \right] \end{aligned}$$

Y, se obtiene la siguiente expresión, donde se puede observar claramente la dependencia sobre los parámetros variacionales:

$$\begin{aligned} \text{ELBO}(\mathbf{m}, \mathbf{s}^2, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) &= \sum_{k=1}^K \sum_{j=0}^{C_k-1} E [\log p(\eta_{k,j}); m_{k,j}, s_{k,j}^2] \\ &\quad + \sum_{i=1}^n (E [\log p(\theta_i); \mu_\theta, \sigma_\theta^2] + E [\log p(y_i | \theta_i, \boldsymbol{\eta}); \mu_\theta, \sigma_\theta^2]) \\ &\quad - \sum_{k=1}^K \sum_{j=0}^{C_k-1} E [\log q(\eta_{k,j}; m_{k,j}, s_{k,j}^2)] - \sum_{i=1}^n E [\log q(\theta_i; \mu_\theta, \sigma_\theta^2)] \quad (2.15) \end{aligned}$$

Considerando las asunciones acerca de las distribuciones de los paráme-

tros detalladas en (2.11), (2.13) y 2.4 se tiene que,

$$\begin{aligned}
\log p(\eta_{k,0}) &= -\log(\sqrt{2\pi}(1 - \Phi(0))) - \frac{\eta_{k,0}^2}{2}, \quad \eta_{k,0} \geq 0 \\
\log p(\eta_{k,j}) &= \log\left(\frac{1}{2}\right) \quad \text{para } j = 1, \dots, C_k - 1 \\
\log p(\theta_i) &= -\frac{1}{2} \log(2\pi) - \frac{\theta_i^2}{2} \\
\log q(\eta_{k,j}) &= -\frac{1}{2} \log(2\pi s_{k,j}^2) - \frac{(\eta_{k,j} - m_{k,j})^2}{2s_{k,j}^2} \\
\log q(\theta_i) &= -\frac{1}{2} \log(2\pi\sigma_\theta^2) - \frac{(\theta_i - \mu_\theta)^2}{2\sigma_\theta^2} \\
\log p(y_i | \theta_i, \boldsymbol{\eta}) &= \sum_{k=1}^K \log p(y_{ik} | \theta_i, \boldsymbol{\eta})
\end{aligned}$$

y al calcular la esperanza de cada uno de estos términos obtenemos,

$$\begin{aligned}
E[\log p(\eta_{k,0})] &= -\log(\sqrt{2\pi}(1 - \Phi(0))) - \frac{E[\eta_{k,0}^2]}{2} = -\log(\sqrt{2\pi}(1 - \Phi(0))) - \frac{(s_{k,0}^2 + m_{k,0}^2)}{2} \\
E[\log p(\eta_{k,j})] &= \log\left(\frac{1}{2}\right) \\
E[\log p(\theta_i)] &= -\frac{1}{2} \log(2\pi) - \frac{E[\theta_i^2]}{2} = -\frac{1}{2} \log(2\pi) - \frac{(\sigma_\theta^2 + \mu_\theta^2)}{2} \\
E[\log q(\eta_{k,j})] &= -\frac{1}{2} \log(2\pi s_{k,j}^2) - \frac{E[(\eta_{k,j} - m_{k,j})^2]}{2s_{k,j}^2} = -\frac{1}{2} \log(2\pi s_{k,j}^2) - \frac{1}{2} \\
E[\log q(\theta_i)] &= -\frac{1}{2} \log(2\pi\sigma_\theta^2) - \frac{E[(\theta_i - \mu_\theta)^2]}{2\sigma_\theta^2} = -\frac{1}{2} \log(2\pi\sigma_\theta^2) - \frac{1}{2}
\end{aligned}$$

Para $E[\log p(y_i | \theta_i, \boldsymbol{\eta})] = \sum_{k=1}^K E[\log p(y_{ik} | \theta_i, \boldsymbol{\eta})]$ donde $\log p(y_{ik} | \theta_i, \boldsymbol{\eta})$ tiene la forma:

$$\log p(y_{ik} | \theta_i, \boldsymbol{\eta}) = \begin{cases} \log\left(1 - \frac{1}{1 + \exp\{\eta_{k,y_{ik}} - \eta_{k,0}\theta_i\}}\right), & \mathbf{si} \ y_{ik} = 1 \\ \log\left(\frac{1}{1 + \exp\{\eta_{k,y_{ik}-1} - \eta_{k,0}\theta_i\}} - \frac{1}{1 + \exp\{\eta_{k,y_{ik}} - \eta_{k,0}\theta_i\}}\right), & \mathbf{si} \ y_{ik} \in \{2, \dots, C_k - 1\} \\ \log\left(\frac{1}{1 + \exp\{\eta_{k,y_{ik}-1} - \eta_{k,0}\theta_i\}}\right), & \mathbf{si} \ y_{ik} = C_k \end{cases} \quad (2.16)$$

Se debe tomar en cuenta que para esta expresión resulta sumamente complicado calcular la esperanza, por lo que se utilizará una aproxima-

ción de $\log p(y_i | \theta_i, \boldsymbol{\eta})$.

Primero, se encontrará una aproximación para el caso en que $y_{ik} = 1$. Llamaremos f , a la función:

$$f(\theta_i, \eta_{k,0}, \eta_{k,y_{ik}}) = \log \left(1 - \frac{1}{1 + \exp \{ \eta_{k,y_{ik}} - \eta_{k,0} \theta_i \}} \right)$$

El polinomio de Taylor de segundo orden para la función $f(x)$ alrededor del punto $a = (\theta_i, \eta_{k,0}, \eta_{k,y_{ik}}) = (\frac{1}{4}, \frac{1}{2}, -\frac{1}{2})$ está dado por (2.19), y se calculan las derivadas parciales:

$$\begin{aligned} d_0 &= \ln \left(1 - \frac{1}{1 + e^{-\frac{5}{8}}} \right) \\ d_1 &= \frac{\partial f}{\partial \theta_i} \Big|_a = -\frac{e^{\frac{5}{8}}}{2(e^{\frac{5}{8}} + 1)}, \quad d_2 = \frac{\partial f}{\partial \eta_{k,0}} \Big|_a = -\frac{e^{\frac{5}{8}}}{4(e^{\frac{5}{8}} + 1)}, \quad d_3 = \frac{\partial f}{\partial \eta_{k,y_{ik}}} \Big|_a = \frac{e^{\frac{5}{8}}}{(e^{\frac{5}{8}} + 1)} \\ d_4 &= \frac{\partial^2 f}{\partial \theta_i^2} \Big|_a = -\frac{e^{\frac{5}{8}}}{4(e^{\frac{5}{8}} + 1)^2}, \quad d_5 = \frac{\partial^2 f}{\partial \eta_{k,0}^2} \Big|_a = -\frac{e^{\frac{5}{8}}}{16(e^{\frac{5}{8}} + 1)^2}, \quad d_6 = \frac{\partial^2 f}{\partial \eta_{k,y_{ik}}^2} \Big|_a = -\frac{e^{\frac{5}{8}}}{(e^{\frac{5}{8}} + 1)^2} \\ d_7 &= \frac{\partial^2 f}{\partial \eta_{k,0} \partial \theta_i} \Big|_a = \frac{\partial^2 f}{\partial \theta_i \partial \eta_{k,0}} \Big|_a = -\frac{8e^{\frac{5}{4}} + 9e^{\frac{5}{8}}}{8(e^{\frac{5}{8}} + 1)^2} \\ d_8 &= \frac{\partial^2 f}{\partial \eta_{k,y_{ik}} \partial \theta_i} \Big|_a = \frac{\partial^2 f}{\partial \theta_i \partial \eta_{k,y_{ik}}} \Big|_a = \frac{e^{\frac{5}{8}}}{2(e^{\frac{5}{8}} + 1)^2} \\ d_9 &= \frac{\partial^2 f}{\partial \eta_{k,y_{ik}} \partial \eta_{k,0}} \Big|_a = \frac{\partial^2 f}{\partial \eta_{k,0} \partial \eta_{k,y_{ik}}} \Big|_a = \frac{e^{\frac{5}{8}}}{4(e^{\frac{5}{8}} + 1)^2} \end{aligned}$$

Así, la aproximación para función $\log p(y_{ik} | \theta_i, \boldsymbol{\eta})$ para el caso en que $y_{ik} = 1$ es:

$$\begin{aligned} \log p(y_{ik} | \theta_i, \boldsymbol{\eta}) &\approx d_0 + d_1 \left(\theta_i - \frac{1}{4} \right) + d_2 \left(\eta_{k,0} - \frac{1}{2} \right) + d_3 \left(\eta_{k,y_{ik}} + \frac{1}{2} \right) \\ &\quad + \frac{1}{2} d_4 \left(\theta_i - \frac{1}{4} \right)^2 + \frac{1}{2} d_5 \left(\eta_{k,0} - \frac{1}{2} \right)^2 + \frac{1}{2} d_6 \left(\eta_{k,y_{ik}} + \frac{1}{2} \right)^2 \\ &\quad + d_7 \left(\theta_i - \frac{1}{4} \right) \left(\eta_{k,0} - \frac{1}{2} \right) + d_8 \left(\theta_i - \frac{1}{4} \right) \left(\eta_{k,y_{ik}} + \frac{1}{2} \right) \\ &\quad + d_9 \left(\eta_{k,0} - \frac{1}{2} \right) \left(\eta_{k,y_{ik}} + \frac{1}{2} \right) \end{aligned} \tag{2.17}$$

luego,

$$\begin{aligned} \log p(y_{ik} | \theta_i, \boldsymbol{\eta}) \approx & a_0 + a_1\theta_i + a_2\eta_{k,0} + a_3\eta_{k,y_{ik}} - a_4\theta_i^2 - a_5\eta_{k,0}^2 - a_6\eta_{k,y_{ik}}^2 \\ & - a_7\theta_i\eta_{k,0} + a_8\theta_i\eta_{k,y_{ik}} + a_9\eta_{k,0}\eta_{k,y_{ik}}, \quad \mathbf{si} \ y_{ik} = 1 \end{aligned} \quad (2.18)$$

los valores de los coeficientes a_i con $i = 1, \dots, 9$, se encuentran en 2.22.

Ahora, se encontrará una aproximación para el caso en que $y_{ik} \in \{2, \dots, C_k\}$. Llamaremos f , a la función:

$$f(\theta_i, \eta_{k,0}, \eta_{k,y_{ik}-1}, \eta_{k,y_{ik}}) = \log \left(\frac{1}{1 + \exp \{ \eta_{k,y_{ik}-1} - \eta_{k,0}\theta_i \}} - \frac{1}{1 + \exp \{ \eta_{k,y_{ik}} - \eta_{k,0}\theta_i \}} \right)$$

El polinomio de Taylor de segundo orden para la función $f(x)$ alrededor del punto $a = (\theta_i, \eta_{k,0}, \eta_{k,y_{ik}-1}, \eta_{k,y_{ik}}) = (\frac{1}{4}, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2})$ está dado por:

$$f(a) + \sum_{l=1}^4 \frac{\partial f}{\partial x_l}(a) \cdot (x_l - a_l) + \frac{1}{2} \sum_{l=1}^4 \sum_{j=1}^4 \frac{\partial^2 f}{\partial x_l \partial x_j}(a) \cdot (x_l - a_l)(x_j - a_j) \quad (2.19)$$

Notemos que,

$$\begin{aligned} d_0 &= \ln \left(\frac{e-1}{1+e+e^{\frac{5}{8}}+e^{\frac{3}{8}}} \right) \quad ; \quad d_1 = \left. \frac{\partial f}{\partial \theta_i} \right|_a = -\frac{(e^{\frac{1}{4}}-1)}{2(e^{\frac{1}{8}}+e^{\frac{1}{2}})(e^{\frac{1}{8}}+e^{-\frac{1}{2}})} \\ d_2 &= \left. \frac{\partial f}{\partial \eta_{k,0}} \right|_a = -\frac{(e^{\frac{1}{4}}-1)}{4(e^{\frac{1}{8}}+e^{\frac{1}{2}})(e^{\frac{1}{8}}+e^{-\frac{1}{2}})} \quad ; \quad d_3 = \left. \frac{\partial f}{\partial \eta_{k,y_{ik}-1}} \right|_a = \frac{(e^{-\frac{3}{8}}+1)}{(e^{-\frac{1}{2}}-e^{\frac{1}{2}})(e^{-\frac{1}{2}}+e^{\frac{1}{8}})} \\ d_4 &= \left. \frac{\partial f}{\partial \eta_{k,y_{ik}}} \right|_a = \frac{(e^{\frac{5}{8}}+1)}{(e^{\frac{1}{2}}-e^{-\frac{1}{2}})(e^{\frac{1}{2}}+e^{\frac{1}{8}})} \quad ; \quad d_5 = \left. \frac{\partial^2 f}{\partial \theta_i^2} \right|_a = -\frac{4e^{\frac{1}{4}}+(e^{-\frac{1}{2}}+e^{\frac{1}{2}})(e^{\frac{1}{4}}+1)e^{\frac{1}{8}}}{4(e^{\frac{1}{8}}+e^{\frac{1}{2}})^2(e^{\frac{1}{8}}+e^{-\frac{1}{2}})^2} \\ d_6 &= \left. \frac{\partial^2 f}{\partial \eta_{k,0}^2} \right|_a = -\frac{4e^{\frac{1}{4}}+(e^{-\frac{1}{2}}+e^{\frac{1}{2}})(e^{\frac{1}{4}}+1)e^{\frac{1}{8}}}{16(e^{\frac{1}{8}}+e^{\frac{1}{2}})^2(e^{\frac{1}{8}}+e^{-\frac{1}{2}})^2} \\ d_7 &= \left. \frac{\partial^2 f}{\partial \eta_{k,y_{ik}-1}^2} \right|_a = -\frac{(e^{-\frac{3}{8}}+1)(e^{-1}+e^{\frac{5}{8}})}{(e^{-\frac{1}{2}}-e^{\frac{1}{2}})^2(e^{-\frac{1}{2}}+e^{\frac{1}{8}})^2} \\ d_8 &= \left. \frac{\partial^2 f}{\partial \eta_{k,y_{ik}}^2} \right|_a = -\frac{(e^{\frac{5}{8}}+1)(e^1+e^{-\frac{3}{8}})}{(e^{\frac{1}{2}}-e^{-\frac{1}{2}})^2(e^{\frac{1}{2}}+e^{\frac{1}{8}})^2} \end{aligned}$$

Y las derivadas mixtas,

$$\begin{aligned}
d_9 &= \frac{\partial^2 f}{\partial \theta_i \partial \eta_{k,0}} \Big|_a = - \frac{\partial^2 f}{\partial \eta_{k,0} \partial \theta_i} \Big|_a = \frac{\frac{1}{2}e^{\frac{1}{4}} + e^{\frac{1}{2}} + (e^{-\frac{1}{2}} + e^{\frac{1}{2}}) \left(\frac{9}{8}e^{\frac{3}{8}} - \frac{7}{8}e^{\frac{1}{8}} \right)}{\left(e^{\frac{1}{8}} + e^{\frac{1}{2}} \right)^2 \left(e^{\frac{1}{8}} + e^{-\frac{1}{2}} \right)^2} \\
d_{10} &= \frac{\partial^2 f}{\partial \theta_i \partial \eta_{k,y_{ik}-1}} = \frac{\partial^2 f}{\partial \eta_{k,y_{ik}-1} \partial \theta_i} \Big|_a = \frac{e^{\frac{5}{8}}}{2 \left(e^{\frac{5}{8}} + 1 \right)^2} \\
d_{11} &= \frac{\partial^2 f}{\partial \theta_i \partial \eta_{k,y_{ik}}} = \frac{\partial^2 f}{\partial \eta_{k,y_{ik}} \partial \theta_i} \Big|_a = \frac{e^{\frac{5}{8}}}{2 \left(e^{\frac{1}{2}} + e^{\frac{1}{8}} \right)^2} \\
d_{12} &= \frac{\partial^2 f}{\partial \eta_{k,y_{ik}} \partial \eta_{k,y_{ik}-1}} \Big|_a = \frac{\partial^2 f}{\partial \eta_{k,y_{ik}-1} \partial \eta_{k,y_{ik}}} \Big|_a = \frac{1}{\left(e^{\frac{1}{2}} - e^{-\frac{1}{2}} \right)^2} \\
d_{13} &= \frac{\partial^2 f}{\partial \eta_{k,0} \partial \eta_{k,y_{ik}}} = \frac{\partial^2 f}{\partial \eta_{k,y_{ik}} \partial \eta_{k,0}} \Big|_a = \frac{e^{\frac{5}{8}}}{4 \left(e^{\frac{1}{2}} + e^{\frac{1}{8}} \right)^2}; \\
d_{14} &= \frac{\partial^2 f}{\partial \eta_{k,0} \partial \eta_{k,y_{ik}-1}} = \frac{\partial^2 f}{\partial \eta_{k,y_{ik}-1} \partial \eta_{k,0}} \Big|_a = \frac{e^{\frac{5}{8}}}{4 \left(e^{\frac{5}{8}} + 1 \right)^2}
\end{aligned}$$

Por tanto, la aproximación de $f(\theta_i, \eta_{k,0}, \eta_{k,y_{ik}-1}, \eta_{k,y_{ik}}) = \log p(y_{ik} | \theta_i, \boldsymbol{\eta})$ para el caso en que $y_{ik} \in \{2, \dots, C_k - 1\}$ es:

$$\begin{aligned}
\log p(y_{ik} | \theta_i, \boldsymbol{\eta}) &\approx d_0 + d_1 \left(\theta_i - \frac{1}{4} \right) + d_2 \left(\eta_{k,0} - \frac{1}{2} \right) + d_3 \left(\eta_{k,y_{ik}-1} + \frac{1}{2} \right) + d_4 \left(\eta_{k,y_{ik}} - \frac{1}{2} \right) \\
&\quad + \frac{1}{2} d_5 \left(\theta_i - \frac{1}{4} \right)^2 + \frac{1}{2} d_6 \left(\eta_{k,0} - \frac{1}{2} \right)^2 + \frac{1}{2} d_7 \left(\eta_{k,y_{ik}-1} + \frac{1}{2} \right)^2 \\
&\quad + \frac{1}{2} d_8 \left(\eta_{k,y_{ik}} - \frac{1}{2} \right)^2 + d_9 \left(\eta_{k,0} - \frac{1}{2} \right) \left(\theta_i - \frac{1}{4} \right) + d_{10} \left(\theta_i - \frac{1}{2} \right) \left(\eta_{k,y_{ik}-1} + \frac{1}{2} \right) \\
&\quad + d_{11} \left(\theta_i - \frac{1}{2} \right) \left(\eta_{k,y_{ik}} - \frac{1}{2} \right) + d_{12} \left(\eta_{k,y_{ik}} - \frac{1}{2} \right) \left(\eta_{k,y_{ik}-1} + \frac{1}{2} \right) \\
&\quad + d_{13} \left(\eta_{k,0} - \frac{1}{2} \right) \left(\eta_{k,y_{ik}} - \frac{1}{2} \right) + d_{14} \left(\eta_{k,0} - \frac{1}{2} \right) \left(\eta_{k,y_{ik}-1} + \frac{1}{2} \right)
\end{aligned}$$

luego,

$$\begin{aligned}
\log p(y_{ik} | \theta_i, \boldsymbol{\eta}) &\approx b_0 + b_1 \theta_i + b_2 \eta_{k,0} - b_3 \eta_{k,y_{ik}-1} + b_4 \eta_{k,y_{ik}} \\
&\quad - b_5 \theta_i^2 - b_6 \eta_{k,0}^2 - b_7 \eta_{k,y_{ik}-1}^2 - b_8 \eta_{k,y_{ik}}^2 \\
&\quad - b_9 \eta_{k,0} \theta_i + b_{10} \theta_i \eta_{k,y_{ik}-1} + b_{11} \theta_i \eta_{k,y_{ik}} + b_{12} \eta_{k,0} \eta_{k,y_{ik}} \\
&\quad + b_{13} \eta_{k,0} \eta_{k,y_{ik}-1} + b_{14} \eta_{k,y_{ik}} \eta_{k,y_{ik}-1} \quad \mathbf{si} \ y_{ik} \in \{2, \dots, C_k - 1\}
\end{aligned}$$

los valores de los coeficientes b_i con $i = 1, \dots, 14$ se encuentran en [2.22](#).

Finalmente, se encontrará una aproximación para el caso en que $y_{ik} = C_k$. Llamaremos f , a la función:

$$f(\theta_i, \eta_{k,0}, \eta_{k,y_{ik}}) = \log \left(\frac{1}{1 + \exp \{ \eta_{k,y_{ik}-1} - \eta_{k,0} \theta_i \}} \right)$$

El polinomio de Taylor de segundo orden para la función $f(x)$ alrededor del punto $a = (\theta_i, \eta_{k,0}, \eta_{k,y_{ik}}) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{2})$ está dado por [\(2.19\)](#), y se calculan las derivadas parciales:

$$\begin{aligned} d_0 &= \ln \left(\frac{1}{1 + e^{\frac{3}{8}}} \right) \\ d_1 &= \frac{\partial f}{\partial \theta_i} \Big|_a = \frac{e^{\frac{1}{2}}}{2 \left(e^{\frac{5}{8}} + e^{\frac{1}{2}} \right)}, \quad d_2 = \frac{\partial f}{\partial \eta_{k,0}} \Big|_a = \frac{e^{\frac{1}{2}}}{4 \left(e^{\frac{1}{8}} + e^{\frac{1}{2}} \right)} \\ d_3 &= \frac{\partial f}{\partial \eta_{k,y_{ik}-1}} \Big|_a = -\frac{e^{\frac{1}{2}}}{\left(e^{\frac{5}{8}} + e^{\frac{1}{2}} \right)}, \quad d_4 = \frac{\partial^2 f}{\partial \theta_i^2} \Big|_a = -\frac{e^{\frac{5}{8}}}{4 \left(e^{\frac{1}{8}} + e^{\frac{1}{2}} \right)^2} \\ d_5 &= \frac{\partial^2 f}{\partial \eta_{k,0}^2} \Big|_a = -\frac{e^{\frac{5}{8}}}{16 \left(e^{\frac{1}{8}} + e^{\frac{1}{2}} \right)^2}, \quad d_6 = \frac{\partial^2 f}{\partial \eta_{k,y_{ik}}^2} \Big|_a = -\frac{e^{\frac{5}{8}}}{\left(e^{\frac{1}{8}} + e^{\frac{1}{2}} \right)^2} \\ d_7 &= \frac{\partial^2 f}{\partial \eta_{k,0} \partial \theta_i} \Big|_a = \frac{7e^{\frac{5}{8}} + 8e}{8 \left(e^{\frac{1}{8}} + e^{\frac{1}{2}} \right)^2}, \quad d_8 = \frac{\partial^2 f}{\partial \eta_{k,y_{ik}-1} \partial \theta_i} \Big|_a = \frac{\partial^2 f}{\partial \theta_i \partial \eta_{k,y_{ik}-1}} \Big|_a = \frac{e^{\frac{5}{8}}}{2 \left(e^{\frac{1}{8}} + e^{\frac{1}{2}} \right)^2} \\ d_9 &= \frac{\partial^2 f}{\partial \eta_{k,y_{ik}-1} \partial \eta_{k,0}} \Big|_a = \frac{\partial^2 f}{\partial \eta_{k,0} \partial \eta_{k,y_{ik}-1}} \Big|_a = \frac{e^{\frac{5}{8}}}{4 \left(e^{\frac{1}{8}} + e^{\frac{1}{2}} \right)^2} \end{aligned}$$

Así, la aproximación para la densidad $\log p(y_{ik} | \theta_i, \boldsymbol{\eta})$ para el caso en que $y_{ik} = C_k$ es:

$$\begin{aligned} \log p(y_{ik} | \theta_i, \boldsymbol{\eta}) &\approx d_0 + d_1 \left(\theta_i - \frac{1}{4} \right) + d_2 \left(\eta_{k,0} - \frac{1}{2} \right) + d_3 \left(\eta_{k,y_{ik}} - \frac{1}{2} \right) \\ &\quad + \frac{1}{2} d_4 \left(\theta_i - \frac{1}{4} \right)^2 + \frac{1}{2} d_5 \left(\eta_{k,0} - \frac{1}{2} \right)^2 + \frac{1}{2} d_6 \left(\eta_{k,y_{ik}} - \frac{1}{2} \right)^2 \\ &\quad + d_7 \left(\theta_i - \frac{1}{4} \right) \left(\eta_{k,0} - \frac{1}{2} \right) + d_8 \left(\theta_i - \frac{1}{4} \right) \left(\eta_{k,y_{ik}} - \frac{1}{2} \right) \\ &\quad + d_9 \left(\eta_{k,0} - \frac{1}{2} \right) \left(\eta_{k,y_{ik}} - \frac{1}{2} \right) \end{aligned} \tag{2.20}$$

luego,

$$\begin{aligned} \log p(y_{ik} | \theta_i, \boldsymbol{\eta}) \approx & c_0 + c_1\theta_i + c_2\eta_{k,0} + c_3\eta_{k,y_{ik}} - c_4\theta_i^2 - c_5\eta_{k,0}^2 - c_6\eta_{k,y_{ik}}^2 \\ & - c_7\theta_i\eta_{k,0} + c_8\theta_i\eta_{k,y_{ik}} + c_9\eta_{k,0}\eta_{k,y_{ik}}, \quad \mathbf{si} \ y_{ik} = C_k \end{aligned} \quad (2.21)$$

los valores c_i con $i = 1, \dots, 9$, se muestran a continuación,

$$\begin{aligned} \mathbf{a}_0 = -0,71047; \quad \mathbf{a}_1 = 0,08515; \quad \mathbf{a}_2 = 0,04257; \quad \mathbf{a}_3 = 0,48103; \quad \mathbf{a}_4 = 0,02838 \\ \mathbf{a}_5 = 0,00709; \quad \mathbf{a}_6 = 0,11354; \quad \mathbf{a}_7 = 0,67974; \quad \mathbf{a}_8 = 0,11354; \quad \mathbf{a}_9 = 0,05677 \end{aligned}$$

$$\begin{aligned} \mathbf{b}_0 = -2,89485; \quad \mathbf{b}_1 = 0,05498; \quad \mathbf{b}_2 = 0,02749; \quad \mathbf{b}_3 = 2,0755; \quad \mathbf{b}_4 = 1,97033 \\ \mathbf{b}_5 = 0,05856; \quad \mathbf{b}_6 = 0,01464; \quad \mathbf{b}_7 = 0,57388; \quad \mathbf{b}_8 = 0,58104; \quad \mathbf{b}_9 = 0,11725 \\ \mathbf{b}_{10} = 0,11354; \quad \mathbf{b}_{11} = 0,1207; \quad \mathbf{b}_{12} = 0,06035; \quad \mathbf{b}_{13} = 0,05677; \quad \mathbf{b}_{14} = 0,92067 \end{aligned}$$

$$\begin{aligned} \mathbf{c}_0 = -0,68341; \quad \mathbf{c}_1 = -0,02017; \quad \mathbf{c}_2 = -0,01508; \quad \mathbf{c}_3 = 0,53231; \quad \mathbf{c}_4 = 0,03017 \\ \mathbf{c}_5 = 0,00754; \quad \mathbf{c}_6 = 0,12070; \quad \mathbf{c}_7 = 0,56248; \quad \mathbf{c}_8 = 0,1207; \quad \mathbf{c}_9 = 0,06035, \end{aligned} \quad (2.22)$$

En resumen, una aproximación de la función:

$$\log p(y_{ik} | \theta_i, \boldsymbol{\eta}) = \begin{cases} \log \left(1 - \frac{1}{1 + \exp \{ \eta_{k,y_{ik}} - \eta_{k,0}\theta_i \}} \right), & \mathbf{si} \ y_{ik} = 1 \\ \log \left(\frac{1}{1 + \exp \{ \eta_{k,y_{ik}-1} - \eta_{k,0}\theta_i \}} - \frac{1}{1 + \exp \{ \eta_{k,y_{ik}} - \eta_{k,0}\theta_i \}} \right), & \mathbf{si} \ y_{ik} \in \{2, \dots, C_k - 1\} \\ \log \left(\frac{1}{1 + \exp \{ \eta_{k,y_{ik}-1} - \eta_{k,0}\theta_i \}} \right), & \mathbf{si} \ y_{ik} = C_k \end{cases} \quad (2.23)$$

utilizando el polinomio de Taylor de segundo grado está dada por:

$$\log p(y_i | \theta_i, \boldsymbol{\eta}, \mathbf{y}) \left\{ \begin{array}{l} \approx \sum_{k=1}^K [a_0 + a_1 \theta_i + a_2 \eta_{k,0} + a_3 \eta_{k,y_{ik}} - a_4 \theta_i^2 - a_5 \eta_{k,0}^2 - a_6 \eta_{k,y_{ik}}^2 \\ - a_7 \theta_i \eta_{k,0} + a_8 \theta_i \eta_{k,y_{ik}} + a_9 \eta_{k,0} \eta_{k,y_{ik}}], \quad \text{si} \quad \mathbf{y}_{ik} = 1 \\ \\ \approx \sum_{k=1}^K [b_0 + b_1 \theta_i + b_2 \eta_{k,0} - b_3 \eta_{k,y_{ik-1}} + b_4 \eta_{k,y_{ik}} \\ - b_5 \theta_i^2 - b_6 \eta_{k,0}^2 - b_7 \eta_{k,y_{ik-1}}^2 - b_8 \eta_{k,y_{ik}}^2 \\ - b_9 \eta_{k,0} \theta_i + b_{10} \theta_i \eta_{k,y_{ik-1}} + b_{11} \theta_i \eta_{k,y_{ik}} + b_{12} \eta_{k,0} \eta_{k,y_{ik}} \\ + b_{13} \eta_{k,0} \eta_{k,y_{ik-1}} + b_{14} \eta_{k,y_{ik}} \eta_{k,y_{ik-1}}], \quad \text{si} \quad \mathbf{y}_{ik} \in \{2, \dots, C_k - 1\} \\ \\ \approx \sum_{k=1}^K [c_0 + c_1 \theta_i + c_2 \eta_{k,0} + c_3 \eta_{k,y_{ik}} - c_4 \theta_i^2 - c_5 \eta_{k,0}^2 - c_6 \eta_{k,y_{ik}}^2 \\ - c_7 \theta_i \eta_{k,0} + c_8 \theta_i \eta_{k,y_{ik}} + c_9 \eta_{k,0} \eta_{k,y_{ik}}], \quad \text{si} \quad \mathbf{y}_{ik} = C_k \end{array} \right. \quad (2.24)$$

De acuerdo al teorema de Taylor presentado en [1], la aproximación se realiza donde la función (2.23) es analítica. Para esta función se puede ver que existe una singularidad cuando $\eta_{k,y_{ik}} = \eta_{k,y_{ik-1}}$, por lo que la aproximación se limitará a una bola que tiene como centro el punto a y como radio la distancia desde a hasta el plano $\eta_{k,y_{ik}} = \eta_{k,y_{ik-1}}$.

Finalmente, el ELBO($\mathbf{m}, \mathbf{s}^2, \mu_\theta, \sigma_\theta^2$) se aproxima por la expresión siguiente:

$$\begin{aligned}
\text{ELBO} = & - \sum_{k=1}^K \left(\log(\sqrt{2\pi}(1 - \Phi(0))) - \frac{(s_{k,0}^2 + m_{k,0}^2)}{2} + \frac{1}{6}(C_k - 1) \right) \\
& + \sum_{i=1}^n \left(-\frac{\log(2\pi)}{2} - \frac{(\sigma_\theta^2 + \mu_\theta^2)}{2} \right) \\
& + \sum_{i=1}^n \sum_{k=1}^K \left([\mathbf{a}_0 + \mathbf{a}_1\mu_\theta + \mathbf{a}_2m_{k,0} + \mathbf{a}_3m_{k,y_{ik}} - \mathbf{a}_4(\mu_\theta^2 + \sigma_\theta^2) - \mathbf{a}_5(m_{k,0}^2 + s_{k,0}^2) \right. \\
& - \mathbf{a}_6(m_{k,y_{ik}}^2 + s_{k,y_{ik}}^2) - \mathbf{a}_7\mu_\theta m_{k,0} + \mathbf{a}_8\mu_\theta m_{k,y_{ik}} + \mathbf{a}_9m_{k,0}m_{k,y_{ik}}] \mathbb{1}_{\{y_{ik}=1\}} \\
& + [\mathbf{b}_0 + \mathbf{b}_1\mu_\theta + \mathbf{b}_2m_{k,0} - \mathbf{b}_3m_{k,y_{ik}-1} + \mathbf{b}_4m_{k,y_{ik}} - \mathbf{b}_5(\mu_\theta^2 + \sigma_\theta^2) - \mathbf{b}_6(m_{k,0}^2 + s_{k,0}^2) \\
& - \mathbf{b}_7(m_{k,y_{ik}-1}^2 + s_{k,y_{ik}-1}^2) - \mathbf{b}_8(m_{k,y_{ik}}^2 + s_{k,y_{ik}}^2) - \mathbf{b}_9m_{k,0}\mu_\theta + \mathbf{b}_{10}\mu_\theta m_{k,y_{ik}-1} + \mathbf{b}_{11}\mu_\theta m_{k,y_{ik}} \\
& + \mathbf{b}_{12}m_{k,0}m_{k,y_{ik}} + \mathbf{b}_{13}m_{k,0}m_{k,y_{ik}-1} + \mathbf{b}_{14}m_{k,y_{ik}}m_{k,y_{ik}-1}] \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} \\
& + [\mathbf{c}_0 + \mathbf{c}_1\mu_\theta + \mathbf{c}_2m_{k,0} - \mathbf{c}_3m_{k,y_{ik}-1} - \mathbf{c}_4(\mu_\theta^2 + \sigma_\theta^2) - \mathbf{c}_5(m_{k,0}^2 + s_{k,0}^2) \\
& - \mathbf{c}_6(m_{k,y_{ik}-1}^2 + s_{k,y_{ik}-1}^2) + \mathbf{c}_7\mu_\theta m_{k,0} + \mathbf{c}_8\mu_\theta m_{k,y_{ik}-1} + \mathbf{c}_9m_{k,0}m_{k,y_{ik}-1}] \mathbb{1}_{\{y_{ik}=C_k\}} \left. \right) \\
& - \sum_{k=1}^K \sum_{j=0}^{C_k-1} \left[-\frac{\log(2\pi s_{k,j}^2)}{2} - \frac{1}{2} \right] - \sum_{i=1}^n \left[-\frac{\log(2\pi\sigma_\theta^2)}{2} - \frac{1}{2} \right]
\end{aligned}$$

Ahora, para encontrar las actualizaciones para los parámetros variacionales se recurre al cálculo variacional [9], de donde se obtiene que los parámetros solución deben satisfacer las ecuaciones de Euler-Lagrange. Para el ELBO en particular, si y_i es un parámetro variacional, debe satisfacer:

$$\frac{\partial}{\partial y_i} \text{ELBO} = 0$$

Así, calculamos las derivadas parciales del ELBO respecto a cada uno de los parámetros.

Parámetro μ_θ :

$$\begin{aligned}
\frac{\partial}{\partial \mu_\theta} \text{ELBO} = & -n\mu_\theta + \sum_{i=1}^n \sum_{k=1}^K \left([\mathbf{a}_1 - 2\mathbf{a}_4\mu_\theta - \mathbf{a}_7m_{k,0} + \mathbf{a}_8m_{k,1}] \mathbb{1}_{\{y_{ik}=1\}} \right. \\
& + [\mathbf{b}_1 - 2\mathbf{b}_5\mu_\theta - \mathbf{b}_9m_{k,0} + \mathbf{b}_{10}m_{k,y_{i,k}-1} + \mathbf{b}_{11}m_{k,y_{i,k}}] \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} \\
& \left. + [\mathbf{c}_1 - 2\mathbf{c}_4\mu_\theta + \mathbf{c}_7m_{k,0} + \mathbf{c}_8m_{k,C_k-1}] \mathbb{1}_{\{y_{ik}=C_k\}} \right) = 0
\end{aligned}$$

luego,

$$\begin{aligned} \mu_\theta = & \frac{\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K ([a_1 - \mathbf{a}_7 m_{k,0} + \mathbf{a}_8 m_{k,1}] \mathbb{1}_{\{y_{ik}=1\}})}{1 + \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^K (\mathbf{a}_4 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_5 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_4 \mathbb{1}_{\{y_{ik}=C_k\}})} \\ & + \frac{\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K ([b_1 - \mathbf{b}_9 m_{k,0} + \mathbf{b}_{10} m_{k,y_{ik}-1} + \mathbf{b}_{11} m_{k,y_{ik}}] \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k-1\}})}{1 + \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^K (\mathbf{a}_4 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_5 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_4 \mathbb{1}_{\{y_{ik}=C_k\}})} \\ & + \frac{\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K [\mathbf{c}_1 + \mathbf{c}_7 m_{k,0} + \mathbf{c}_8 m_{k,C_k-1}] \mathbb{1}_{\{y_{ik}=C_k\}}}{1 + \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^K (\mathbf{a}_4 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_5 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_4 \mathbb{1}_{\{y_{ik}=C_k\}})} \end{aligned}$$

Parámetro σ_θ^2 :

$$\frac{\partial}{\partial \sigma_\theta^2} \text{ELBO} = -\frac{n}{2} + \sum_{i=1}^n \sum_{k=1}^K \left(-\mathbf{a}_4 \mathbb{1}_{\{y_{ik}=1\}} - \mathbf{b}_5 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} - \mathbf{c}_4 \mathbb{1}_{\{y_{ik}=C_k\}} \right) + \frac{n}{2\sigma_\theta^2} = 0$$

de donde,

$$\sigma_\theta^2 = \frac{1}{1 + \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^K (\mathbf{a}_4 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_5 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_4 \mathbb{1}_{\{y_{ik}=C_k\}})}$$

Parámetro $m_{k,0}$:

$$\begin{aligned} \frac{\partial}{\partial m_{k,0}} \text{ELBO} = & -m_{k,0} + \sum_{i=1}^n \left([\mathbf{a}_2 - 2\mathbf{a}_5 m_{k,0} - \mathbf{a}_7 \mu_\theta + \mathbf{a}_9 m_{k,1}] \mathbb{1}_{\{y_{ik}=1\}} \right. \\ & + [\mathbf{b}_2 - 2\mathbf{b}_6 m_{k,0} - \mathbf{b}_9 \mu_\theta + \mathbf{b}_{12} m_{k,y_{i,k}} + \mathbf{b}_{13} m_{k,y_{ik}-1}] \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} \\ & \left. + [\mathbf{c}_2 - 2\mathbf{c}_5 m_{k,0} + \mathbf{c}_7 \mu_\theta + \mathbf{c}_9 m_{k,C_k-1}] \mathbb{1}_{\{y_{ik}=C_k\}} \right) = 0 \end{aligned}$$

luego,

$$\begin{aligned}
m_{k,0} = & \frac{\sum_{i=1}^n ([a_2 - a_7\mu_\theta + a_9m_{k,1}] \mathbb{1}_{\{y_{ik}=1\}})}{1 + 2 \sum_{i=1}^n (a_5 \mathbb{1}_{\{y_{ik}=1\}} + b_6 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + c_5 \mathbb{1}_{\{y_{ik}=C_k\}})} \\
& + \frac{\sum_{i=1}^n ([b_2 - b_9\mu_\theta + b_{12}m_{k,y_{ik}} + b_{13}m_{k,y_{ik}-1}] \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}})}{1 + 2 \sum_{i=1}^n (a_5 \mathbb{1}_{\{y_{ik}=1\}} + b_6 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + c_5 \mathbb{1}_{\{y_{ik}=C_k\}})} \\
& + \frac{\sum_{i=1}^n [c_2 + c_7\mu_\theta + c_9m_{k,C_k-1}] \mathbb{1}_{\{y_{ik}=C_k\}}}{1 + 2 \sum_{i=1}^n (a_5 \mathbb{1}_{\{y_{ik}=1\}} + b_6 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + c_5 \mathbb{1}_{\{y_{ik}=C_k\}})}
\end{aligned}$$

Parámetro $s_{k,0}^2$:

$$\frac{\partial}{\partial s_{k,0}^2} \text{ELBO} = -\frac{1}{2} + \sum_{i=1}^n \left(-a_5 \mathbb{1}_{\{y_{ik}=1\}} - b_6 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} - c_5 \mathbb{1}_{\{y_{ik}=C_k\}} \right) + \frac{1}{2s_{k,0}^2} = 0$$

de donde,

$$s_{k,0}^2 = \frac{1}{1 + 2 \sum_{i=1}^n (a_5 \mathbb{1}_{\{y_{ik}=1\}} + b_6 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + c_5 \mathbb{1}_{\{y_{ik}=C_k\}})}$$

Parámetro $m_{k,1}$:

$$\begin{aligned}
\frac{\partial}{\partial m_{k,1}} \text{ELBO} = & \sum_{i=1}^n \left([a_3 - 2a_6m_{k,1} + a_8\mu_\theta + a_9m_{k,0}] \mathbb{1}_{\{y_{ik}=1\}} \right. \\
& \left. + [-b_3 - 2b_7m_{k,1} + b_{10}\mu_\theta + b_{13}m_{k,0} + b_{14}m_{k,2}] \mathbb{1}_{\{y_{ik}=2\}} \right) = 0
\end{aligned}$$

luego,

$$m_{k,1} = \frac{\sum_{i=1}^n \left([a_3 + a_8\mu_\theta + a_9m_{k,0}] \mathbb{1}_{\{y_{ik}=1\}} + [-b_3 + b_{10}\mu_\theta + b_{13}m_{k,0} + b_{14}m_{k,2}] \mathbb{1}_{\{y_{ik}=2\}} \right)}{2 \sum_{i=1}^n (a_6 \mathbb{1}_{\{y_{ik}=1\}} + b_7 \mathbb{1}_{\{y_{ik}=2\}})}$$

Parámetro $s_{k,1}^2$:

$$\frac{\partial}{\partial s_{k,1}^2} \text{ELBO} = \sum_{i=1}^n \left(-\mathbf{a}_6 \mathbb{1}_{\{y_{ik}=1\}} - \mathbf{b}_7 \mathbb{1}_{\{y_{ik}=2\}} \right) + \frac{1}{2s_{k,1}^2} = 0$$

de donde,

$$s_{k,1}^2 = \frac{1}{2 \sum_{i=1}^n (\mathbf{a}_6 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_7 \mathbb{1}_{\{y_{ik}=2\}})}$$

Parámetro $m_{k,j}$: para $j \in 2, \dots, C_k - 2$

$$\begin{aligned} \frac{\partial}{\partial m_{k,j}} \text{ELBO} &= \sum_{i=1}^n \left([b_4 - 2b_8 m_{k,j} + b_{11} \mu_\theta + b_{12} m_{k,0} + b_{14} m_{k,y_{ik}-1}] \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k - 2\}} \right) \\ &= 0 \end{aligned}$$

luego,

$$m_{k,j} = \frac{\sum_{i=1}^n \left([b_4 + b_{11} \mu_\theta + b_{12} m_{k,0} + b_{14} m_{k,y_{ik}-1}] \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k - 2\}} \right)}{2 \sum_{i=1}^n (\mathbf{b}_8 \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k - 2\}})}$$

Parámetro $s_{k,j}^2$: para $j \in 2, \dots, C_k - 2$

$$\frac{\partial}{\partial s_{k,j}^2} \text{ELBO} = \sum_{i=1}^n \left(-\mathbf{b}_8 \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k - 2\}} \right) + \frac{1}{2s_{k,j}^2} = 0$$

de donde,

$$s_{k,j}^2 = \frac{1}{2 \sum_{i=1}^n (\mathbf{b}_8 \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k - 2\}})}$$

Parámetro m_{k,C_k-1} :

$$\begin{aligned} \frac{\partial}{\partial m_{k,C_k-1}} \text{ELBO} &= \sum_{i=1}^n \left([b_4 - 2b_8 m_{k,C_k-1} + b_{11} \mu_\theta + b_{12} m_{k,0} + b_{14} m_{k,C_k-2}] \mathbb{1}_{\{y_{ik}=C_k-1\}} \right. \\ &\quad \left. + [-c_3 - 2c_6 m_{k,C_k-1} + c_8 \mu_\theta + c_9 m_{k,0}] \mathbb{1}_{\{y_{ik}=C_k\}} \right) \\ &= 0 \end{aligned}$$

luego,

$$m_{k,C_k-1} = \frac{\sum_{i=1}^n \left([b_4 + b_{11}\mu_\theta + b_{12}m_{k,0} + b_{14}m_{k,C_k-2}] \mathbf{1}_{\{y_{ik}=C_k-1\}} \right)}{2 \sum_{i=1}^n \left(b_8 \mathbf{1}_{\{y_{ik}=C_k-1\}} + c_6 \mathbf{1}_{\{y_{ik}=C_k\}} \right)} + \frac{\sum_{i=1}^n \left([-c_3 + c_8\mu_\theta + c_9m_{k,0}] \mathbf{1}_{\{y_{ik}=C_k\}} \right)}{2 \sum_{i=1}^n \left(b_8 \mathbf{1}_{\{y_{ik}=C_k-1\}} + c_6 \mathbf{1}_{\{y_{ik}=C_k\}} \right)}$$

Parámetro s_{k,C_k-1}^2 :

$$\frac{\partial}{\partial s_{k,C_k-1}^2} \text{ELBO} = \sum_{i=1}^n \left(-b_8 \mathbf{1}_{\{y_{ik}=C_k-1\}} - c_6 \mathbf{1}_{\{y_{ik}=C_k\}} \right) + \frac{1}{2s_{k,C_k-1}^2} = 0$$

de donde,

$$s_{k,C_k-1}^2 = \frac{1}{2 \sum_{i=1}^n \left(b_8 \mathbf{1}_{\{y_{ik}=C_k-1\}} + c_6 \mathbf{1}_{\{y_{ik}=C_k\}} \right)}$$

En resumen, las varianzas σ_θ y $s_{k,j}$ para $k = 1, \dots, K$ y $j \in \{0, \dots, C_k - 1\}$, son:

$$\left\{ \begin{array}{l} \sigma_\theta^2 = \frac{1}{1 + \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^K \left(a_4 \mathbf{1}_{\{y_{ik}=1\}} + b_5 \mathbf{1}_{\{y_{ik} \in \{2, \dots, C_k-1\}\}} + c_4 \mathbf{1}_{\{y_{ik}=C_k\}} \right)} \\ s_{k,0}^2 = \frac{1}{1 + 2 \sum_{i=1}^n \left(a_5 \mathbf{1}_{\{y_{ik}=1\}} + b_6 \mathbf{1}_{\{y_{ik} \in \{2, \dots, C_k-1\}\}} + c_5 \mathbf{1}_{\{y_{ik}=C_k\}} \right)} \\ s_{k,1}^2 = \frac{1}{2 \sum_{i=1}^n \left(a_6 \mathbf{1}_{\{y_{ik}=1\}} + b_7 \mathbf{1}_{\{y_{ik}=2\}} \right)} \\ s_{k,j}^2 = \frac{1}{2 \sum_{i=1}^n \left(b_8 \mathbf{1}_{\{y_{ik} \in \{2, \dots, C_k-2\}\}} \right)} \quad \text{si } j \in \{2, \dots, C_k - 2\} \\ s_{k,C_k-1}^2 = \frac{1}{2 \sum_{i=1}^n \left(b_8 \mathbf{1}_{\{y_{ik}=C_k-1\}} + c_6 \mathbf{1}_{\{y_{ik}=C_k\}} \right)} \end{array} \right. \quad (2.25)$$

Y las actualizaciones para las medias μ_θ y $m_{k,j}$ para $k = 1, \dots, K$, y

$j \in \{0, \dots, C_k - 1\}$, son:

$$\left\{ \begin{aligned}
\mu_\theta &= \frac{\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K ([a_1 - \mathbf{a}_7 m_{k,0} + \mathbf{a}_8 m_{k,1}] \mathbb{1}_{\{y_{ik}=1\}})}{1 + \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^K (\mathbf{a}_4 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_5 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_4 \mathbb{1}_{\{y_{ik}=C_k\}})} \\
&+ \frac{\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K ([b_1 - \mathbf{b}_9 m_{k,0} + \mathbf{b}_{10} m_{k,y_{ik}-1} + \mathbf{b}_{11} m_{k,y_{ik}}] \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k-1\}})}{1 + \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^K (\mathbf{a}_4 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_5 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_4 \mathbb{1}_{\{y_{ik}=C_k\}})} \\
&+ \frac{\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K [c_1 + \mathbf{c}_7 m_{k,0} + \mathbf{c}_8 m_{k,C_k-1}] \mathbb{1}_{\{y_{ik}=C_k\}}}{1 + \frac{2}{n} \sum_{i=1}^n \sum_{k=1}^K (\mathbf{a}_4 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_5 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_4 \mathbb{1}_{\{y_{ik}=C_k\}})} \\
m_{k,0} &= \frac{\sum_{i=1}^n ([a_2 - a_7 \mu_\theta + a_9 m_{k,1}] \mathbb{1}_{\{y_{ik}=1\}})}{1 + 2 \sum_{i=1}^n (\mathbf{a}_5 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_6 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_5 \mathbb{1}_{\{y_{ik}=C_k\}})} \\
&+ \frac{\sum_{i=1}^n [b_2 - b_9 \mu_\theta + b_{12} m_{k,y_{ik}} + b_{13} m_{k,y_{ik}-1}] \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k-1\}}}{1 + 2 \sum_{i=1}^n (\mathbf{a}_5 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_6 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_5 \mathbb{1}_{\{y_{ik}=C_k\}})} \\
&+ \frac{\sum_{i=1}^n [c_2 + c_7 \mu_\theta + c_9 m_{k,C_k-1}] \mathbb{1}_{\{y_{ik}=C_k\}}}{1 + 2 \sum_{i=1}^n (\mathbf{a}_5 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_6 \mathbb{1}_{\{y_{ik} \in [2, \dots, C_k-1]\}} + \mathbf{c}_5 \mathbb{1}_{\{y_{ik}=C_k\}})}, \\
m_{k,1} &= \frac{\sum_{i=1}^n ([a_3 + a_8 \mu_\theta + a_9 m_{k,0}] \mathbb{1}_{\{y_{ik}=1\}} + [-b_3 + b_{10} \mu_\theta + b_{13} m_{k,0} + b_{14} m_{k,2}] \mathbb{1}_{\{y_{ik}=2\}})}{2 \sum_{i=1}^n (\mathbf{a}_6 \mathbb{1}_{\{y_{ik}=1\}} + \mathbf{b}_7 \mathbb{1}_{\{y_{ik}=2\}})} \\
m_{k,j} &= \frac{\sum_{i=1}^n ([b_4 + b_{11} \mu_\theta + b_{12} m_{k,0} + b_{14} m_{k,y_{ik}-1}] \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k-2\}})}{2 \sum_{i=1}^n (\mathbf{b}_8 \mathbb{1}_{\{y_{ik} \in 2, \dots, C_k-2\}})}, j \in \{2, \dots, C_k - 2\} \\
m_{k,C_k-1} &= \frac{\sum_{i=1}^n ([b_4 + b_{11} \mu_\theta + b_{12} m_{k,0} + b_{14} m_{k,C_k-2}] \mathbb{1}_{\{y_{ik}=C_k-1\}})}{2 \sum_{i=1}^n (\mathbf{b}_8 \mathbb{1}_{\{y_{ik}=C_k-1\}} + \mathbf{c}_6 \mathbb{1}_{\{y_{ik}=C_k\}})} \\
&+ \frac{\sum_{i=1}^n ([-c_3 + c_8 \mu_\theta + c_9 m_{k,0}] \mathbb{1}_{\{y_{ik}=C_k\}})}{2 \sum_{i=1}^n (\mathbf{b}_8 \mathbb{1}_{\{y_{ik}=C_k-1\}} + \mathbf{c}_6 \mathbb{1}_{\{y_{ik}=C_k\}})}
\end{aligned} \right. \tag{2.26}$$

Algorithm 2: CAVI for a mixture of Gaussians

Entrada: La matriz de Datos $Y_{n \times K}$, con n individuos y K ítems;
Salida: Densidades normales $q(\eta_{k,j}; m_{k,j}, s_{k,j}^2)$ y $q(\theta; \mu_\theta, \sigma_\theta^2)$;
Inicialización: Cálculo de las varianzas σ_θ^2 y $s_{k,j}^2$ de acuerdo a (2.25) y valores iniciales para los parámetros variacionales μ_θ^2 y $m_{k,j}$;

```
while el ELBO no converge do  
  for  $i \in \{1, \dots, N\}$  do  
     $\mu_\theta \propto \exp \{E_{-\theta} [\log p(z_j | z_{-j}, \mathbf{y})]\}$   
  end  
  for  $k \in \{1, \dots, K\}$  do  
    for  $j \in \{0, \dots, C_k - 1\}$  do  
      Actualizar  $m_{k,j}$  de acuerdo a (2.26)  
    end  
  end  
  Calcular ELBO  
end  
return  $q(\eta_{k,j}; m_{k,j}, s_{k,j}^2)$  y  $q(\theta; \mu_\theta, \sigma_\theta^2)$ 
```

El código utilizado para la implementación se encuentra en Anexos.

2.3. Bondad de Ajuste

Para evaluar el ajuste del modelo, en [12] se sugiere comparar la proporción observada de respuestas en cada categoría y los valores que proporciona el modelo en cada uno de los ítems. En el siguiente capítulo se encuentra una tabla con estos valores.

Capítulo 3

Resultados, conclusiones y recomendaciones

3.1. Resultados

Con la finalidad de medir el interés político de los franceses se consideran las preguntas realizadas sobre el ámbito político en la Encuesta Social Europea, ésta se puede encontrar en la página web de la ESS [6]. La pregunta es la siguiente:

- 1 ¿How interested would you say you are in politics?
(Very, quite, hardly, not at all interested)
- 2 ¿How able do you think you are to take an active role in a group involved with political issues?
(Not at all, a little, quite, very, completely able)
- 3 How confident are you in your own ability to participate in politics?
(Not at all, a little, quite, very, completely confident)

Se trabajó con las respuestas que proporcionaron 1938 individuos en la ESS [7], a los 3 ítems descritos anteriormente. Para el ítem $k = 1$ se tienen $C_1 = 4$ categorías, para el ítem $k = 2$ se tienen $C_2 = 5$ categorías, y para el ítem $k = 3$ se tienen $C_3 = 5$ categorías. El algoritmo empleado es el *CAVI for a Gaussian mixture model*, y se lo puede encontrar en [4]. La implementación se realizó haciendo uso del lenguaje de programación Python. Como resultado se obtienen la estimaciones de los parámetros

del modelo (1.4) y la distribución de la variable latente θ . Los resultados se presentan a continuación.

En el cuadro 3.1 se pueden encontrar las estimaciones de los parámetros *step difficulty*.

		Discri- minación	Umbral ítem 1 y 2	Umbral ítem 2 y 3	Umbral ítem 3 y 4	Umbral ítem 4 y 5
Ítem k	Descripción	a_k	$\kappa_{k,1}$	$\kappa_{k,2}$	$\kappa_{k,3}$	$\kappa_{1,4}$
1	interés	1.12	-0.6	-0.19	1.23	
2	capacidad	0.51	-0.67	-0.33	0.06	1.12
3	confianza	0.89	-0.66	-0.30	1.29	1.30

Cuadro 3.1: Estimaciones parámetros GRM

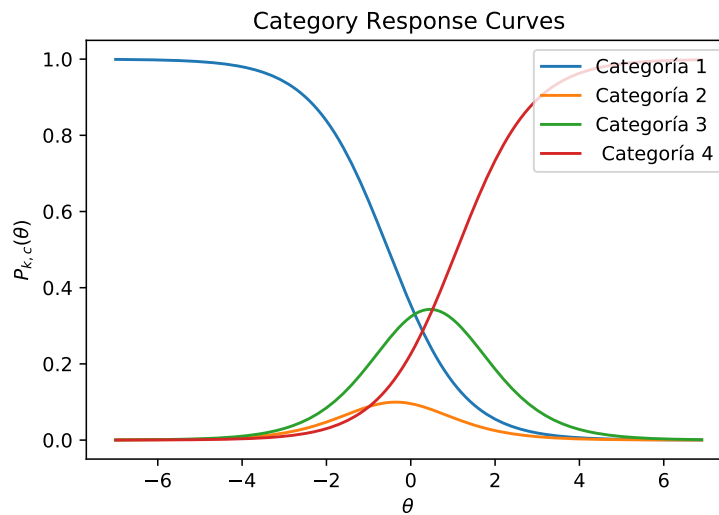


Figura 3.1: Category Response Curves Ítem 1

En los gráficos 3.1, 3.2 y 3.3 se muestran las curvas notadas como $P_{k,c}$ para los ítems $k = 1, 2, 3$ y las categorías $c = 1, 2, 3, 4$ o 5, denominadas *Category Response Curves* (CRC), [8]. Estas curvas representan la probabilidad de que un individuo con habilidad θ responda en la categoría c .

Ahora, los parámetros *step difficulty* conocidos también como parámetros de umbral para cada categoría [12], se pueden interpretar de la

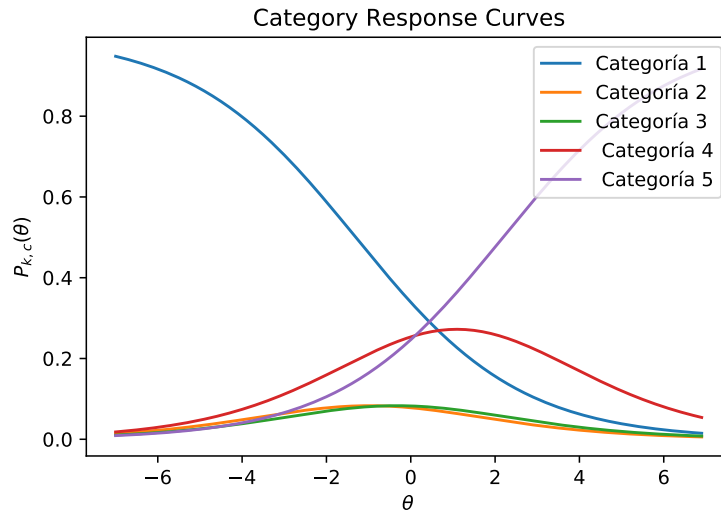


Figura 3.2: Category Response Curves Ítem 2

siguiente forma. Para el ítem 1, $\kappa_{1,2}$ es el parámetro de umbral entre las categorías 2 y 3, y es el punto en la escala de habilidad (θ) para el que un individuo que contesta el test tiene un 50% de posibilidad de responder por encima del umbral; es decir, en las categorías 3 y 4. Esto se puede observar en la figura 3.4(a) y en el cuadro 3.1. Ahora, los resultados de las estimaciones de la media y varianza de la distribución que sigue la variable latente, el interés político, determinan que ésta es de la forma:

$$\theta \sim N(\mu_{\theta} = -0,24, \sigma_{\theta}^2 = -0,24) \quad (3.1)$$

En los gráficos 3.4(a), 3.4(b) y 3.4(c) se puede observar la distribución de la variable latente junto con los valores de los parámetros $\kappa_{k,c}$ para el ítem k y la categoría c . Las rectas sobre la gráfica determinan la proporción de individuos que responden al test entre las categorías.

Para probar qué tan bien se ajusta el modelo se procede a comparar la proporción de individuos que responde al ítem en cada categoría y la proporción que sugiere el modelo. Los datos se presentan en el cuadro 3.1.

Se puede observar que la proporción estimada para la mayoría de las categorías no es adecuada, esto sugiere que la estimación de los pará-

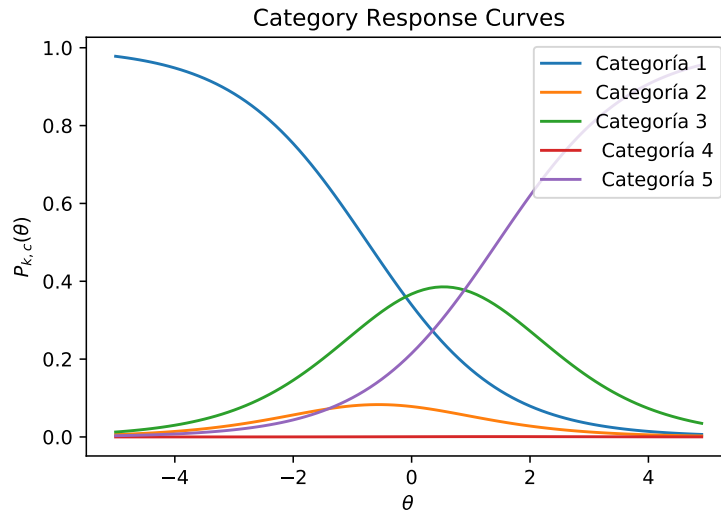
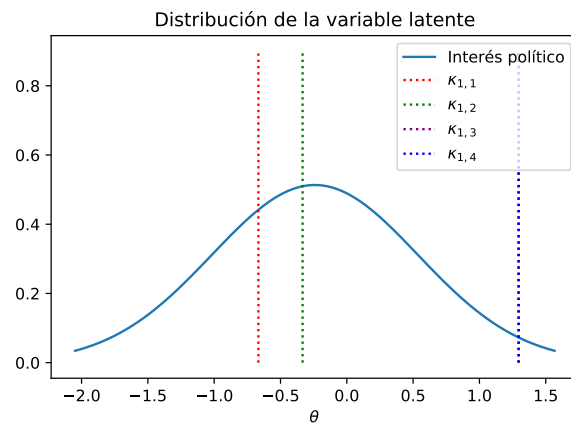


Figura 3.3: Category Response Curves Ítem 3

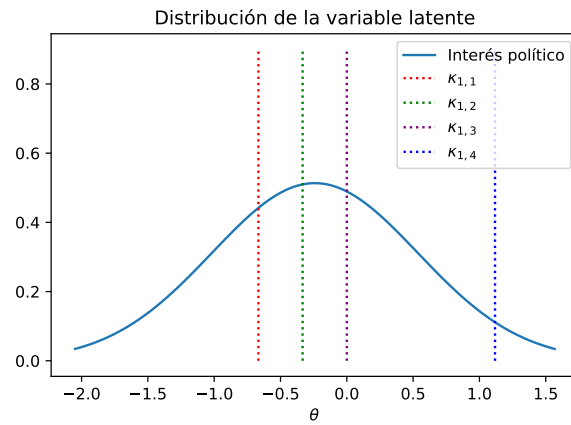
Proporción de individuos en la muestra <i>vs</i> proporción estimada						
	Ítem	categoria 1	categoria 2	categoria 3	categoria 4	categoria 5
muestra	1	0.1507	0.2451	0.4030	0.2012	
modelo	1	0.3418	0.1766	0.4345	0.0475	
muestra	2	0.3999	0.2946	0.2095	0.0521	0.0439
modelo	2	0.3145	0.1437	0.1494	0.3307	0.0617
muestra	3	0.2606	0.3891	0.2730	0.0490	0.0284
modelo	3	0.3140	0.1440	0.5009	0.0003	0.0406

Cuadro 3.2: Proporción de respuestas para cada categoría

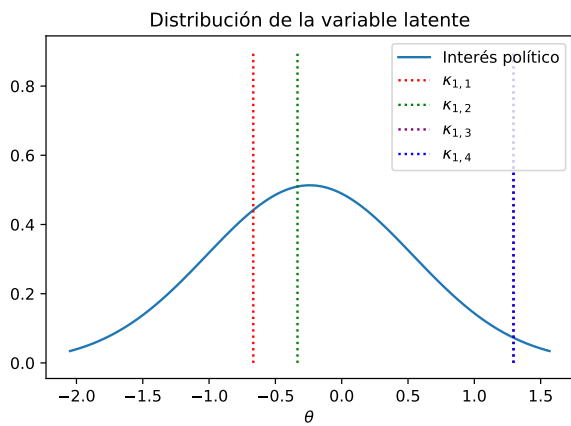
metros no es buena. Ahora, hay que tomar en cuenta a que esto puede deberse a que se utiliza una aproximación del ELBO, debido a la complejidad de la función (2.16) se realizó el desarrollo de Taylor de segundo orden para la misma. Es posible que existan mejores aproximaciones y que con esto a su vez se obtenga mejores resultados en las estimaciones de los parámetros expuestos anteriormente.



(a) Distribución de la variable latente y $\kappa_{k,1}$



(b) Distribución de la variable latente y $\kappa_{k,2}$



(c) Distribución de la variable latente y $\kappa_{k,3}$

Figura 3.4: Distribución de la variable latente

3.2. Conclusiones y recomendaciones

3.2.1. Conclusiones

En el trabajo se utilizan métodos de inferencia variacional para aproximar la distribución de una variable latente, el interés político en este caso, tomando en cuenta su eficiencia computacional en comparación con otros métodos. Se pudo observar que en cuestión de segundos y pocas iteraciones convergía el algoritmo y se obtuvieron los resultados de las estimaciones.

De entre una variedad de algoritmos de inferencia variacional como son: *Coordinate Ascent Variational Inference*, *Gradient Ascent Variational Inference*, *Stochastic Variational Inference* entre otros [3], se decidió trabajar con el primero, CAVI, que a diferencia de los demás algoritmos, no requiere el cálculo del gradiente de la función que se desea optimizar en cada iteración.

La familia de densidades sobre la cual se optimizó tiene las características de una *Mixture of Gaussians*; como su nombre lo indica, es una familia de distribuciones normales a través de la cual nos acercamos lo máximo posible y considerando variables independientes.

La obtención, el tratamiento de los datos y la implementación del algoritmo se realizaron con éxito haciendo uso del lenguaje de programación python a través de los cuadernos *Jupyter Notebooks*; sin embargo, los resultados no fueron satisfactorios, pues se tuvo que recurrir a una aproximación de la función objetivo debido a la complejidad de la misma lo que por supuesto afectó los resultados de las estimaciones de los parámetros del modelo. Otra razón por la que pudo verse afectada la estimación es el supuesto de que las variables latentes consideradas en este trabajo son independientes, [14].

3.2.2. Recomendaciones

Para futuros trabajos se recomienda, en primer lugar, considerar otras formas de aproximar el ELBO, para que de esta forma mejore la estima-

ción de los parámetros, o de ser posible, realizar el cálculo exacto del mismo; también se podría considerar otros algoritmos de inferencia variacional, los que se puede encontrar en [3] por ejemplo, y comparar los resultados. Finalmente, con el objetivo de mejorar las estimaciones de los parámetros se recomienda considerar más ítems que midan la variable latente y que ayuden a describir el comportamiento de la misma.

Referencias bibliográficas

- [1] Mark J Ablowitz, Athanassios S Fokas, and Athanassios S Fokas. *Complex variables: introduction and applications*. Cambridge University Press, 2003.
- [2] Frank B Baker and Seock-Ho Kim. *Item response theory: Parameter estimation techniques*. CRC press, 2004.
- [3] Matthew James Beal. *Variational algorithms for approximate Bayesian inference*. University of London, University College London (United Kingdom), 2003.
- [4] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112, 2017.
- [5] Rafael Jaime De Ayala. *The theory and practice of item response theory*. Guilford Publications, 2013.
- [6] ESS ERIC. European social survey european research infrastructure ess10 data documentation. sikt - norwegian agency for shared services in education and research, 2022.
- [7] ESS ERIC. European social survey european research infrastructure, ess20 - integrated file, edition 1.2 [data set]. sikt - norwegian agency for shared services in education and research, 2022.
- [8] Jean-Paul Fox. *Bayesian item response modeling: Theory and applications*. Springer, 2010.

- [9] Hansjörg Kielhöfer. Calculus of variations. *Texts in Applied Mathematics*, 67:85, 2018.
- [10] Iwin Leenen. Virtudes y limitaciones de la teoría de respuesta al ítem para la evaluación educativa en las ciencias médicas. *Investigación en educación médica*, 3(9):40–55, 2014.
- [11] Antonio Matas Terrón. *Introducción al análisis de la Teoría de Respuesta al Ítem*. Aidesoc.net, 2010.
- [12] Mariagiulia Matteucci and Luisa Stracqualursi. Student assessment via graded response model. *Statistica*, 66(4):435–447, 2006.
- [13] Bengt Muthén. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1):115–132, 1984.
- [14] Christian Andersson Naesseth. *Machine learning using approximate inference: Variational and sequential Monte Carlo methods*, volume 1969. Linköping University Electronic Press, 2018.
- [15] Sean Plummer, Debdeep Pati, and Anirban Bhattacharya. Dynamics of coordinate ascent variational inference: A case study in 2d ising models. *Entropy*, 22(11):1263, 2020.
- [16] W.J. Van der Linden and R.K Hambleton. *Handbook of Modern Item Response Theory*. Springer Science+Business Media New York, 1997.
- [17] M Wu, Davis R.L., B.W. Domingue, C. Piech, and N. Goodman. Variational item response theory: Fast, accurate, and expressive. *Department of Computer Science, Education, and Psychology. Stanford University*, 2020.
- [18] Eric P Xing, Michael I Jordan, and Stuart Russell. A generalized mean field algorithm for variational inference in exponential families. *arXiv preprint arXiv:1212.2512*, 2012.

Capítulo A

Título anexo

A.1. Óptimo de los parámetros Variacionales

A continuación se presenta el proceso mediante el cual se obtiene la expresión:

$$q_j^*(z_j) \propto \exp \{E_{-j} [\log p(z_j | z_{-j}, \mathbf{y})]\} \quad (\text{A.1})$$

donde $z = (z_j)_1^m$, $q_j^*(z_j)$ es el óptimo y los parámetros variacionales distintos a j son fijos. Para una mejor visualización se realizará el proceso para 3 variables latentes. Así, la distribución de la variable latente se escribe como sigue,

$$q(z) = \prod_{j=1}^3 q_j(z_j) = q_1(z_1)q_2(z_2)q_3(z_3) \quad (\text{A.2})$$

Se buscan q_1^* , q_2^* y q_3^* tales que,

$$q_1^*, q_2^*, q_3^* = \arg \min_{q_1, q_2, q_3} \text{ELBO}(q)$$

donde,

$$\text{ELBO}(q) = \int_{z_1} \int_{z_2} \int_{z_3} q_1(z_1)q_2(z_2)q_3(z_3) [\log p(z, \mathbf{y}) - \log q(z)] dz_1 dz_2 dz_3$$

por facilidad se nota: $q_1 = q_1(z_1)$, $q_2 = q_2(z_2)$, $q_3 = q_3(z_3)$, $p = p(z|\mathbf{y})$, y E_{-j} a la esperanza respecto a todas las variables distintas de j . Luego,

$$\begin{aligned} \text{ELBO}(q) &= \int_{z_1} \int_{z_2} \int_{z_3} q_1 q_2 q_3 [\log p - \log q_1 - \log q_2 - \log q_3] dz_1 dz_2 dz_3 \\ &= \int_{z_1} q_1 \left[\int_{z_2} \int_{z_3} q_2 q_3 [\log p] dz_2 dz_3 \right] dz_1 - \int_{z_1} q_1 \log q_1 dz_1 \\ &\quad - \int_{z_2} \int_{z_3} q_2 q_3 [\log q_1 + \log q_2] dz_2 dz_3 \\ &= \int_{z_1} q_1 E_{2,3}(\log p - \log q_1) dz_1 - E_{2,3}(\log q_2 + \log q_3) \end{aligned}$$

Si fijamos q_2 y q_3 obtenemos una función que depende únicamente de q_1

$$\text{ELBO}(q) = \text{ELBO}(q_1) = \int_{z_1} q_1 (E_{2,3}(\log(p)) - \log(q_1)) dz_1 + C$$

Consideramos

$$I(q_1) = \int_{z_1} l(q_1) dz_1, \text{ donde } l(q_1) = q_1 (E_{2,3}(\log(p)) - \log(q_1))$$

volviendo a la optimización

$$q_1^* = \arg \min_{q_1} \text{ELBO}(q_1)$$

y tenemos que las ecuaciones de Euler-Lagrange para este problema son las siguientes, [9]

$$\frac{\partial I}{\partial q_1} = 0, \text{ y } \frac{\partial l}{\partial q_1} = 0$$

de donde obtenemos que

$$\log(q_1) = E_{2,3}(\log(p)) - 1 \tag{A.3}$$

y de forma general tenemos

$$\log(q_j) = E_{-j}(\log(p(z|\mathbf{y}))) \tag{A.4}$$

A.2. Implementación CAVI en Python

```
1 import numpy as np
```

```

2 import math as ma
3 import pandas as pd
4 from scipy.stats import norm
5
6 DATOS=pd.read_csv('ESS10.csv',index_col='name')
7 DATOS=DATOS[(DATOS["cntry"]=="FR")]
8 DATOS.index=DATOS['idno']
9 DATOS=DATOS.loc[:,['polintr','actrolga','cptppola']]
10 DATOS=DATOS.drop(DATOS[DATOS["cptppola"]==7].index)
11 DATOS=DATOS.drop(DATOS[DATOS["cptppola"]==8].index)
12 DATOS=DATOS.drop(DATOS[DATOS["actrolga"]==8].index)
13 #DATOS=DATOS.iloc[range(15),:]
14 print(len(DATOS.index))
15
16 # Proporcion de individuos que responden al item 1
17 mu=0.10
18 sigma2=0.73
19 D=DATOS
20 D=D.drop(D[D["polintr"]==2].index)
21 D=D.drop(D[D["polintr"]==3].index)
22 D=D.drop(D[D["polintr"]==4].index)
23 print(D['polintr'].unique())
24 print('----proporcion individuos que responden en la categoria c=1 ----'
      )
25 p1=len(D.index)/len(DATOS.index)
26 print(p1," val -> ",norm.ppf(p1, loc=mu, scale=ma.sqrt(sigma2)))
27 D=DATOS
28 D=D.drop(D[D["polintr"]==1].index)
29 D=D.drop(D[D["polintr"]==3].index)
30 D=D.drop(D[D["polintr"]==4].index)
31 print(D['polintr'].unique())
32 print('----proporcion individuos que responden en la categoria c=2 ----'
      )
33 p2=len(D.index)/len(DATOS.index)
34 print(p2," val -> ",norm.ppf(p1+p2, loc=mu, scale=ma.sqrt(sigma2)))
35 D=DATOS
36 D=D.drop(D[D["polintr"]==1].index)
37 D=D.drop(D[D["polintr"]==2].index)
38 D=D.drop(D[D["polintr"]==4].index)
39 print(D['polintr'].unique())
40 print('----proporcion individuos que responden en la categoria c=3 ----'
      )
41 p3=len(D.index)/len(DATOS.index)

```

```

42 print(p3, ", val -> ", norm.ppf(p1+p2+p3, loc=mu, scale=ma.sqrt(sigma2)))
43 D=DATOS
44 D=D.drop(D[D["polintr"]==1].index)
45 D=D.drop(D[D["polintr"]==2].index)
46 D=D.drop(D[D["polintr"]==3].index)
47 print(D['polintr'].unique())
48 print('----proporcion individuos que responden en la categoria c=4 ----'
      )
49 p4=len(D.index)/len(DATOS.index)
50 print(p4)
51 print(p1+p2+p3+p4)
52
53 #Proporcion de individuos que responden al item 2
54 D=DATOS
55 D=D.drop(D[D["actrolga"]==2].index)
56 D=D.drop(D[D["actrolga"]==3].index)
57 D=D.drop(D[D["actrolga"]==4].index)
58 D=D.drop(D[D["actrolga"]==5].index)
59 print(D['actrolga'].unique())
60 print('----proporcion individuos que responden en la categoria c=1 ----'
      )
61 p1=len(D.index)/len(DATOS.index)
62 print(p1, ", val -> ", norm.ppf(p1, loc=mu, scale=ma.sqrt(sigma2)))
63 D=DATOS
64 D=D.drop(D[D["actrolga"]==1].index)
65 D=D.drop(D[D["actrolga"]==3].index)
66 D=D.drop(D[D["actrolga"]==4].index)
67 D=D.drop(D[D["actrolga"]==5].index)
68 print(D['actrolga'].unique())
69 print('----proporcion individuos que responden en la categoria c=2 ----'
      )
70 p2=len(D.index)/len(DATOS.index)
71 print(p2, ", val -> ", norm.ppf(p1+p2, loc=mu, scale=ma.sqrt(sigma2)))
72 D=DATOS
73 D=D.drop(D[D["actrolga"]==1].index)
74 D=D.drop(D[D["actrolga"]==2].index)
75 D=D.drop(D[D["actrolga"]==4].index)
76 D=D.drop(D[D["actrolga"]==5].index)
77 print(D['actrolga'].unique())
78 print('----proporcion individuos que responden en la categoria c=3 ----'
      )
79 p3=len(D.index)/len(DATOS.index)
80 print(p3, ", val -> ", norm.ppf(p1+p2+p3, loc=mu, scale=ma.sqrt(sigma2)))

```

```

81 D=DATOS
82 D=D.drop(D[D["actrolga"]==1].index)
83 D=D.drop(D[D["actrolga"]==2].index)
84 D=D.drop(D[D["actrolga"]==3].index)
85 D=D.drop(D[D["actrolga"]==5].index)
86 print(D['actrolga'].unique())
87 print('----proporcion individuos que responden en la categoria c=4 ----'
      )
88 p4=len(D.index)/len(DATOS.index)
89 print(p4," , val -> ",norm.ppf(p1+p2+p3+p4, loc=mu, scale=ma.sqrt(sigma2
      )))
90 D=DATOS
91 D=D.drop(D[D["actrolga"]==1].index)
92 D=D.drop(D[D["actrolga"]==2].index)
93 D=D.drop(D[D["actrolga"]==3].index)
94 D=D.drop(D[D["actrolga"]==4].index)
95 print(D['actrolga'].unique())
96 print('----proporcion individuos que responden en la categoria c=5 ----'
      )
97 p5=len(D.index)/len(DATOS.index)
98 print(p5)
99 print(p1+p2+p3+p4+p5)
100
101 #Proporcion de individuos que responden al item 3
102 D=DATOS
103 D=D.drop(D[D["cptppola"]==2].index)
104 D=D.drop(D[D["cptppola"]==3].index)
105 D=D.drop(D[D["cptppola"]==4].index)
106 D=D.drop(D[D["cptppola"]==5].index)
107 print(D['cptppola'].unique())
108 print('----proporcion individuos que responden en la categoria c=1 ----'
      )
109 p1=len(D.index)/len(DATOS.index)
110 print(p1," , val -> ",norm.ppf(p1, loc=mu, scale=ma.sqrt(sigma2)))
111 D=DATOS
112 D=D.drop(D[D["cptppola"]==1].index)
113 D=D.drop(D[D["cptppola"]==3].index)
114 D=D.drop(D[D["cptppola"]==4].index)
115 D=D.drop(D[D["cptppola"]==5].index)
116 print(D['cptppola'].unique())
117 print('----proporcion individuos que responden en la categoria c=2 ----'
      )
118 p2=len(D.index)/len(DATOS.index)

```

```

119 print(p2, ", val -> ", norm.ppf(p1+p2, loc=mu, scale=ma.sqrt(sigma2)))
120 D=DATOS
121 D=D.drop(D[D["cptppola"]==1].index)
122 D=D.drop(D[D["cptppola"]==2].index)
123 D=D.drop(D[D["cptppola"]==4].index)
124 D=D.drop(D[D["cptppola"]==5].index)
125 print(D['cptppola'].unique())
126 print('----proporcion individuos que responden en la categoria c=3 ----'
      )
127 p3=len(D.index)/len(DATOS.index)
128 print(p3, ", val -> ", norm.ppf(p1+p2+p3, loc=mu, scale=ma.sqrt(sigma2)))
129 D=DATOS
130 D=D.drop(D[D["cptppola"]==1].index)
131 D=D.drop(D[D["cptppola"]==2].index)
132 D=D.drop(D[D["cptppola"]==3].index)
133 D=D.drop(D[D["cptppola"]==5].index)
134 print(D['cptppola'].unique())
135 print('----proporcion individuos que responden en la categoria c=4 ----'
      )
136 p4=len(D.index)/len(DATOS.index)
137 print(p4, ", val -> ", norm.ppf(p1+p2+p3+p4, loc=mu, scale=ma.sqrt(sigma2
      )))
138 D=DATOS
139 D=D.drop(D[D["cptppola"]==1].index)
140 D=D.drop(D[D["cptppola"]==2].index)
141 D=D.drop(D[D["cptppola"]==3].index)
142 D=D.drop(D[D["cptppola"]==4].index)
143 print(D['cptppola'].unique())
144 print('----proporcion individuos que responden en la categoria c=5 ----'
      )
145 p5=len(D.index)/len(DATOS.index)
146 print(p5)
147 print(p1+p2+p3+p4+p5)
148
149 #CAVI
150 class ALG_CAVI(object):
151     '''Gaussian Model - CAVI'''
152     #La funcion __init__() asigna valores a las propiedades del objeto
153     def __init__(self, Y, K=3, Ck=[4,5,5]):
154         self.Y = Y
155         self.K = K
156         self.Ck = Ck

```

```

157         self.N = self.Y.shape[0] #numero de filas ( de individuos que
responden al test)
158
159     def _init(self): #se define la funcion init que inicializa los
parametros
160         #inicializamos los valores de mu \sim N(0,1)
161         np.random.seed(3)
162         self.mu = np.random.normal(0.0, 1)
163         # inicializamos los sum(Ck) valores de m_k
164         #np.random.seed(100) #26
165         self.m=[]
166         for k in range(self.K):
167             L=[]
168             L.append(np.random.uniform(low=0, high=1))
169             for j in range(1,self.Ck[k]):
170                 L.append(-1+j*(2/(self.Ck[k]+1)))
171             self.m.append(L)
172         print('Valores iniciales de m_kj: ',self.m)
173         #coeficientes
174         #a=[ma.log(0.5)-1/8,1/4,1/4,1/4,1/8,1/8,1/8,3/4,1/4,1/4]
175         #b=[ma.log((ma.e-1)/(2*(ma.e+1)))-(ma.e**4+36*ma.e**3-26*ma.e
**2+4*ma.e+1)/(8*((ma.e+1)**2)*((ma.e-1)**2)),(ma.e**2+10*ma.e+1)
/(4*(ma.e+1)**2),(ma.e**2+10*ma.e+1)/(4*(ma.e+1)**2),(ma.e*(5*ma.e
**2-2*ma.e+1))/((ma.e**2-1)**2),(ma.e**2+6*ma.e-3)/(4*(ma.e-1)**2)
,(ma.e**2+6*ma.e+1)/(8*(ma.e+1)**2),(ma.e**2+6*ma.e+1)/(8*(ma.e+1)
**2),(ma.e*(ma.e**2+1))/((ma.e**2-1)**2),((ma.e+1)**2)/(8*(ma.e-1)
**2),(3*ma.e**2+6*ma.e-1)/(4*(ma.e+1)**2),(ma.e)/((ma.e+1)**2)
,1/4,1/4,(ma.e)/((ma.e+1)**2),(ma.e)/((ma.e-1)**2)]
176         #c=[ma.log((ma.e)/(ma.e+1))-(3*ma.e+1)/((ma.e+1)**2),(2*ma.e)
/((ma.e+1)**2),(2*ma.e)/((ma.e+1)**2),(3*ma.e+1)/((ma.e+1)**2),(ma.
e)/(2*(ma.e+1)**2),(ma.e)/(2*(ma.e+1)**2),(ma.e)/(2*(ma.e+1)**2)
,1/((ma.e+1)**2),(ma.e)/((ma.e+1)**2),(ma.e)/((ma.e+1)**2)]
177         a=[-0.71047,0.08515,0.04257,0.48103,0.02838,0.00709,0.11354,
178         0.67974,0.11354,0.05677]
179         b=[-2.89485,0.05498,0.02749,2.0755,1.97033,0.05856,0.01464,
180         0.57388,0.58104,0.11725,0.11354,0.1207,0.06035,0.05677,0.92067]
181         c=[-0.68341,-0.03017,-0.01508,0.53231,0.03017,0.00754,0.1207,
182         0.56248,0.1207,0.06035]
183         #DETERMINAMOS LOS VALORES DE LA VARIANZA Sigma^2
184         self.sigma2=0
185         s=0
186         for i in range(self.N):
187             for k in range(self.K):

```

```

188         if self.Y.iloc[i,k]==1:
189             s=s+a[4]
190         elif self.Y.iloc[i,k]==self.Ck[k]:
191             s=s+c[4]
192         else:
193             s=s+b[5]
194     self.sigma2=1/(1+(2/self.N)*s)
195     #DETERMINAMOS LOS VALORES DE LAS VARIANZAS Sk^2
196     self.s2=[]
197     for k in range(self.K):
198         L=[]
199         for j in range(self.Ck[k]):
200             if j==0:
201                 s=0
202                 for i in range(self.N):
203                     if self.Y.iloc[i,k]==1:
204                         s=s+a[5]
205                     elif self.Y.iloc[i,k]==self.Ck[k]:
206                         s=s+c[5]
207                     else:
208                         s=s+b[6]
209                 L.insert(j,1/(1+2*s))
210             elif j==1:
211                 s=0
212                 for i in range(self.N):
213                     if self.Y.iloc[i,k]==1:
214                         s=s+a[6]
215                     elif self.Y.iloc[i,k]==2:
216                         s=s+b[7]
217                 L.insert(j,1/(2*s))
218             elif j in range(2,self.Ck[k]-1):
219                 s=0
220                 for i in range(self.N):
221                     if self.Y.iloc[i,k] in range(2,self.Ck[k]-1):
222                         s=s+b[8]
223                 L.insert(j,1/(2*s))
224             else:
225                 s=0
226                 for i in range(self.N):
227                     if self.Y.iloc[i,k]==self.Ck[k]-1:
228                         s=s+b[8]
229                     elif self.Y.iloc[i,k]==self.Ck[k]:
230                         s=s+c[6]

```



```

231         L.insert(j,1/(2*s))
232     self.s2.append(L)
233     print('Valores iniciales de s_kj: ',self.s2)
234
235     #imprimimos los valores iniciales de las medias y varianzas
236     print('Medias')
237     print('---m---')
238     print(self.m)
239     print('---mu---')
240     print(self.mu)
241     print('s2 y sigma 2 calculados')
242
243
244     def get_elbo(self):
245         #a=[ma.log(0.5)-1/8,1/4,1/4,1/4,1/8,1/8,1/8,3/4,1/4,1/4]
246         #b=[ma.log((ma.e-1)/(2*(ma.e+1)))-(ma.e**4+36*ma.e**3-26*ma.e
**2+4*ma.e+1)/(8*((ma.e+1)**2)*((ma.e-1)**2)),(ma.e**2+10*ma.e+1)
/(4*(ma.e+1)**2),(ma.e**2+10*ma.e+1)/(4*(ma.e+1)**2),(ma.e*(5*ma.e
**2-2*ma.e+1))/((ma.e**2-1)**2),(ma.e**2+6*ma.e-3)/(4*(ma.e-1)**2)
,(ma.e**2+6*ma.e+1)/(8*(ma.e+1)**2),(ma.e**2+6*ma.e+1)/(8*(ma.e+1)
**2),(ma.e*(ma.e**2+1))/((ma.e**2-1)**2),((ma.e+1)**2)/(8*(ma.e-1)
**2),(3*ma.e**2+6*ma.e-1)/(4*(ma.e+1)**2),(ma.e)/((ma.e+1)**2)
,1/4,1/4,(ma.e)/((ma.e+1)**2),(ma.e)/((ma.e-1)**2)]
247         #c=[ma.log((ma.e)/(ma.e+1))-(3*ma.e+1)/((ma.e+1)**2),(2*ma.e)
/((ma.e+1)**2),(2*ma.e)/((ma.e+1)**2),(3*ma.e+1)/((ma.e+1)**2),(ma.
e)/(2*(ma.e+1)**2),(ma.e)/(2*(ma.e+1)**2),(ma.e)/(2*(ma.e+1)**2)
,1/((ma.e+1)**2),(ma.e)/((ma.e+1)**2),(ma.e)/((ma.e+1)**2)]
248         a=[-0.71047,0.08515,0.04257,0.48103,0.02838,0.00709,
249             0.11354,0.67974,0.11354,0.05677]
250         b=[-2.89485,0.05498,0.02749,2.0755,1.97033,0.05856,
251             0.01464,0.57388,0.58104,0.11725,0.11354,0.1207,0.06035,0.05677,0.92067]
252
253         c=[-0.68341,-0.03017,-0.01508,0.53231,0.03017,
254             0.00754,0.1207,0.56248,0.1207,0.06035]
255         #calculamos el termino 1
256         ter1=0
257         for k in range(self.K):
258             ter1=ter1-ma.log(ma.sqrt(2*ma.pi)*(0.5))-(self.s2[k][0]+
self.m[k][0]**2)/2+ma.log(1/2)*(self.Ck[k]-1)
259         #calculamos el termino 4
260         ter4=0
261         for k in range(self.K):

```

```

261         for j in range(self.Ck[k]):
262             ter4=ter4+(ma.log(2*ma.pi*self.s2[k][j]))/2 + 0.5
263         #calculamos el termino 2
264         ter2=0
265         for i in range(self.N):
266             ter2=ter2-(ma.log(2*ma.pi))/2 -(self.sigma2+self.mu**2)/2
267         #calculamos el termino 5
268         ter5=0
269         for i in range(self.N):
270             ter5=ter5+(ma.log(2*ma.pi*self.sigma2))/2 +0.5
271         #calculamos el termino 3
272         ter3=0
273         for i in range(self.N):
274             for k in range(self.K):
275                 if DATOS.iloc[i,k]==1:
276                     ter3=ter3+a[0]+(a[1]*self.mu)+(a[2]*self.m[k][0])+(
a[3]*self.m[k][1])
277                     ter3=ter3-(a[4]*(self.mu**2+self.sigma2))-(a[5]*(
self.m[k][0]**2+self.s2[k][0]))-(a[6]*(self.m[k][1]**2+self.s2[k
][1]))
278                     ter3=ter3-(a[7]*self.mu*self.m[k][0])+(a[8]*self.mu
*self.m[k][1])+(a[9]*self.m[k][0]*self.m[k][1])
279                 elif DATOS.iloc[i,k]==self.Ck[k]:
280                     ter3=ter3+c[0]+(c[1]*self.mu)+(c[2]*self.m[k][0])-(
c[3]*self.m[k][self.Ck[k]-1])
281                     ter3=ter3-(c[4]*(self.mu**2+self.sigma2))-(c[5]*(
self.m[k][0]**2+self.s2[k][0]))-(c[6]*(self.m[k][self.Ck[k]-1]**2+
self.s2[k][self.Ck[k]-1]))
282                     ter3=ter3+(c[7]*self.mu*self.m[k][0])+(c[8]*self.mu
*self.m[k][self.Ck[k]-1])+(c[9]*self.m[k][0]*self.m[k][self.Ck[k
]-1])
283                 else:
284                     ter3=ter3+b[0]+(b[1]*self.mu)+(b[2]*self.m[k][0])-(
b[3]*self.m[k][int(self.Y.iloc[i,k])-1])+(b[4]*self.m[k][int(self.Y
.iloc[i,k])])
285                     ter3=ter3-(b[5]*(self.mu**2+self.sigma2))-(b[6]*(
self.m[k][0]**2+self.s2[k][0]))-(b[7]*(self.m[k][int(self.Y.iloc[i,
k])-1]**2+self.s2[k][int(self.Y.iloc[i,k])-1]))-(b[8]*(self.m[k][
int(self.Y.iloc[i,k])]**2+self.s2[k][int(self.Y.iloc[i,k])]))
286                     ter3=ter3-(b[9]*self.mu*self.m[k][0])+(b[10]*self.
mu*self.m[k][int(self.Y.iloc[i,k])-1])+(b[11]*self.mu*self.m[k][int
(self.Y.iloc[i,k])])+(b[12]*self.m[k][0]*self.m[k][int(self.Y.iloc[
i,k])])+(b[13]*self.m[k][0]*self.m[k][int(self.Y.iloc[i,k])-1])+(b

```

```

[14]*self.m[k][int(self.Y.iloc[i,k])]*self.m[k][int(self.Y.iloc[i,k
287         return ter1 + ter2 + ter3 + ter4 + ter5
288
289     def fit(self, max_iter=100, tol=0.0000001):
290         self._init()
291         self.elbo_values = [self.get_elbo()]
292         self.m_history = [self.m]
293         self.mu_history = [self.mu]
294         for iter_ in range(1, max_iter+1):
295             self._cavi()
296             self.m_history.append(self.m)
297             self.mu_history.append(self.mu)
298             self.elbo_values.append(self.get_elbo())
299             print('ITER:',iter_, ' -> ELBO:',self.elbo_values )
300             if iter_ % 1 == 0:
301                 print('--- m ---')
302                 print(iter_, self.m_history[iter_])
303                 print('--- mu ---')
304                 print(iter_, self.mu_history[iter_])
305                 if np.abs(self.elbo_values[-2] - self.elbo_values[-1]) <=
tol:
306                     print('ELBO converge con  %.3f at iteration %d'%(self.
elbo_values[-1],
307
iter_))
308                     break
309
310             if iter_ == max_iter:
311                 print('ELBO finalizo con  %.3f'%(self.elbo_values[-1]))
312
313
314     def _cavi(self):
315         self._update_m()
316         self._update_mu()
317
318     def _update_m(self):
319         #a=[ma.log(0.5)-1/8,1/4,1/4,1/4,1/8,1/8,1/8,3/4,1/4,1/4]
320         #b=[ma.log((ma.e-1)/(2*(ma.e+1)))-(ma.e**4+36*ma.e**3-26*ma.e
**2+4*ma.e+1)/(8*((ma.e+1)**2)*((ma.e-1)**2)),(ma.e**2+10*ma.e+1)
/(4*(ma.e+1)**2),(ma.e**2+10*ma.e+1)/(4*(ma.e+1)**2),(ma.e*(5*ma.e
**2-2*ma.e+1))/((ma.e**2-1)**2),(ma.e**2+6*ma.e-3)/(4*(ma.e-1)**2)
,(ma.e**2+6*ma.e+1)/(8*(ma.e+1)**2),(ma.e**2+6*ma.e+1)/(8*(ma.e+1)

```

```

**2), (ma.e*(ma.e**2+1))/((ma.e**2-1)**2), ((ma.e+1)**2)/(8*(ma.e-1)
**2), (3*ma.e**2+6*ma.e-1)/(4*(ma.e+1)**2), (ma.e)/((ma.e+1)**2)
,1/4,1/4, (ma.e)/((ma.e+1)**2), (ma.e)/((ma.e-1)**2)]
321     #c=[ma.log((ma.e)/(ma.e+1))-(3*ma.e+1)/((ma.e+1)**2), (2*ma.e)
/((ma.e+1)**2), (2*ma.e)/((ma.e+1)**2), (3*ma.e+1)/((ma.e+1)**2), (ma.
e)/((ma.e+1)**2), (ma.e)/((ma.e+1)**2), (ma.e)/((ma.e+1)**2), 1/((ma.e
+1)**2), (ma.e)/((ma.e+1)**2), (ma.e)/((ma.e+1)**2)]
322     a=[-0.71047,0.08515,0.04257,0.48103,0.02838,0.00709,
323     0.11354,0.67974,0.11354,0.05677]
324     b=[-2.89485,0.05498,0.02749,2.0755,1.97033,0.05856,
325     0.01464,0.57388,0.58104,0.11725,0.11354,0.1207,
326     0.06035,0.05677,0.92067]
327     c=[-0.68341,-0.03017,-0.01508,0.53231,0.03017,
328     0.00754,0.1207,0.56248,0.1207,0.06035]
329     #aux_m=self.m
330     R=7/2
331     for k in range(self.K):
332         for j in range(self.Ck[k]):
333             if j==0:
334                 s=0
335                 for i in range(self.N):
336                     if self.Y.iloc[i,k]==1:
337                         s=s+a[2]-a[7]*self.mu+a[9]*self.m[k][1]
338                     elif self.Y.iloc[i,k]==self.Ck[k]:
339                         s=s+c[2]+c[7]*self.mu+c[9]*self.m[k][self.
Ck[k]-1]
340                     else:
341                         s=s+b[2]-b[9]*self.mu+b[12]*self.m[k][int(
self.Y.iloc[i,k])+b[13]*self.m[k][int(self.Y.iloc[i,k])-1]
342                     #aproximacion en una vecindad del punto al rededor
del cual se hizo el desarrollo de Taylor-----
343                     for l in range(1,self.Ck[k]+1):
344                         if l==1:
345                             if (s*self.s2[k][0]-1/2)**2+(self.mu-1/4)
**2+(self.m[k][1]-1/2)**2<R:
346                                 val=True
347                             else:
348                                 val=False
349                         elif l in range(2,self.Ck[k]):
350                             if (s*self.s2[k][0]-1/2)**2<R-(self.mu-1/4)
**2-(self.m[k][l-1]-1/2)**2-(self.m[k][1]-1/2)**2:
351                                 val=True
352                     else:

```

```

353         val=False
354     elif l ==self.Ck[k]:
355         if (s*self.s2[k][0]-1/2)**2+(self.mu-1/4)
**2+(self.m[k][l-1]-1/2)**2<R:
356             val=True
357         else:
358             val=False
359     #
-----

360     if val:
361         self.m[k][0]=s*self.s2[k][0]
362     elif j==1:
363         s=0
364         for i in range(self.N):
365             if self.Y.iloc[i,k]==1:
366                 s=s+a[3]+a[8]*self.mu+a[9]*self.m[k][0]
367             elif self.Y.iloc[i,k]==2:
368                 s=s-b[3]+b[10]*self.mu+b[13]*self.m[k][0]+b
[14]*self.m[k][2]
369         #aproximacion en una vecindad del punto al rededor
del cual se hizo el desarrollo de Taylor-----
370         if (s*self.s2[k][1]-1/2)**2<R-(self.mu-1/4)
**2-(self.m[k][2]-1/2)**2-(self.m[k][0]-1/2)**2:
371             val=True
372         else:
373             val=False
374     #
-----

375     #condicion de monotonia para avanzar
376     if (s*self.s2[k][1]<=self.m[k][2])&val:
377         self.m[k][1]=s*self.s2[k][1]
378     elif j in range(2,self.Ck[k]-1):
379         s=0
380         for i in range(self.N):
381             if self.Y.iloc[i,k] in range(2,self.Ck[k]-1):
382                 s=s+b[4]+b[11]*self.mu+b[12]*self.m[k][0]+b
[14]*self.m[k][int(self.Y.iloc[i,k])-1]
383         #aproximacion en una vecindad del punto al rededor
del cual se hizo el desarrollo de Taylor-----
384         if (s*self.s2[k][j]-1/2)**2<R-(self.mu-1/4)**2-(
self.m[k][j-1]-1/2)**2-(self.m[k][0]-1/2)**2:

```

```

385         val=True
386     else:
387         val=False
388     #
-----
389     #condicion de monotonia para avanzar
390     if (s*self.s2[k][j]<=self.m[k][j+1])&val:
391         self.m[k][j]=s*self.s2[k][j]
392     elif j==self.Ck[k]-1:
393         s=0
394         for i in range(self.N):
395             if self.Y.iloc[i,k]==self.Ck[k]-1:
396                 s=s+b[4]+b[11]*self.mu+b[12]*self.m[k][0]+b
[14]*self.m[k][int(self.Y.iloc[i,k])-1]
397             elif self.Y.iloc[i,k]==self.Ck[k]:
398                 s=s-c[3]+c[8]*self.mu+c[9]*self.m[k][0]
399     #aproximacion en una vecindad del punto al rededor
del cual se hizo el desarrollo de Taylor-----
400     if (s*self.s2[k][j]-1/2)**2<R-(self.mu-1/4)**2-(
self.m[k][j-1]-1/2)**2-(self.m[k][0]-1/2)**2:
401         val=True
402     else:
403         val=False
404     #
-----
405     #condicion de monotonia para avanzar
406     if (s*self.s2[k][j]>self.m[k][j-1])&val:
407         self.m[k][j]=s*self.s2[k][j]
408
409     def _update_mu(self):
410         #a=[ma.log(0.5)-1/8,1/4,1/4,1/4,1/8,1/8,1/8,3/4,1/4,1/4]
411         #b=[ma.log((ma.e-1)/(2*(ma.e+1)))-(ma.e**4+36*ma.e**3-26*ma.e
**2+4*ma.e+1)/(8*((ma.e+1)**2)*((ma.e-1)**2)),(ma.e**2+10*ma.e+1)
/(4*(ma.e+1)**2),(ma.e**2+10*ma.e+1)/(4*(ma.e+1)**2),(ma.e*(5*ma.e
**2-2*ma.e+1))/((ma.e**2-1)**2),(ma.e**2+6*ma.e-3)/(4*(ma.e-1)**2)
,(ma.e**2+6*ma.e+1)/(8*(ma.e+1)**2),(ma.e**2+6*ma.e+1)/(8*(ma.e+1)
**2),(ma.e*(ma.e**2+1))/((ma.e**2-1)**2),((ma.e+1)**2)/(8*(ma.e-1)
**2),(3*ma.e**2+6*ma.e-1)/(4*(ma.e+1)**2),(ma.e)/((ma.e+1)**2)
,1/4,1/4,(ma.e)/((ma.e+1)**2),(ma.e)/((ma.e-1)**2)]
412         #c=[ma.log((ma.e)/(ma.e+1))-(3*ma.e+1)/((ma.e+1)**2),(2*ma.e)
/((ma.e+1)**2),(2*ma.e)/((ma.e+1)**2),(3*ma.e+1)/((ma.e+1)**2),(ma.

```

```

e)/(2*(ma.e+1)**2), (ma.e)/(2*(ma.e+1)**2), (ma.e)/(2*(ma.e+1)**2)
,1/((ma.e+1)**2), (ma.e)/((ma.e+1)**2), (ma.e)/((ma.e+1)**2)]
413     a=[-0.71047,0.08515,0.04257,0.48103,0.02838,0.00709,
414         0.11354,0.67974,0.11354,0.05677]
415     b=[-2.89485,0.05498,0.02749,2.0755,1.97033,0.05856,
416         0.01464,0.57388,0.58104,0.11725,0.11354,0.1207,
417         0.06035,0.05677,0.92067]
418     c=[-0.68341,-0.03017,-0.01508,0.53231,0.03017,
419         0.00754,0.1207,0.56248,0.1207,0.06035]
420     s=0
421     for i in range(self.N):
422         for k in range(self.K):
423             if self.Y.iloc[i,k]==1:
424                 s=s+a[1]-a[7]*self.m[k][0]+a[8]*self.m[k][1]
425             elif self.Y.iloc[i,k]==self.Ck[k]:
426                 s=s+c[1]+c[7]*self.m[k][0]+c[8]*self.m[k][self.Ck[k
]-1]
427             else:
428                 s=s+b[1]-b[9]*self.m[k][0]+b[10]*self.m[k][int(self
.Y.iloc[i,k])-1]+b[11]*self.m[k][int(self.Y.iloc[i,k])]
429     self.mu=(1/self.N)*s*self.sigma2

```

Listing A.1: CAVI mezcla de Gaussianas