



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

**Construcción de un índice de desarrollo humano para
México utilizando el análisis bayesiano de
componentes principales**

Rosa Angela Pacheco Gil

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA
OBTENER EL GRADO DE:

MAESTRA EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO

2010

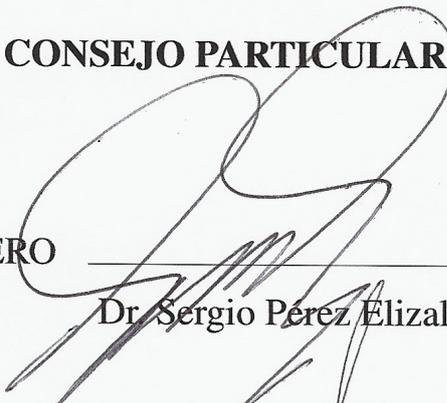
La presente tesis titulada: **Construcción de un índice de desarrollo humano para México utilizando el análisis bayesiano de componentes principales**, realizada por la alumna: **Rosa Angela Pacheco Gil.**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

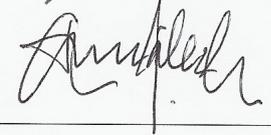
CONSEJO PARTICULAR

CONSEJERO



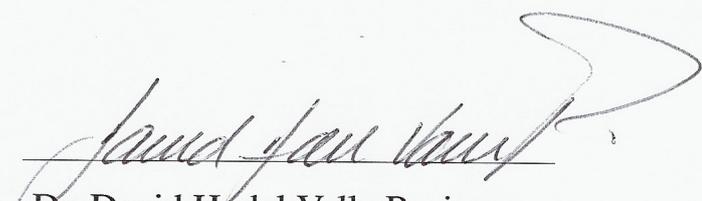
Dr. Sergio Pérez Elizalde

ASESOR



Dr. Félix V. González Cossio

ASESOR



Dr. David H. del Valle Paniagua

RESUMEN

Construcción de un índice de desarrollo humano para México utilizando el análisis bayesiano de componentes principales

Rosa Angela Pacheco Gil., M.C.

Colegio de Postgraduados, 2010

En numerosos problemas se desea explicar las fuentes de variación de los datos y un método utilizado para dicho fin es el análisis de componentes principales (ACP). Se denomina primera componente principal de una distribución n -dimensional, a la combinación lineal de componentes que posee máxima varianza; análogamente, la segunda componente principal es la combinación lineal que presenta mayor desviación, una vez descontada la parte de varianza atribuible a la primera componente. El objetivo de este procedimiento es explicar la variación muestral en términos de q combinaciones lineales de las variables originales. Uno de los problemas del método de componentes principales es la inferencia estadística sobre q . En este trabajo se aplicó un enfoque bayesiano para hacer inferencia sobre q , y los otros parámetros del modelo probabilístico de las componentes principales, al problema de la determinación de un índice de desarrollo humano (*idh*). Debido a que la dimensión del espacio paramétrico depende de q , para obtener una aproximación de la distribución a posteriori conjunta se utilizó el método de cadenas de Markov Monte Carlo con saltos reversibles (RJMCMC, por sus siglas en inglés). Como resultado se obtuvo un *idh* que permite clasificar objetivamente a los municipios de México de acuerdo con su nivel de bienestar.

Palabras clave: Desarrollo humano, componentes principales bayesianas, MCMC.

ABSTRACT

Construction of a human development index for México using the bayesian analysis of principal components

Rosa Angela Pacheco Gil., M.C.

Colegio de Postgraduados, 2010

In many problems you want to explain the sources of variation in the data and methodology used for this purpose is the principal component analysis (PCA). The first principal component of a distribution n -dimensional, is the linear combination of components has maximum variance, similarly, the second principal component is the linear combination that presents the greatest deviation, once discounted the variance attributable to the first component. The purpose of this procedure is to explain the sample variation in terms of q linear combinations of the original variables. One of the problems of the principal component method is the statistical inference about q . In this work we applied a bayesian approach to inference about q , and other parameters of the probabilistic model of principal components, to the problem of determining human development index (*idh*). Because the dimension of the parameter space depends on q , to approximate the joint posterior distribution method was used Reversible jump Markov chain Monte Carlo (RJMCMC). The result was an objectively *idh* for classifying the municipalities of Mexico according to their welfare level.

Keywords: Human development, principal components bayesian, MCMC.

*A mis padres: María Isabel y José Arturo,
porque me lo han dado todo.*

A mis hermanos

A mis cuñadas

A mis sobrinos

por sus enseñanzas y apoyo incondicional.

*A mi pequeña gran familia Ramses y Giovanni
porque son la luz de mi vida.*

¡¡Mil gracias!!

Los amo

AGRADECIMIENTOS

A los millones de mexicanos que pagan impuestos, quienes, a través del Consejo Nacional de la Ciencia y la Tecnología (CONACYT) y el Colegio de Postgraduados me dieron la oportunidad y el apoyo financiero para la realización y termino de este postgrado.

A los integrantes de mi consejo particular por su tiempo, paciencia y consejos para la elaboración de este trabajo.

A todos los profesores que han tenido que ver con mi formación académica por regalarme su conocimiento.

A mis compañeros y amigos de la maestría: Blanca, Dayna, Luis y Alex por compartir y su apoyo incondicional.

A Rene y Paulino por sus valiosas intervenciones en la realización de este trabajo.

Al todo el personal que labora en Estadística, en especial a Emma, Isa, Carmen y Olga por su orientación y apoyo.

Índice

1. Introducción	1
2. Objetivo	4
3. Revisión bibliográfica	5
3.1. El <i>idh</i>	5
3.2. El <i>idh</i> en México	7
3.3. Construcción del <i>idh</i> de acuerdo a PNUD y Magaña	8
3.4. Componentes principales probabilísticas y bayesianas	11
3.4.1. Componentes principales probabilísticas	11
3.4.2. Inferencia Bayesiana	12
3.4.3. Métodos MCMC	13
3.4.3.1. Gibbs Sampling	14
3.4.3.2. Metropolis-Hasting	15

ÍNDICE

3.4.4.	Análisis de convergencia de Gelman y Rubin	15
3.4.5.	Diagnóstico de Raftery y Lewis	17
3.4.6.	Algoritmo Reversible Jump Markov Chain Monte Carlo (RJMCMC)	18
4.	Planteamiento del problema	21
5.	Metodología	24
5.1.	Modelo de componentes principales bayesianas	24
5.2.	Algoritmo RJMCMC utilizado para estimar las componentes principales	28
6.	Simulación	31
7.	Resultados y Discusión	33
7.1.	Componentes principales probabilísticas	33
7.2.	Componentes principales bayesianas	34
8.	Conclusiones	42
	Referencias	43
A.	Apéndice	46

Índice de tablas

3.1. Máximos y Mínimos de las variables del <i>idh</i>	9
4.1. Probabilidad <i>a posteriori</i> para los diferentes números de CP y los valores estimados del ejemplo de Zhang et al. (2004). .	23
6.1. Resultados de la simulación.	31
7.1. Ponderaciones estimadas para las variables que componen al <i>idh</i>	33
7.2. Resultados del diagnóstico de Raftery y Lewis para 1995 y 2000.	34
7.3. Cuantiles de la distribución <i>a posteriori</i> de σ^2	35
7.4. Intervalos de credibilidad para la estimación de l_1 en 1995 y 2000.	35
7.5. Ponderaciones para cada variable en 1995 y 2000.	38

Índice de figuras

7.1. Distribución <i>a posteriori</i> de σ^2 para 1995 y 2000	36
7.2. Distribución <i>a posteriori</i> de l_1 para 1995 y 2000	37
7.3. Mapas comparativos del índice de desarrollo humano en México en 1995	40
7.4. Mapas comparativos del índice de desarrollo humano en México en el 2000	41
A.1. Distribución <i>a posteriori</i> del número de componentes principales para 1995(a) y 2000(b)	46
A.2. Diagnóstico de convergencia de Gelman y Rubin para σ^2 para 1995 y 2000	47
A.3. Gráfica de la densidad estimada de σ^2 para 1995 y 2000	48
A.4. Diagnóstico de convergencia de Gelman y Rubin para l_1 en 1995 y 2000	49
A.5. Gráfica de la densidad estimada de l_1 para 1995 y 2000	50

Introducción

En 1990 el Programa de las Naciones Unidas para el Desarrollo (PNUD) introduce la idea de desarrollo humano y como medida de la calidad de vida de las personas propone al índice de desarrollo humano (*idh*).

PNUD define al desarrollo humano como un proceso mediante el cual se potencian las capacidades de las personas expandiendo sus opciones para que tengan una vida saludable, creativa y con los medios adecuados para desenvolverse en su entorno social (PNUD, 1990); note que de la definición podemos considerar muchas variables que nos pueden ayudar a describir al desarrollo humano como: esperanza de vida, educación, servicios con los que cuenta una vivienda, ingreso familiar, índices delictivos, índice de contaminación, si el lugar donde vive se considera rural o urbano, etc.

Sin embargo, el PNUD solo considera tres aspectos en los que se pretende incluir todas las variables que se mencionaron anteriormente, estas son: la salud, la educación y el acceso a recursos necesarios para tener un nivel de bienestar aceptable; estos tienen como variable de medición el índice de esperanza de vida (*InEsp*), el índice de educación (*InEdu*) y el índice del producto interno bruto per-cápita (*InPibp*).

1. Introducción

Socialmente, cuando alguien tiene un poder adquisitivo alto, es razonable pensar que se tiene acceso a educación de calidad, a un nivel cultural alto y, por tanto, se tiene conciencia de lo importante que es cuidar la salud. Así nos podemos dar cuenta de como estan interrelacionados salud, educación y el acceso a recursos necesarios para tener un nivel de bienestar aceptable, en otras palabras, se puede suponer que son variables correlacionadas.

El *idh* lo construye PNUD como el promedio ponderado del *InEsp*, *InEdu* e *InPibp* y tiene como objetivo caracterizar los países de acuerdo a su nivel de desarrollo.

El análisis de componentes principales (ACP) consiste en estandarizar combinaciones lineales de variables cuantitativas correlacionadas, como las que componen al *idh*, y transformarlas en nuevas variables no correlacionadas entre sí, maximizando la varianza de las combinaciones lineales resultantes, lo cual nos permite ver con más claridad las diferencias entre individuos, que es parte del objetivo del *idh*.

Existen trabajos en donde se ha aplicado la técnica de componentes principales en el *idh* para diferentes casos como [Biswas and Caliendo \(2002\)](#) y [Lai \(2002\)](#). En este trabajo se utilizó un análisis bayesiano sobre las componentes principales (ABCP), para construir un nuevo *idh*, utilizando indicadores socioeconómicos a nivel municipal de México correspondientes a los años 1995 y 2000, cuyo fin sea medir mas objetivamente las características socioeconómicas, permitiendo de esta manera jeraquizar los municipios de México.

Se utiliza el ABCP para hacer inferencia sobre las componentes principales (CP), teniendo como ventaja adicional el poder incorporar información *a priori* y la obtención del número de componentes principales más probable *a*

1. Introducción

posteriori.

Además, con este enfoque se puede obtener la distribución *a posteriori* del número de componentes principales a utilizar, de la varianza del error y de los eigenvalores; con la que se calculan intervalos de credibilidad y las estimaciones puntuales.

Como resultado de la aplicación de dicha técnica se obtienen nuevas ponderaciones para el *idh* siendo esta la primera componente principal. Los resultados no coinciden con la propuesta del PNUD, pues resulta que se le da una mayor ponderación a los servicios públicos en especial al drenaje, pero permiten jerarquizar mejor a los municipios de México de acuerdo con su nivel de desarrollo humano.

Objetivo

- Constatar la eficiencia del ABCP, al construir un nuevo *idh* el cual pondere de forma objetiva las características socioeconómicas, permitiendo jerarquizar mas eficientemente a los municipios de México de acuerdo con su nivel de desarrollo humano.
- Utilizar como herramienta para obtener las CP bayesianas las cadenas de Markov Monte Carlo con saltos reversibles (RJMCMC, por sus siglas en inglés)

Revisión bibliográfica

3.1. El *idh*

Desde 1990, cuando PNUD publica su primer reporte sobre desarrollo humano a nivel mundial, se inicia una polémica, ya que la forma de medir el desarrollo humano que proponía PNUD no era muy acertada desde el punto de vista conceptual y de instrumentación (Kelley, 1991; Ravallion, 1997; Srinivasan, 1994), debido a esto, el *idh* ha sufrido siete modificaciones sustanciales en su procedimiento de cálculo, realizadas con ayuda de Amartya Kuar Sen, quien fue nombrado en 1999 asesor distinguido de desarrollo humano del PNUD, por aportar las principales ideas sobre desarrollo humano y por bosquejar y redefinir el *idh*.

En 1991 en vez de usar en la dimensión de conocimiento solo la alfabetización en adultos, se agrega la escolaridad promedio, combinándolas 2/3 y 1/3 respectivamente; se establece una nueva fórmula para ponderar distintos niveles de ingreso, de forma que entre más alto sea el Pibp en relación con la línea de pobreza de los países desarrollados, el exceso registrado tiene contribuciones positivas al índice pero cada vez menores (PNUD, 2002).

3.1. El *idh*

En 1994 en vez de corregir el ingreso con una línea de pobreza internacional, se utiliza como umbral al Pibp real medio mundial de 5 mil 120 dólares en paridad de poder de compra (PPC); y se fijan nuevos valores normativos máximos y mínimos: para la esperanza de vida 85 y 25 años, para la tasa de alfabetización 100 % y 0 %; para el promedio de años de escolaridad 15 y 0 años; para el ingreso 40 mil y 200 dólares PPC (PNUD, 2002).

En 1995 se reemplaza la escolaridad promedio por la tasa de matriculación combinada: primaria, secundaria y media superior, modificación que obedece a lo complicado y demandante del cálculo de la escolaridad promedio, que obliga a usar estimaciones no siempre confiables; también en este año se modifica el valor mínimo del ingreso a 100 dólares PPC debido a que éste es el valor mínimo observado para las mujeres y es necesario utilizarlo para mantener la coherencia entre los índices específicos para cada género (PNUD, 2002).

En 1999 se sustituye la fórmula para ponderar el Pibp, de manera que se descuenta en toda la distribución y no sólo por encima de cierto nivel, evitando reducir indebidamente su contribución al desarrollo en los países de ingreso medio (PNUD, 2002).

Biswas and Caliendo (2002) proponen aplicar CP sobre el *idh* para los 162 países que PNUD considera en su reporte de desarrollo humano 2001; plantean un razonamiento totalmente sociológico con el fin de justificar la aplicación de CP por el hecho de que las variables son interdependientes entre sí.

Utilizan la primera componente principal (CP1) para asignar nuevos pesos ya que es la que concentra la mayor variación de los datos originales, sin embargo, es necesario hacer un reescalamiento ya que el *idh* por ser índice toma valores entre cero y uno, por tanto, las ponderaciones deben sumar uno.

3.2. El *idh* en México

Un dato que les llama la atención, es que las varianzas de los datos originales son muy semejantes, lo que les lleva a pensar desde ese momento del análisis, que las nuevas ponderaciones serán muy parecidas a las propuestas por PNUD. Como resultado final concluyen que efectivamente como lo plantea PNUD la ponderación para el *InEsp*, *InEdu* e *InPibp*, que conforma al *idh* debe ser 1/3.

En el mismo año [Lai \(2002\)](#) utiliza componentes principales ponderadas por el tamaño de la población para medir y analizar el progreso del desarrollo humano en china desde el año 1990; concluye que este es un mejor método, dada la variabilidad del tamaño de las poblaciones en china esta técnica diferencia más objetivamente las áreas con mayor y menor desarrollo humano.

3.2. El *idh* en México

En México se han realizado trabajos sobre desarrollo humano utilizando la metodología propuesta por PNUD como es el caso de el Primer Informe sobre Desarrollo Humano de México 1995 ([Sánchez et al., 2000](#)) a escala municipal, estatal y nacional, y el Segundo Informe sobre Desarrollo Humano de México: Análisis comparativo del Desarrollo Humano en el lustro 1995-2000 ([Sánchez et al., 2001](#)); los cuales se difundieron en diciembre de 2000 y 2001, siendo publicados hasta febrero de 2003 y 2004 respectivamente.

Por otro lado [Magaña \(1999\)](#) propone el índice de desarrollo humano con servicios (*idhs*), el cual considera, al igual que el *idh* de PNUD, los mismos tres aspectos esenciales, solo que el acceso a recursos necesarios para tener un nivel de bienestar aceptable lo mide a través de los servicios a los que tiene acceso la población como es el agua entubada (*InAgu*), la electricidad

3.3. Construcción del *idh* de acuerdo a PNUD y Magaña

(*InEle*) y el drenaje (*InDre*).

No podía faltar la aportación de [PNUD \(2002\)](#), quien hace un análisis a nivel estatal y luego a nivel municipal usando los datos del año 2000; los aspectos relevantes a mencionar sobre este informe es, que existe una alta desigualdad regional en el país, siendo el indicador del Pib el que detectara la mayor, seguido del indicador de educación y finalmente el indicador de esperanza de vida, además, que los municipios que ocupan los diez primeros lugares en México, respecto al *idh*, están concentrados en el Distrito Federal, Nuevo León, Estado de México, Morelos y Oaxaca. Por otra parte, los diez municipios con menor *idh* pertenecen a los estados de Oaxaca, Chiapas, Veracruz y Guerrero.

Como antecedente de la teoría de componentes principales probabilísticas está el trabajo de [Tipping and Bishop \(1999\)](#) quienes en su artículo proponen un método para la estimación de máxima verosimilitud de los parámetros en un modelo con variables latentes. De aquí [Green \(1995\)](#) parte para hacer el análisis bayesiano de componentes principales, e ilustra la metodología del RJMCMC. Con esta base lo han podido aplicar, con sus variaciones y para resolver distintos casos [Zhang et al. \(2004\)](#) y [Mohamed et al. \(2002\)](#).

3.3. Construcción del *idh* de acuerdo a PNUD y Magaña

Un vector x se dice que es combinación lineal de un conjunto de vectores $A = \{x_1, x_2, \dots, x_n\}$ si existe una forma de expresarlo como suma de parte o todos los vectores de A multiplicados cada uno de ellos por un coeficiente

3.3. Construcción del *idh* de acuerdo a PNUD y Magaña

escalar a_1, a_2, \dots, a_n , de forma que:

$$x = a_1x_1 + a_2x_2 + \dots + a_nx_n = \sum_{i=1}^n a_ix_i$$

Para construir cada uno de los índices que constituyen al *idh* es necesario contar con las siguientes variables (X): esperanza de vida (Esp), alfabetización (Alf), matrícula infantil (Mat), producto interno bruto per-capita ($Pibp$), servicio de agua entubada (Agu), servicio de electricidad (Ele) y servicio de drenaje (Dre) para posteriormente aplicar a cada variable la siguiente fórmula:

$$InX = \frac{\text{valor}X - \text{mínimo}X}{\text{máximo}X - \text{mínimo}X}$$

En el caso del $Pibp$ se requiere el logaritmo natural (Ln)

$$InPibp = \frac{\text{Ln}(Pibp) - \text{Ln}(\text{mínimo})}{\text{Ln}(\text{máximo}) - \text{Ln}(\text{mínimo})}$$

donde, en la Tabla 3.1 se pueden observar los valores máximos y mínimos que se le han asignado a cada variable.

Tabla 3.1: Máximos y Mínimos de las variables del *idh*

Variable	Mínimo	Máximo
<i>Esp</i>	30 años	90 años
<i>Alf</i>	0 %	100 %
<i>Mat</i>	0 %	100 %
<i>Pibp</i>	100 dolares PPC	40000 dolares PPC
<i>Agu</i>	0 %	100 %
<i>Ele</i>	0 %	100 %
<i>Dre</i>	0 %	100 %

Ya con los índices calculados, el *idh* se calcula como sigue:

$$idh = \frac{1}{3}InEsp + \underbrace{\frac{1}{3} \frac{2}{3}InAlf + \frac{1}{3}InMat}_{InEdu} + \frac{1}{3}InPibp$$

3.3. Construcción del *idh* de acuerdo a PNUD y Magaña

donde:

InEsp: Índice de esperanza de vida

InAlf: Índice de alfabetización

InMat: Índice de matriculación infantil

InEdu: Índice de educación

InPibp: Índice del producto interno bruto per cápita

Por otra parte, el *idhs* de Magaña se construye como sigue:

$$idhs = \frac{1}{3}InEsp + \underbrace{\frac{1}{3} \frac{2}{3}InAlf + \frac{1}{3}InMat}_{InEdu} + \underbrace{\frac{1}{3} \frac{1}{3}InAgu + \frac{1}{3}InEle + \frac{1}{3}InDre}_{InServ}$$

donde:

InEsp: Índice de esperanza de vida

InAlf: Índice de alfabetización

InMat: Índice de matriculación infantil

InEdu: Índice de educación

InAgu: Índice de habitantes con agua entubada

InEle: Índice de habitantes con electricidad

InDre: Índice de habitantes con drenaje

InServ: Índice de servicios

Se uso la proporción de niños de 6 a 14 años que saben leer y escribir como proxy del *InMat*.

Todos los datos de las variables señaladas anteriormente se obtuvieron del Instituto de Estadística, Geografía e Informática (INEGI) en el Sistema Municipal de Bases de Datos (SIMBAD) y el Banco de Información Económica para los correspondientes años.

3.4. Componentes principales probabilísticas y bayesianas

3.4.1. Componentes principales probabilísticas

El análisis de componentes principales es un método que fue introducido inicialmente por Karl Pearson en 1901 y desarrollado por Hotelling en 1933 como sigue: Para un vector d -dimensional con datos observados $t_n, n \in \{1, \dots, N\}$, los q ejes principales $w_j, j \in \{1, \dots, q\}$ son ejes ortonormales, los cuales tienen varianza máxima bajo la proyección. Están dados por los correspondientes eigenvectores de la matriz de varianzas y covarianzas

$$S = \sum_n (t_n - \bar{t})(t_n - \bar{t})^T / N$$

donde \bar{t} es la media muestral, tal que, $Sw_j = \lambda_j w_j$. Así las q componentes principales observadas del vector t_n están dadas por $x_n = W^T(t_n - \bar{t})$ variables no correlacionadas, donde $W = (w_1, w_2, \dots, w_q)$.

En otras palabras el análisis de componentes principales es una técnica multivariada utilizada principalmente para reducir la dimensionalidad de un modelo, la cual consiste en hacer combinaciones lineales estandarizadas de las variables cuantitativas originales, correlacionadas en alguna medida, transformándolas en nuevas variables cuantitativas no correlacionadas entre sí, conocidas como componentes principales; dichas componentes tienen características deseables como ver con más claridad las diferencias entre individuos, jerarquizándolos por el criterio que se usa en la construcción de éstas sobre las variables originales, puesto que la varianza entre los individuos, de las combinaciones lineales resultantes, se hace máxima en la primera de ellas y decrece secuencialmente en las siguientes ([Tipping and Bishop, 1999](#)).

3.4. Componentes principales probabilísticas y bayesianas

Por sus propiedades ha sido un método utilizado para resolver problemas en muchas ramas de las ciencias, sin embargo, no ha cubierto todas las necesidades, como el hecho de saber cuantas componentes principales se deben utilizar para obtener la máxima información de estas nuevas variables, además es una herramienta solamente descriptiva, es por eso que con ayuda del análisis bayesiano se ha podido modificar el método.

3.4.2. Inferencia Bayesiana

Para realizar inferencias bayesianas acerca de un parámetro θ se supone un modelo paramétrico:

$$x \sim f(x | \theta); \theta \sim \Pi(\theta)$$

Donde los datos u observaciones proponen cierta distribución $f(\cdot)$ dependiente del parámetro de interés θ . También, debemos proponer una distribución *a priori* para θ la cual contiene información previa relevante sobre la cantidad de interés.

Luego, dado el modelo y los datos observados, la inferencia se basa completamente en la distribución final del parámetro condicionado en el vector de observaciones, la cual se obtiene a partir del teorema de Bayes, esto es:

$$\Pi(\theta | x) = \frac{f(x | \theta)\Pi(\theta)}{\int f(x | \theta)\Pi(\theta)d\theta} \quad (3.1)$$

Cabe notar que:

$$\Pi(\theta | x) \propto f(x | \theta)\Pi(\theta)$$

puesto que la integral del denominador en 3.1 no depende de θ , es solamente

3.4. Componentes principales probabilísticas y bayesianas

una constante de normalización para la distribución *a posteriori*.

En términos estadísticos, el teorema de Bayes actualiza la información sobre el parámetro a través de las observaciones obtenidas (experimento). Lo que se puede hacer debido a que los parámetros y observaciones tienen, conceptualmente, el mismo nivel; ambos tienen distribuciones (Gelman et al., 2004).

Para la obtención de la distribución *a posteriori* de los parámetros es necesario implementar métodos de simulación los cuales se basan en Cadenas de Markov Monte Carlo (MCMC, por sus siglas en inglés).

3.4.3. Métodos MCMC

Una cadena de Markov es un modelo matemático de sistemas estocásticos donde los estados dependen de probabilidades de transición; el estado actual solo depende del estado anterior:

$$X_t \mid X_{t-1}, \dots, X_0 \sim P(X_t \mid X_{t-1}, \dots, X_0) = P(X_t \mid X_{t-1})$$

El método Monte Carlo es un método no determinístico o estadístico numérico usado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud; a diferencia de los métodos numéricos que se basan en evaluaciones en N puntos en un espacio M dimensional para producir una solución aproximada, el método Monte Carlo tiene un error absoluto de la estimación que decrece como $\frac{1}{\sqrt{(x)}}$ de acuerdo al teorema del límite central. La idea principal del método es hacer muchas repeticiones basándose en un generador de números aleatorios con el fin de reconocer el comportamiento del sistema.

3.4. Componentes principales probabilísticas y bayesianas

De las definiciones anteriores podemos decir que los métodos MCMC son técnicas que simulan una cadena de Markov cuyos estados siguen una distribución probabilística dada en un espacio de estados de grandes dimensiones. Después de un gran número de iteraciones se utiliza el estado de la cadena, como una muestra de la distribución deseada. La calidad de la muestra mejora en función del número de pasos de la cadena.

Por lo general, no es difícil construir una cadena de Markov con las propiedades deseadas. El problema más difícil es determinar cuántos pasos son necesarios para la convergencia a la distribución estacionaria dentro de un error aceptable. Una buena cadena tendrá una "rapid mixture", es decir, se llega rápidamente a la distribución estacionaria a partir de una posición arbitraria (Brooks, 1998).

Los métodos MCMC mas conocidos son el algoritmo de muestreo de Gibbs y el Metropolis-Hasting.

3.4.3.1. Gibbs Sampling

El algoritmo Gibbs Sampling es un caso especial del algoritmo Metropolis-Hasting, su objetivo es obtener una distribución multivariada a partir de una muestra de la distribución condicional que resulta de integrar una distribución conjunta. Supongamos que se desea muestrear k valores de x de la distribución conjunta $p(x, y)$. Se comienza con un valor y_0 y una muestra x obtenida de $x_i \sim p(x | y = y_{i-1})$. Una vez que se calcula el valor de x , se repite el muestreo para la próxima $y_i \sim p(y | x = x_i)$.

3.4. Componentes principales probabilísticas y bayesianas

3.4.3.2. Metropolis-Hasting

Este algoritmo diseña una cadena de Markov que se desarrolla en dos etapas. En la primera etapa se propone un nuevo punto para alguna distribución propuesta, y en la segunda etapa este punto propuesto es aceptado o rechazado. Si el punto propuesto es aceptado, la cadena de Markov es desplazado a este punto, y si el punto es rechazado, la cadena de Markov permanece en la posición en que se encuentra. Seleccionando correctamente la probabilidad de aceptación, terminamos por crear una cadena de Markov que tiene una distribución estacionaria.

3.4.4. Análisis de convergencia de Gelman y Rubin

Gelman y Rubin en 1992 proponen un enfoque general para monitorear la convergencia de un proceso MCMC cuyas salidas son $m > 1$ cadenas paralelas las cuales se ejecutan con valores iniciales muy dispersos en relación con la distribución *a posteriori*. La convergencia es diagnosticada cuando las cadenas han "olvidado" sus valores iniciales, y la salida de todas las cadenas no se puede distinguir. El diagnóstico se aplica a cada una de las variables de la cadena. Se basa en una comparación de las varianzas dentro la cadena y entre las m cadenas.

Hay dos maneras de estimar la varianza de la distribución estacionaria: la media de la varianza de los datos empíricos dentro de cada cadena, W , y la varianza empírica combinada de todas las cadenas, que puede ser expresada como:

$$\hat{\sigma}^2 = \frac{(n-1)W}{n} + \frac{B}{n}$$

donde n es el número de repeticiones y $\frac{B}{n}$ es la varianza empírica entre cade-

3.4. Componentes principales probabilísticas y bayesianas

nas.

Si las cadenas convergen, entonces, ambas estimaciones son parecidas. De lo contrario, el primer método podría subestimar la varianza, ya que cada cadena no ha tenido tiempo para muestrear toda la distribución estacionaria, y el segundo método podría sobreestimar la varianza, puesto que los puntos de inicio son dispersos.

El diagnóstico de convergencia se basa en el supuesto de que la distribución objetivo es normal. Un intervalo creíble bayesiano puede ser construido utilizando una distribución t-student con media

$$\hat{\mu} = \text{la media simple de todas las cadenas combinadas}$$

y varianza

$$\hat{V} = \hat{\sigma}^2 + \frac{B}{mn}$$

y con grados de libertad estimados por el metodo de momentos

$$d = \frac{2\hat{V}^2}{\text{Var}(\hat{V})}$$

El estadístico para el diagnóstico de convergencia por sí mismo es:

$$R = \sqrt{\frac{(d+3)\hat{V}}{(d+1)W}}$$

Valores sustancialmente mayores que 1 indican la falta de convergencia. Si las cadenas no convergen, los intervalos creíbles bayesianos basados en la distribución t-student son demasiado amplios, y tienen el potencial de reducirse por este factor si la corrida del proceso MCMC es continua. O bien si el valor del estadístico de Gelman-Rubin o "factor de reducción potencial

3.4. Componentes principales probabilísticas y bayesianas

de escala” (SRF por su nombre en inglés) está por debajo de 1.2 nos indica convergencia (Raftery and Lewis, 1995).

3.4.5. Diagnóstico de Raftery y Lewis

La estimación del tamaño de muestra para una variable U se basa en el proceso $Z_t = I(U_t \leq u)$, donde I es la función indicadora y u es el q -ésimo cuantil de U . El proceso Z_t se deriva de la obtención de una cadena de Markov por medio de la marginalización y el truncamiento, pero por sí mismo no es una cadena de Markov. Sin embargo, Z_t puede comportarse como una cadena de Markov si es suficientemente dispersa. Con este procedimiento se calcula el valor más pequeño k tal que $Z_t^{(k)}$ se comporta como una cadena de Markov. El tamaño de la muestra se calcula a partir de estas secuencias. Mientras que algunos datos se ”tiran”, en tanto, que la muestra de tamaño N se conserva.

El criterio para las m primeras iteraciones que se vayan a desechar (”burn in”), es que la distribución condicional de Z_m^k dado Z_0 debe estar dentro del equilibrio de la distribución de la cadena Z_t^k .

Este diagnóstico da como resultado M que es el número de observaciones que se deben ”tirar”o periodo de calentamiento, N el tamaño de la muestra requerido, N_{min} el tamaño de muestra mínimo requerido basado en la autocorrelación cero e $I = \frac{(M+N)}{N_{min}}$ el ”factor de dependencia” el cual nos indica autocorrelación fuerte si sus valores son mas grandes que 5 (Raftery and Lewis, 1995).

3.4.6. Algoritmo Reversible Jump Markov Chain Monte Carlo (RJMCMC)

El algoritmo RJMCMC es un método MCMC propuesto por [Green \(1995\)](#), este algoritmo se basa en la creación de cadenas de Markov irreducibles y aperiódicas, las cuales puedan ajustar a varios modelos con espacio de parámetros de diferentes dimensiones, manteniendo un equilibrio, y asegurando la distribución límite correcta.

El algoritmo RJMCMC ha sido usado en diferentes situaciones, especialmente en la estimación de las distribuciones *a posteriori*, como ejemplo se tiene a [Richardson and Green \(1997\)](#) que aplican el algoritmo para hacer inferencia bayesiana acerca de un modelo mixto con el número de componentes mixtos desconocido. Implementan dos movimientos: uno en el cuál se crean nuevas componentes, o una ya existente se elimina, y el otro donde una componente se divide en dos, o dos son combinadas para formar una.

En un problema estadístico que involucre inferencias acerca de cierta distribución de interés, que se define a través de la unión de subespacios C_r de diferentes n_r dimensiones, como las encontradas en el problema de selección de modelos. [Green \(1995\)](#) desarrolla el algoritmo RJMCMC como una aplicación típica a este tipo de problemas. En la aplicación es necesario realizar diferentes tipos de movimientos entre los subespacios. Estos se combinan para formar lo que llaman una muestra híbrida, por la selección aleatoria entre los movimientos disponibles en cada transición, de la combinación de espacio de parámetros $C = \bigcup_{i=1}^L C_i$. Los cambios de espacios de parámetro del modelo se hacen empleando la distribución propuesta, y la probabilidad de aceptación destinada a preservar el equilibrio, y con esto, se garantiza la convergencia a la distribución correcta.

3.4. Componentes principales probabilísticas y bayesianas

Sea $M_0, M_1, \dots, M_{kmax}$ una colección de posibles modelos y suponga que M_r es el modelo con θ_r el vector de parámetros desconocidos de tamaño n_r , el cual puede variar de un modelo a otro. Suponga que se desea pasar del modelo M_i al nuevo modelo M_j , entonces una versión típica del algoritmo RJMCMC es el siguiente:

1. Se propone un nuevo modelo M_j con $r_{ij}(\theta_i)$ la probabilidad de que el movimiento del modelo sea de M_i a M_j
2. Se genera un vector de variables aleatoria u de una densidad específica propuesta $q(u)$
3. Se necesita un movimiento reversible entre los modelos M_i y M_j . Esto se hace mediante una función biyectiva g que transforme los parámetros $q(\theta_i, u) = (\theta_j, u')$, y conserve las dimensiones $dim(\theta_i) + dim(u) = dim(\theta_j) + dim(u')$. La transformación inversa g^{-1} nos proporciona el movimiento en la otra dirección. Donde los vectores u y u' se utilizan como dimensiones de los espacios de parámetros de M_i y M_j .
4. Se acepta moverse al modelo M_j con probabilidad

$$\alpha = \min \left(1, \frac{f(M_j, \theta_j | y) r_{ij}(\theta_j) q'(u')}{f(M_i, \theta_i | y) r_{ij}(\theta_i) q(u)} \right) \left| \frac{dg(\theta_i, u)}{d(\theta_i, u)} \right|$$

5. Si no se acepta el movimiento, permanecemos con el modelo M_i con el vector de parámetros θ_i .
6. Repetir los pasos del 1 al 5, m veces bajo el criterio del diagnóstico de Raftery y Lewis .

Donde $q(u')$ es la densidad de u' independiente de θ_j (con dimensión m_j) y $|\cdot|$ es el jacobiano. El movimiento inverso, es decir, pasar del modelo M_j al

3.4. Componentes principales probabilísticas y bayesianas

M_i tiene probabilidad de aceptación α' dada por:

$$\alpha = \min \left(1, \frac{f(M_i, \theta_i | y) r_{ij}(\theta_i) q(u)}{f(M_j, \theta_j | y) r_{ij}(\theta_j) q'(u')} \right) \left| \frac{dg^{-1}(\theta_i, u)}{d(\theta_j, u')} \right|$$

En la aplicación del algoritmo RJMCMC se encuentran serias dificultades, particularmente en la obtención de la cadena de salto de un modelo a otro. Hay dos cosas que se deben considerar cuando se construyen los movimientos entre los modelos de dimensiones diferentes: la densidad propuesta y la función biyectiva g (que corresponda a la función).

Existen propuestas y sugerencia de como seleccionar la densidad, por ejemplo, [Brooks et al. \(2003\)](#) proponen un sistema de selección de la densidad q mas eficiente y de como construir la función g correspondiente.

Planteamiento del problema

Como se ha mencionado, la idea del desarrollo humano y su forma de medirlo ha sido un tema de discusión, básicamente por que surgen preguntas como ¿Que variables deben incluirse en el índice? ¿En que proporción deben tomarse en cuenta las variables? ¿Como homogeneizar las bases de datos a utilizar?.

Como parte de este debate [Ravallion \(1997\)](#) y [Kelley \(1991\)](#) señalan que el promedio simple de la formula de agregación hace posible la sustitución entre sus tres variables. [Kelley \(1991\)](#) por su parte dice que PNUD no justifica el uso de la formula de agregación que proponen y sugiere se le de más peso al Pibp, pues este refleja la capacidad que tienen los individuos de escoger aspectos en su vida que se relacionan con el desarrollo humano, sin embargo, hay que notar que esta sugerencia nos translada al concepto de ingreso y nos aleja del concepto de capacidades, que es donde el desarrollo humano hace mucho énfasis.

[Srinivasan \(1994\)](#) afirma que no existe correspondencia entre el concepto de las capacidades y su instrumentación a través del *idhp* debido a que las capacidades no tienen una métrica homogénea, a diferencia del enfoque del in-

4. Planteamiento del problema

greso que sí tiene esa característica; por otro lado, el autor opina que pueden existir problemas en la forma de crear las bases de datos, como ejemplos tenemos a los procedimientos de extrapolación de datos de algunos países a todos los demás para poder construir una medida del Pib homogénea a través del poder de paridad de compra (PPC); para algunos países menos desarrollados no existe el indicador de esperanza de vida y se tiene que estimar a partir de modelos matemáticos. También, resalta la dificultad que representa el que la alfabetización se defina de distinta forma entre los países, así como no poder catalogar la matrícula escolar por calidad de las escuelas.

Otro problema, es que el *idhp* tiene serios problemas para ser comparable en el tiempo y en el espacio, principalmente debido a las variaciones que ha sufrido su metodología de cálculo, lo que dificulta las inferencias acerca de los procesos de desarrollo.

Debido a los problemas de tipo conceptual y de instrumentación que se han señalado anteriormente se ha tenido la inquietud de obtener una medida más objetiva del desarrollo humano, con especial interés en México, por eso se propone construir un índice haciendo uso del análisis bayesiano.

Retomando la idea de [Biswas and Caliendo \(2002\)](#), en este trabajo se utilizan CP con base en la información a nivel municipal en México de los años 1995 y 2000. Se toman en consideración las variables que se han propuesto para medir el desarrollo humano, las cuales son *InEsp*, *InAlf*, *InMat*, *InPibp*, *InEle*, *InAgu*, *InDre*; note que es apropiado usar esta técnica, ya que las variables que se utilizan están correlacionadas, entonces las CP sintetiza esa información, por lo que cada una proporciona un índice de desarrollo humano que va a diferenciar los municipios de México más objetivamente.

Desde un enfoque bayesiano podemos construir CP con el fin de mejorar

4. Planteamiento del problema

la exactitud de las estimaciones y mediciones del modelo de CP, mediante la incorporación del conocimiento previo sobre estas cantidades, a través de una función de densidad *a priori*. Además nos permite inferir el número de componentes principales a considerarse como índices, esto es importante, porque necesitamos un *idh* que distinga mejor a los municipios, y esto solo se consigue encontrando la combinación lineal óptima.

En su artículo [Zhang et al. \(2004\)](#) demuestran la eficacia del ACPB con un ejemplo en el cual siguen las recomendaciones de [Richardson and Green \(1997\)](#) de escoger los hiperparámetros $r > 1 > \alpha$ como $r = 3$, $\alpha = 0.5$ y $\eta = 1.2/V$ donde V es la desviación estándar de los datos. Ellos corrieron cadenas de 20000 observaciones y después de realizar el diagnóstico de Raftery y Lewis solo se quedaron con las últimas 10000.

El problema consta de 6 variables con varianza igual a 10, 7, 5, 3, 1, 1 respectivamente para cada variable. Los valores propios que se generaron a partir de la matriz de varianzas y covarianzas son: 8.958, 7.2862, 5.3011, 2.8964, 1.1012, 0.9876, respectivamente. En la Tabla 4.1 se muestran las probabilidades *a posteriori* para los diferentes números de CP (q) y sus correspondientes valores estimados de los eigenvalores (l_j) y la varianza del error (σ^2). Observe que la probabilidad de $q = 4$ es la más alta, lo que nos sugiere 4 dimensiones dominantes en esta serie de datos.

Tabla 4.1: Probabilidad *a posteriori* para los diferentes números de CP y los valores estimados del ejemplo de [Zhang et al. \(2004\)](#).

q	$p(q D)$	σ^2	l_1	l_2	l_3	l_4	l_5
4	0.8666	1.0573	9.0342	7.3198	5.2214	2.942	
5	0.1334	1.0263	9.0322	7.3187	5.2201	2.9403	1.093

Metodología

5.1. Modelo de componentes principales bayesianas

Para la aplicación del RJMCMC a componentes principales se planteará el modelo y el análisis bayesiano de acuerdo a la propuesta de [Zhang et al. \(2004\)](#).

Se tiene el siguiente modelo, el cual expresa una combinación lineal:

$$x = \sum_{j=1}^q h_j w_j + \mathbf{m} + \epsilon$$

$$x = \mathbf{H}\mathbf{w} + \mathbf{m} + \epsilon \quad (5.1)$$

$$\epsilon \sim N(0, \mathbf{V})$$

donde $x \in \mathfrak{R}^d$, $\mathbf{w} = (w_1, \dots, w_q)^T \in \mathfrak{R}^q, q \leq d$ y $\mathbf{H} = [h_1, \dots, h_q]$ es una matriz $d \times q$ que relaciona las variables x y \mathbf{w} . El vector \mathbf{m} permite que el modelo no tenga media cero y $\mathbf{V} = \sigma^2 \mathbf{I}_d$ es una matriz de varianzas y covarianzas hiperesférica. Las variables aleatorias latentes w_1, \dots, w_q

5.1. Modelo de componentes principales bayesianas

son normales independientes con media cero y cuya matriz de varianzas y covarianzas es la identidad, esto es, $\mathbf{w} \sim N(0, \mathbf{I}_q)$.

Los estimadores de máxima verosimilitud que obtuvieron [Tipping and Bishop \(1999\)](#) para \mathbf{m} y \mathbf{H} son:

$$\hat{\mathbf{m}} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\hat{\mathbf{H}} = \hat{\mathbf{U}}_q (\hat{\mathbf{L}}_q - \sigma^2 \mathbf{I}_q)^{\frac{1}{2}} \mathbf{R}$$

donde $\hat{\mathbf{U}}_q$ es una matriz ortogonal en la cual los q vectores columna corresponden a los eigenvectores de la matriz de varianzas y covarianzas de los datos ($\hat{\mathbf{S}}$), $\hat{\mathbf{L}}_q$ es una matriz diagonal $q \times q$ que contiene los correspondientes eigenvalores $\hat{l}_1, \dots, \hat{l}_q$, \mathbf{R} es una matriz ortogonal $q \times q$ arbitraria. El estimador de máxima verosimilitud de σ^2 esta dado por:

$$\hat{\sigma}^2 = \frac{1}{d - q} \sum_{j=q+1}^d \lambda_j$$

Los parámetros que se desean estimar son $\theta = (H, m, \sigma^2)$ y q el número de componentes principales que explican la mayor proporción de la varianza en los datos. [Mohamed et al. \(2002\)](#) en su artículo proporcionan una modificación del método anterior y el algoritmo para aplicar RJMCMC a las estimaciones bayesianas de las componentes principales. Se supone que la densidad conjunta de los datos $D = \{x_1, \dots, x_N\}$ y los parámetros tienen la forma

$$p(q, \theta, D) = p(q)p(\theta | q)p(D | \theta, q),$$

5.1. Modelo de componentes principales bayesianas

donde

$$\begin{aligned} p(D | \theta) &= \prod_{i=1}^N p(x_i | \mathbf{H}, \mathbf{m}, \sigma^2) \\ &= (2\pi)^{-Nd/2} |\mathbf{H}\mathbf{H}^T + \sigma^2\mathbf{I}|^{-N/2} e^{-\frac{N}{2}\text{tr}((\mathbf{H}\mathbf{H}^T + \sigma^2\mathbf{I})^{-1}\mathbf{S})}. \end{aligned} \quad (5.2)$$

La matriz \mathbf{H} puede descomponerse como

$$\mathbf{H} = \mathbf{U}_q(\mathbf{L}_q - \sigma^2\mathbf{I}_q)^{1/2}\mathbf{R}$$

donde $\mathbf{U}_q^T\mathbf{U}_q = \mathbf{I}_q$, $\mathbf{R}^T\mathbf{R} = \mathbf{I}_q$, y $\mathbf{L}_q = \text{diag}(l_1, \dots, l_q)$. Note que si \mathbf{U} es ortogonal $d \times d$, entonces $\mathbf{U} = (\mathbf{U}_q, \mathbf{U}_{d-q})$ y por tanto $\mathbf{U}_{d-q}^T\mathbf{U}_{d-q} = \mathbf{I}_{d-q}$. Entonces, si

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_q - \sigma^2\mathbf{I}_q & 0 \\ 0 & 0 \end{pmatrix}$$

Se tiene que:

$$\begin{aligned} \mathbf{U}\mathbf{L}\mathbf{U}^T &= \begin{pmatrix} \mathbf{U}_q & \mathbf{U}_{d-q} \end{pmatrix} \begin{pmatrix} \mathbf{L}_q - \sigma^2\mathbf{I}_q & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{U}_q^T \\ \mathbf{U}_{d-q}^T \end{pmatrix} \\ &= \mathbf{U}_q(\mathbf{L}_q - \sigma^2\mathbf{I}_q)\mathbf{U}_q^T \\ &= \mathbf{H}\mathbf{H}^T \end{aligned}$$

Entonces,

$$\begin{aligned} \mathbf{H}\mathbf{H}^T + \sigma^2\mathbf{I}_d &= \mathbf{U}\mathbf{L}\mathbf{U}^T + \sigma^2\mathbf{I}_d \\ &= \mathbf{U} \begin{pmatrix} \mathbf{L}_q & 0 \\ 0 & \sigma^2\mathbf{I}_{d-q} \end{pmatrix} \mathbf{U}^T \end{aligned} \quad (5.3)$$

Con la transformación en 5.3 se puede ver que el objetivo es estimar los parámetros $l = (l_1, \dots, l_q)$, σ^2 y q ; combinando 5.2 y 5.3 es posible escribir

5.1. Modelo de componentes principales bayesianas

la función de verosimilitud como sigue:

$$\begin{aligned}
 p(D | l_1^{-1}, \dots, l_q^{-1}, \sigma^{-2}) &= \prod_{i=1}^N p(x_i | l_1^{-1}, \dots, l_q^{-1}, \sigma^{-2}) \\
 &= (2\pi)^{-Nd/2} \prod_{j=1}^q l_j^{-N/2} \sigma^{-N(d-q)} e^{-\frac{N}{2} \sum_{j=1}^q l_j^{-1} g_j} e^{-\frac{N\sigma^{-2}}{2} \sum_{j=q+1}^d g_j}
 \end{aligned} \tag{5.4}$$

De acuerdo con el Teorema de Bayes, la distribución *a posteriori* de $l_1, \dots, l_q, \sigma^2$ se obtiene como el producto de la función de verosimilitud en 5.4 por la distribución *a priori* conjunta. Asumiendo que l_1, \dots, l_q son independientes e idénticamente distribuidas como:

$$l_1^{-1}, l_2^{-1}, \dots, l_q^{-1} \sim \text{Ga}(r, \tau)$$

Así como $q \sim U(1, d-1)$ y $\sigma^{-2} \sim \text{Ga}(r, \tau)$, donde $r > 0$ y $\tau \sim \text{Ga}(\alpha, \eta)$ con $\alpha > 0$ y $\eta > 0$. Entonces, la densidad *a priori* conjunta de los parámetros ordenados $l_1 > l_2 > \dots > l_q > \sigma^2$ dado q está dada por:

$$p(l_1, \dots, l_q, \sigma^2 | q) = (q+1)! p(l_1, \dots, l_q, \sigma^2) \mathbf{I}_{l_1 > l_2 > \dots > l_q > \sigma^2}(l, \sigma^2)$$

de donde, la densidad *a posteriori* conjunta es

$$\begin{aligned}
 p(l_1, \dots, l_q, \sigma^2, q | D) &\propto p(l_1, \dots, l_q, \sigma^2 | q) p(D | l_1^{-1}, \dots, l_q^{-1}, \sigma^{-2}) \\
 &\quad \mathbf{I}_{l_1 > l_2 > \dots > l_q > \sigma^2}(l, \sigma^2) \\
 &= \prod_{i=1}^{q+1} \text{Ga}(r, \tau) p(D | l_1^{-1}, \dots, l_q^{-1}, \sigma^{-2}) \mathbf{I}_{l_1 > l_2 > \dots > l_q > \sigma^2}(l, \sigma^2)
 \end{aligned} \tag{5.5}$$

5.2. Algoritmo RJMCMC utilizado para estimar las componentes principales

Dada una muestra inicial se actualizan los parámetros mediante el muestreo Gibbs, donde para simular de la distribución *a posteriori* conjunta se utilizan las densidades condicionales completas obtenidas a partir de 5.5:

$$l_j^{-1} \mid l_{-j}, \sigma^2, q, \tau \sim \text{Ga} \left(\frac{N}{2} + r, \frac{N g_j}{2} + \tau \right) \mathbf{I}_{[l_{j-1}, l_{j+1}]}(l_j)$$

$$\sigma^{-2} \mid l, q, \tau \sim \text{Ga} \left(\frac{N(d-q)}{2} + r, \frac{N \sum_{j=q+1}^d g_j}{2} + \tau \right) \mathbf{I}_{(0, l_q]}(\sigma^2)$$

$$\tau \mid l, \sigma^2, q \sim \text{Ga} \left((q+1)r + \alpha, \sum_{j=1}^q l_j^{-1} + \sigma^{-2} + \eta \right)$$

Para tomar la decisión de pasar al procedimiento de generar una nueva l_j^{-1} (nacimiento) o removerla (muerte), se genera una variable $u \sim U(0, 1)$.

- 1) Si $u \leq b_q$, donde d_k y b_k son respectivamente las probabilidades de muerte y nacimiento, las cuales toman los valores $b_j = d_j = 0.5$ cuando $j = 2, \dots, d-2$, $d_1 = b_{d-1} = 0$ y $b_1 = d_{d-1} = 1$; entonces, el algoritmo de muestreo es:

- i) Generar la nueva l_{q+1}^{-1} de la distribución *a priori* $\text{Ga}(r, \tau) \mathbf{I}_{[l_q, \sigma^2]}(l_{q+1})$.
- ii) Calcular la probabilidad de aceptación $\alpha = \min(1, R)$, donde R es el cociente de la verosimilitud cuando se tiene la nueva l_{q+1}^{-1} , que es el estado nuevo, y la verosimilitud del estado actual, quedando R como sigue:

$$R = \frac{\tau^r}{\Gamma(r)} l_{q+1}^{-1} r^{-1} e^{-\tau l_{q+1}^{-1}}$$

5.2. Algoritmo RJMCMC utilizado para estimar las componentes principales

- iii) Generar una variable aleatoria $v \sim U(0, 1)$.
 - iv) Si $\alpha > v$, entonces se acepta incorporar la nueva l_{q+1}^{-1} , de otra forma, mantener el estado actual y poner a prueba.
- 2) Si $u \leq b_q + d_q$, entonces:
- i) Se remueve la q -ésima componente principal.
 - ii) Calcular la probabilidad de aceptación $\alpha = \min(1, R^{-1})$, donde R es el cociente de la verosimilitud del estado actual y la verosimilitud cuando se tiene l_{q-1}^{-1} , quedando R como sigue:

$$R = \frac{\tau^r}{\Gamma(r)} l_q^{-1r-1} e^{-\tau l_q^{-1}}$$

- iii) Generar una variable aleatoria $v \sim U(0, 1)$.
- iv) Si $\alpha > v$, entonces se acepta remover la l_q^{-1} , de otra forma, mantener el estado actual y poner a prueba.

Repetiendo el algoritmo anterior se simula una cadena de Markov que converge a la distribución *a posteriori* conjunta, generando 2 cadenas por parámetro con longitud 10000 y de acuerdo con el diagnóstico de Raftery y Lewis, se estimó el tamaño de la muestra y el periodo de calentamiento.

Una vez que se estiman los eigenvalores l_1, \dots, l_q se estimaron los eigenvec-tores utilizando la formula

$$(\hat{L}_q I - \hat{S}) \hat{U}_q = 0 \quad (5.6)$$

La programación de este algoritmo se realizó en el lenguaje de programación R versión 2.10.1 (<http://cran.r-project.org/bin/windows/base/>). Para el análisis se incluyen siete variables correlacionadas: *InEsp*, *InAlf*, *InMat*, *InPibp*,

5.2. Algoritmo RJMCMC utilizado para estimar las componentes principales

InAgu, InEle, InDre utilizando como bases de datos los registros municipales para los años 1995 y 2000.

Los mapas comparativos del *idh* se construyeron con el programa Iris 4.0.2

Simulación

Con el propósito de analizar la capacidad predictiva del ABCP, se realizó una simulación, la cual consistió en generar conjuntos de 1000 datos para 5 variables aleatorias normales con media 1. Las varianzas son (10, 1, 1, 1, 1) para el primer conjunto de datos; (10, 5, 1, 1, 1) para el segundo; y (10, 5, 3, 1, 1) para el tercero. Se utilizó el algoritmo RJMCMC descrito en la sección 5.2 para estimar el número de componentes principales y los otros parámetros del modelo dado en (5.1).

Tabla 6.1: Resultados de la simulación.

Conjunto	q		l_1	l_2	l_3	l_4
1	1	Eigenvalor	9.48			
		Valor estimado	9.12			
		Varianza del error	0.0978			
		$P(q = 1 D) = 0.99$				
2	2	Eigenvalor	10.5	5.2		
		Valor estimado	8.063	4.319		
		Varianza del error	0.1089			
		$P(q = 2 D) = 0.91$				
3	3	Eigenvalor	10.3	4.56	2.78	
		Valor estimado	9.054	4.991	3.354	
		Varianza del error	0.0999			
		$P(q = 3 D) = 0.89$				

6. Simulación

La Tabla 6.1 proporciona los eigenvalores obtenidos a partir de la matriz de varianzas y covarianzas del conjunto de datos, el número de componentes a utilizar (q), el eigenvalor estimado (l_i), la varianza del error (σ^2) y la probabilidad del número de componentes principales.

Resultados y Discusión

7.1. Componentes principales probabilísticas

En la Tabla 7.1 se exhiben las ponderaciones estimadas por máxima verosimilitud (MV), que se le asignan a las variables que componen al *idh* de acuerdo con el procedimiento propuesto por [Tipping and Bishop \(1999\)](#) para calcular las CP. Se utilizan como ponderaciones los elementos estandarizados del primer eigenvector, para que al sumar ponderaciones el resultado sea 1 tal y como lo propone el PNUD.

Tabla 7.1: Ponderaciones estimadas para las variables que componen al *idh*.

Año	<i>InEsp</i>	<i>InAlf</i>	<i>InMat</i>	<i>InPibp</i>	<i>InAgu</i>	<i>InEle</i>	<i>InDre</i>
1995	0.008	0.124	0.076	0.089	0.214	0.130	0.358
2000	0.004	0.126	0.072	0.102	0.199	0.111	0.386

Observe que el valor estimado por MV de las ponderaciones es muy semejante a los valores de las ponderaciones cuando se utiliza el enfoque bayesiano, sin embargo, con MV no es posible estimar el número de componentes principales que son necesarias para maximizar la varianza de la combinación lineal del *idh*.

7.2. Componentes principales bayesianas

En la sección 5.2 se mencionó que, como muestra inicial se generaron 2 cadenas con 10000 observaciones cada una, con las cuales se realizó el diagnóstico de Raftery y Lewis. De acuerdo con los resultados que se exhiben en la Tabla 7.2, se decidió que, para propósitos del análisis, se utilizaran 2 cadenas con al menos 1000 observaciones, considerando las primeras 100 como periodo de calentamiento.

Tabla 7.2: Resultados del diagnóstico de Raftery y Lewis para 1995 y 2000.

Año	M	N	Nmin	I
1995	1	390	385	1
2000	2	392	385	1.02

Note que el periodo de calentamiento (M) resultante es considerablemente pequeño en ambos años, así como el tamaño de la muestra (N) y el tamaño mínimo de la muestra (N_{mín}) requerido; aún con valores tan pequeños, el factor de dependencia (I) es menor que 5, lo que es indicativo de una independencia entre las observaciones de las cadenas.

De la salida del RJMCMC, se obtuvo una muestra de la distribución *a posteriori* del número de componentes (q), los eigenvalores (l_j) y la varianza del error σ^2 .

Se observa en la Figura A.1 del Apéndice que la probabilidad *a posteriori* del número de componentes principales es aproximadamente igual a 1. Por lo anterior puede utilizarse sólo la primera componente principal como estimación del *idh* y con esto se maximiza la varianza de la combinación lineal resultante.

7.2. Componentes principales bayesianas

Las distribuciones *a posteriori* para σ^2 se grafican en la Figura 7.1(a) y 7.1(b); éstas, son una buena aproximación de la densidad de σ^2 ya que el diagnóstico de convergencia para σ^2 resultó que en ambos años el SRF=1.

Los valores de los cuantiles para σ^2 se muestra en la Tabla 7.3.

Tabla 7.3: Cuantiles de la distribución *a posteriori* de σ^2 .

Año	Cuantiles				
	2.5 %	25 %	50 %	75 %	97.5 %
1995	0.0115	0.0117	0.0118	0.0119	0.0121
2000	0.0088	0.009	0.0091	0.0092	0.0093

Con lo anterior, podemos decir, que por ejemplo, el valor verdadero para σ^2 en 1995 se encuentra entre 0.0115 y 0.0121 con probabilidad *a posteriori* de 0.95 y que, bajo perdida cuadrática, el valor estimado para σ^2 en el 2000 es 0.0091; además, que este valor es notablemente menor que el correspondiente al año 1995.

Ya con conocimiento de las distribuciones de q y σ^2 , resta analizar la de l_1 .

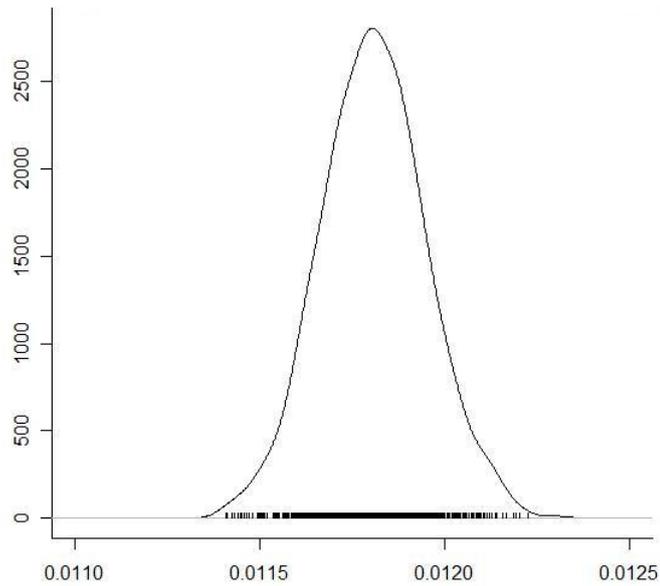
Con base en la muestra MCMC de la distribución *a posteriori* se estiman los cuantiles y con esto se pueden definir intervalos de credibilidad para l_1 (Tabla 7.4).

Tabla 7.4: Intervalos de credibilidad para la estimación de l_1 en 1995 y 2000.

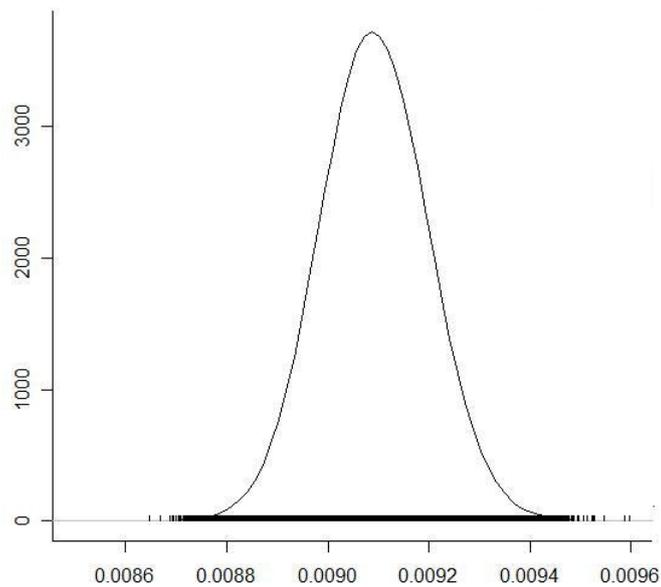
Año	Cuantiles				
	2.5 %	25 %	50 %	75 %	97.5 %
1995	0.129	0.134	0.136	0.139	0.144
2000	0.117	0.121	0.123	0.126	0.130

En promedio el valor para el primer eigenvalor en 1995 es 0.136 y para el 2000 es 0.123; además, con los cuantiles 2.5 % y 97.5 % se tiene un intervalo de credibilidad al 95 % para l_1 .

7.2. Componentes principales bayesianas



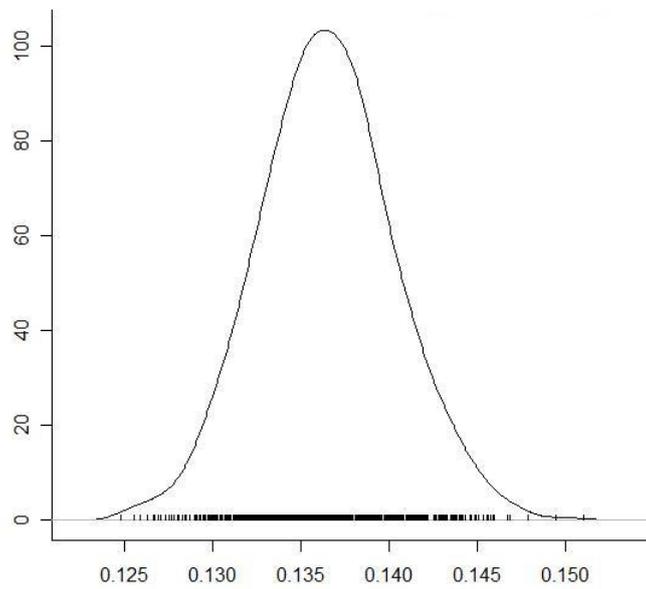
(a) Distribución *a posteriori* de σ^2 para 1995



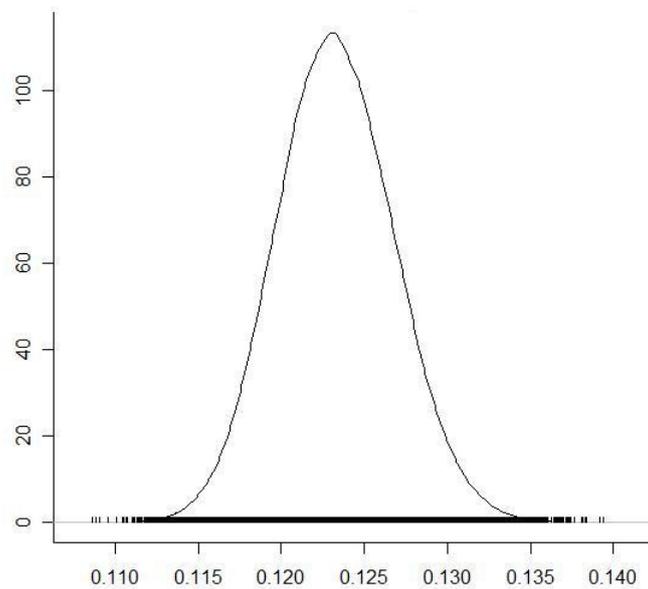
(b) Distribución *a posteriori* de σ^2 para el 2000

Figura 7.1: Distribución *a posteriori* de σ^2 para 1995 y 2000

7.2. Componentes principales bayesianas



(a) Distribución *a posteriori* de l_1 para 1995



(b) Distribución *a posteriori* de l_1 para el 2000

Figura 7.2: Distribución *a posteriori* de l_1 para 1995 y 2000

7.2. Componentes principales bayesianas

Es interesante observar en las Figuras 7.2(a) y 7.2(b) la distribución *a posteriori*, la cual se puede concluir que es una buena aproximación de la densidad de l_1 ya que en el diagnóstico de convergencia para l_1 también resultó que en ambos años el SRF=1. Se puede ver gráficamente la evolución del SRF y como conforme aumenta el número de iteraciones en las Figuras A.4(a) y A.4(b) del Apéndice; el comportamiento de la mediana del SRF tiende a uno y está por debajo del cuantil 0.975.

Una vez estimado l_1 es posible obtener, por medio de (5.6), su correspondiente eigenvector, el cual proporciona los nuevos pesos para cada variable utilizada para el cálculo del *idh*. Recuerde que es necesario estandarizar para que la suma de los nuevos pesos sea igual a uno, pues es una de las condiciones para la construcción del *idh*.

Note en la Tabla 7.5 que a la variable que más peso se le asignó fue al *InDre*, seguido por el *InAgu* e *InEle*; esto es, porque en México estos servicios no llegan en igual proporción a todos los municipios. También se registra una alta ponderación para el *InAlf*, lo cual indica que en algunas zonas del país aún existe una alta proporción de adultos analfabetas.

Tabla 7.5: Ponderaciones para cada variable en 1995 y 2000.

Año	<i>InEsp</i>	<i>InAlf</i>	<i>InMat</i>	<i>InPibp</i>	<i>InAgu</i>	<i>InEle</i>	<i>InDre</i>
1995	0.008	0.124	0.076	0.089	0.214	0.139	0.358
2000	0.004	0.126	0.072	0.102	0.199	0.111	0.386

PNUD también ha propuesto una jerarquización de acuerdo al valor del *idh* definiéndose como desarrollo humano bajo para aquellos valores del *idh* menores a 0.5, como desarrollo humano medio el intervalo [0.5, 0.8) y como desarrollo humano alto el intervalo [0.8, 1].

En las Figuras 7.3(a) y 7.4(a) se exhibe el mapa de los municipios, jerarquiza-

7.2. Componentes principales bayesianas

dos de acuerdo con el índice de desarrollo humano. El *idhb95* y el *idhb00* es el índice de desarrollo humano para 1995 y 2000, respectivamente, que se calcula con los pesos de la Tabla 7.5 como sigue:

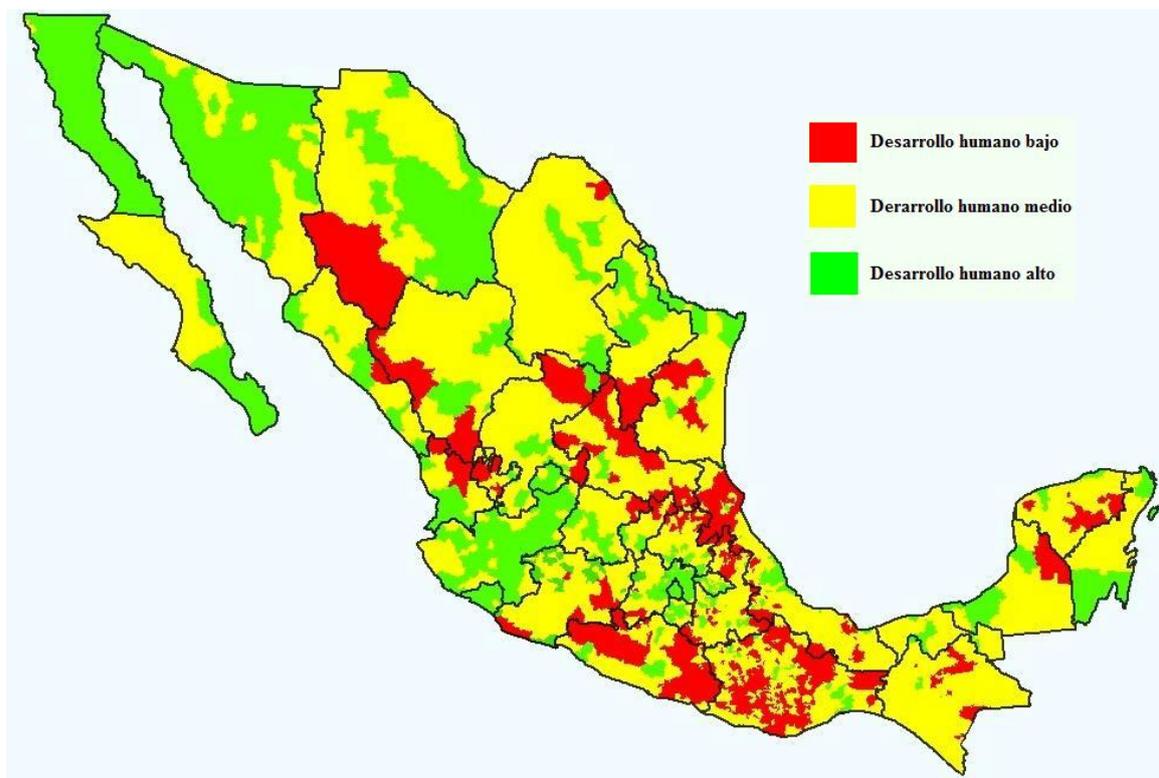
$$\begin{aligned} idhb95 = & 0.008InEsp + \underbrace{0.124InAlf + 0.076InMat}_{InEdu} + 0.089InPibp \\ & + \underbrace{0.214InAgu + 0.139InEle + 0.358InDre}_{InServ} \end{aligned}$$

$$\begin{aligned} idhb00 = & 0.004InEsp + \underbrace{0.126InAlf + 0.072InMat}_{InEdu} + 0.102InPibp \\ & + \underbrace{0.199InAgu + 0.111InEle + 0.386InDre}_{InServ} \end{aligned}$$

La figuras 7.3(b) y 7.4(b) representan el *idh* propuesto por el PNUD. De acuerdo a éste, la mayoría de los municipios tienen desarrollo humano alto y medio; sin embargo, debido al mayor peso que se le asigna a los servicios con el ABCP, los municipios en la zona de la sierra, que es donde regularmente se hace más difícil el acceso a los servicios de agua potable, electricidad y drenaje, son clasificados como de bajo desarrollo humano.

También observe que los estados con una mayor cantidad de municipios con desarrollo humano bajo son Oaxaca, Guerrero, Veracruz, Chiapas y los que tienen más municipios con desarrollo humano alto son Baja California Norte y Sur, Sonora, Distrito Federal, Jalisco, Chihuahua.

7.2. Componentes principales bayesianas



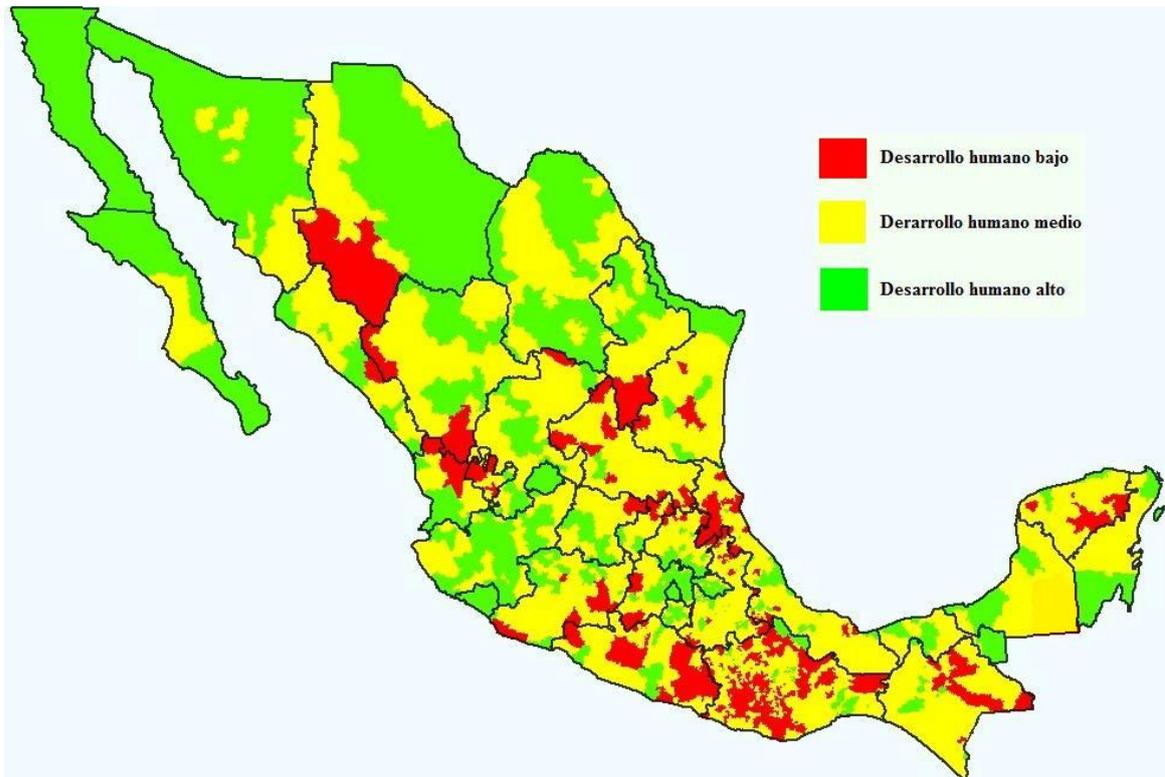
(a) Mapa del *idhb* en México en 1995 (ABCP)



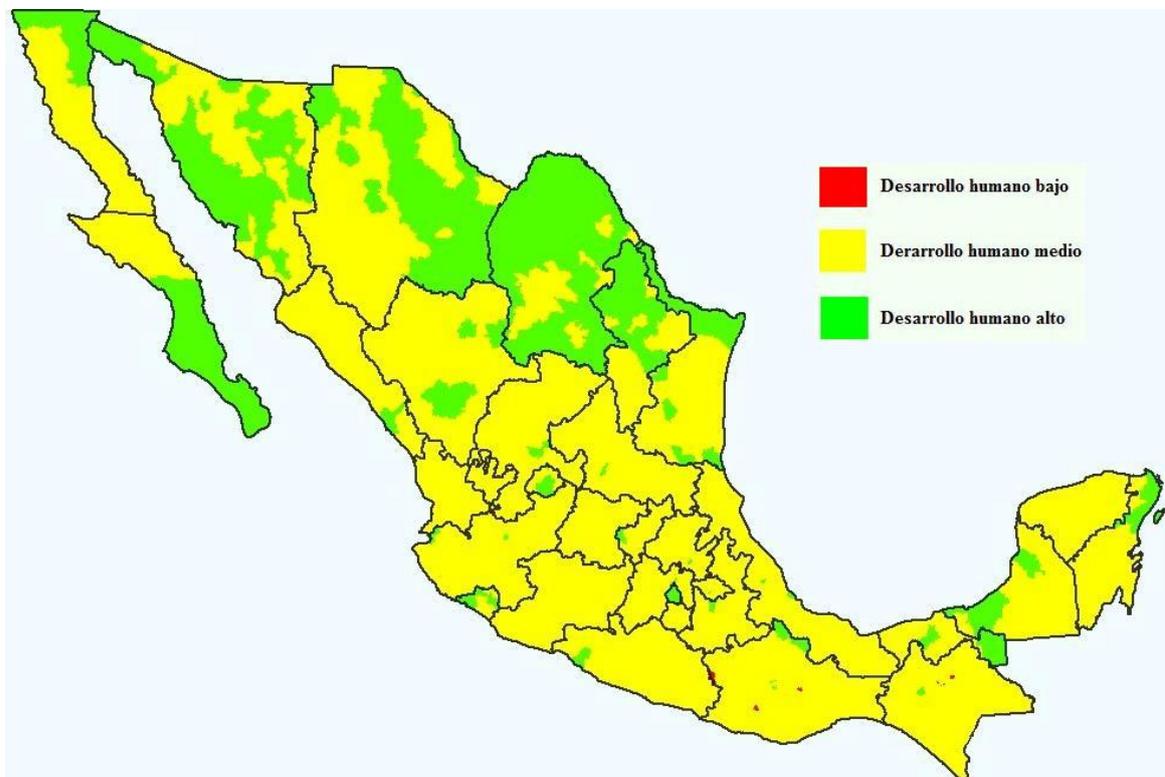
(b) Mapa del *idh* en México en 1995 (PNUD)

Figura 7.3: Mapas comparativos del índice de desarrollo humano en México en 1995

7.2. Componentes principales bayesianas



(a) Mapa del *idhb* en México en el 2000 (ABCP)



(b) Mapa del *idh* en México en el 2000 (PNUD)

Figura 7.4: Mapas comparativos del índice de desarrollo humano en México en el 2000

Conclusiones

Con el enfoque bayesiano para el análisis de componentes principales utilizado en este trabajo, se obtuvo:

1. La distribución *a posteriori* de los parámetros de interés: número de componentes principales, varianza residual y eigenvalores.
2. A partir de la distribución *a posteriori* se estimaron puntual y por intervalo los parámetros.
3. La estimación puntual del número de componentes principales se utilizó para la construcción de un nuevo índice de desarrollo humano, tomándose solo la primera componente principal.
4. Se propone un índice que, con base en la variabilidad de los datos, permite jerarquizar el desarrollo humano en México, observándose que los pesos estimados de las variables no concuerdan con las definiciones de desarrollo humano que propone el Programa de las Naciones Unidas para el Desarrollo; sin embargo, dado que los ponderadores se estiman a partir de los datos, este nuevo índice proporciona una clasificación más objetiva de los municipios de acuerdo con el desarrollo humano.

Referencias

- Biswas, B. and F. Caliendo (2002). A multivariate analysis of the human development index. *The Indian economic journal* 49, 96–100.
- Brooks, S. P. (1998). Markov chain monte carlo method and its application. *The statistician* 47, 69–100.
- Brooks, S. P., P. Giudici, and G. O. Roberts (2003). Efficient construction of reversible jump markov chain monte carlo proposal distributions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 65, 3–55.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). Bayesian Data Analysis.
- Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Journal Biometrika* 82, 711–732.
- Kelley, A. C. (1991). The human development index: 'handle with care'. *Population and Development Review* 17, 315–324.
- Lai, D. (2002). Principal component analysis on human development indicators of china. *Social Indicators Research* 61, 319–330.
- Magaña, A. R. (1999). Índice de desarrollo humano del estado de guanajuato. *Revista del centro de desarrollo humano de Guanajuato A.C* 3, 9–28.

REFERENCIAS

- Mohamed, N. N., B. R. Bhavik, G. K. Prem, and S. Xiaotong (2002). *Bayesian principal component analysis*. *Journal of Chemometrics* 16, 576–595.
- PNUD (1990). *Human development Report 1990*, Chapter 1 Defining and measuring human development, pp. 9–16.
- PNUD (2002). *Informe sobre desarrollo humano México 2002*.
- Raftery, A. and S. Lewis (1995). *The number of iterations, convergence diagnostics and generic Metropolis algorithms*. In *Practical Markov Chain Monte Carlo*.
- Ravallion, M. (1997). Good and bad growth: The human development reports. *World Development* 25, 631–638.
- Richardson, S. and P. J. Green (1997). *On bayesian analysis of mixtures with an unknown number of components (with discussion)*. *Journal of the Royal Statistical Society. Series B (Methodological)* 59, 731–739.
- Sánchez, F. J. Z., M. G. P. Miranda, and C. A. V. Tames (2000). *Primer Informe Sobre Desarrollo Humano de México 1995*.
- Sánchez, F. J. Z., M. G. P. Miranda, and C. A. V. Tames (2001). *Segundo Informe Sobre Desarrollo Humano de México: Análisis comparativo del Desarrollo Humano en el lustro 1995-2000*.
- Srinivasan, T. (1994). Human development: A new paradigm or the reinvention of the wheel? *American Economic Review, Papers and Proceedings* 84, 238–243.
- Tipping, M. E. and C. M. Bishop (1999). *Probabilistic principal component analysis*. *Journal of the Royal Statistical Society Series B* 61, 611–622.
- Zhang, Z., K. L. Chan, J. T. Kwok, and D.-Y. Yeung (2004). Bayesian inference on principal component analysis using reversible jump markov chain

REFERENCIAS

monte carlo. *Proceedings of the 19th National Conference on Artificial Intelligence, San Jose, California, USA, 25-29 July 2004*, AAAI Press, Menlo Park, California, 372–377.

Apéndice

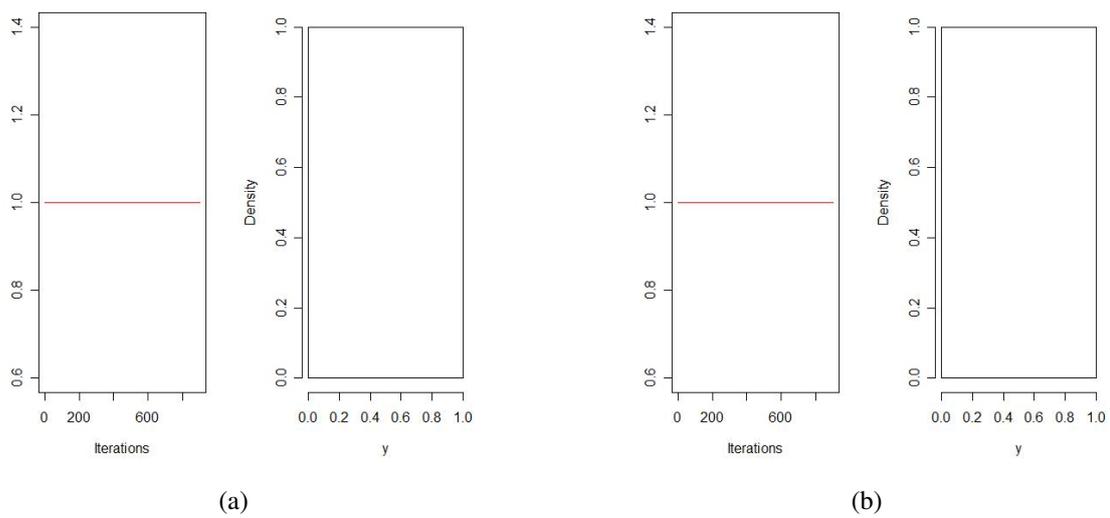
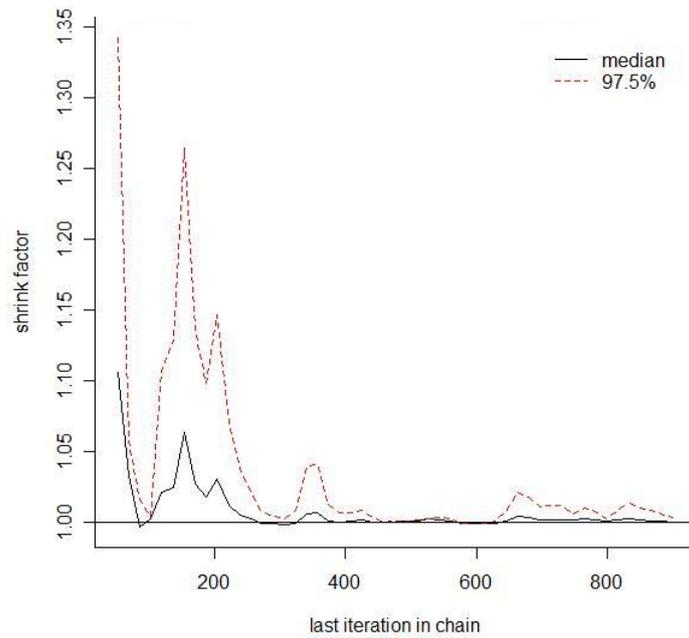
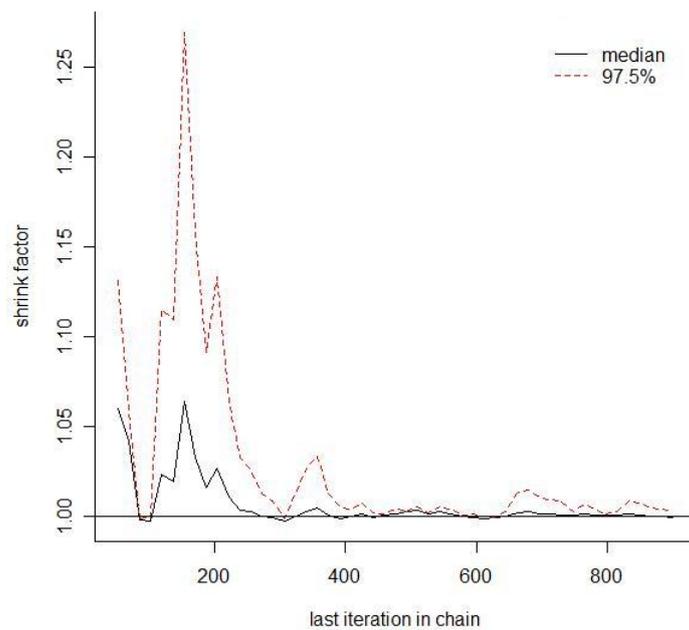


Figura A.1: Distribución *a posteriori* del número de componentes principales para 1995(a) y 2000(b)

A. Apéndice

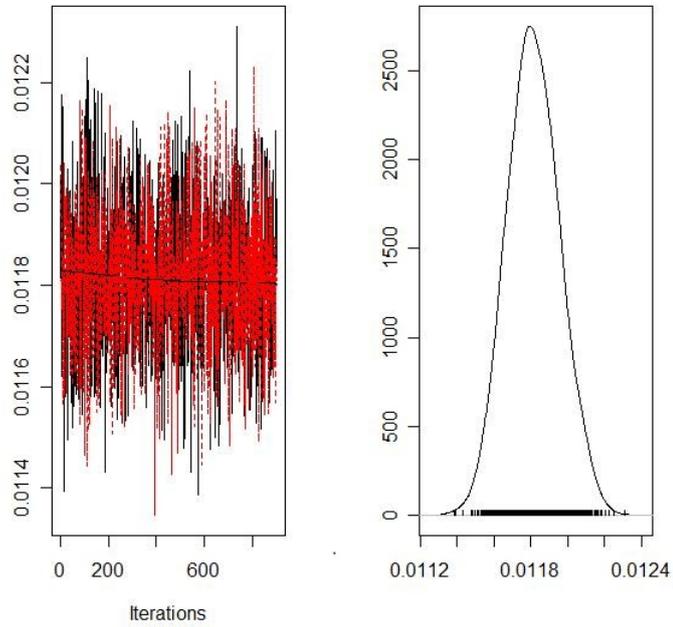


(a) Diagnóstico de convergencia de Gelman y Rubin para σ^2 en 1995

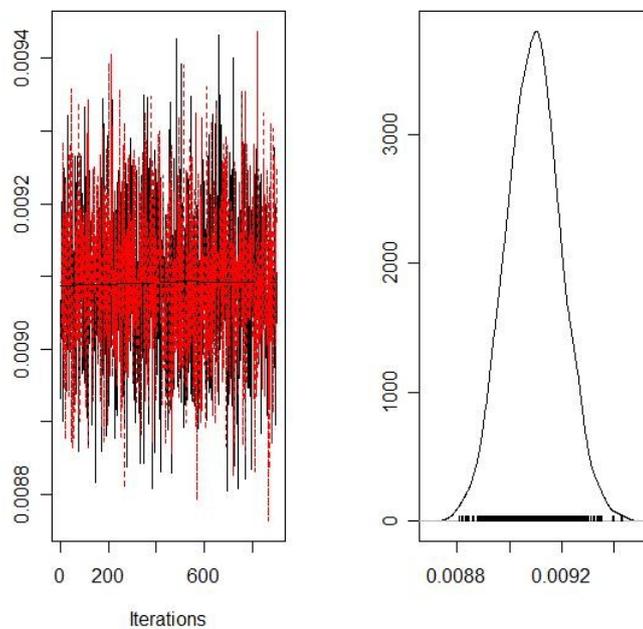


(b) Diagnóstico de convergencia de Gelman y Rubin para σ^2 en el 2000

Figura A.2: Diagnóstico de convergencia de Gelman y Rudin para σ^2 para 1995 y 2000



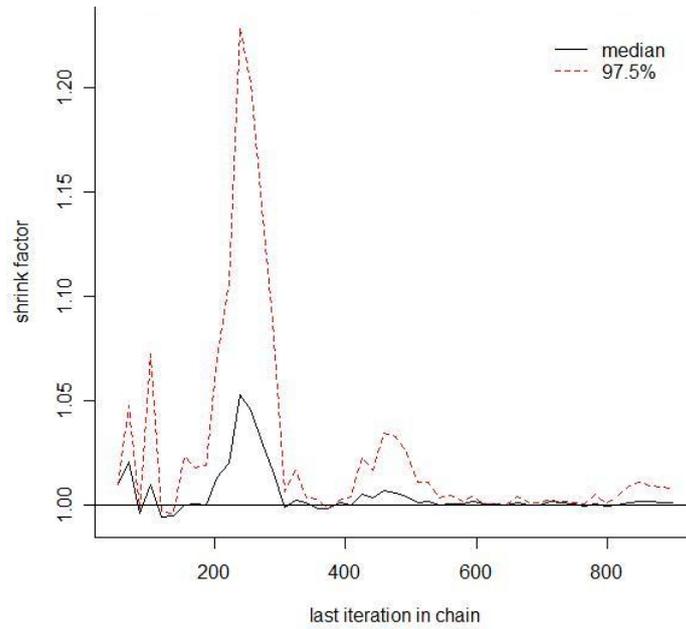
(a) Gráfica de la densidad estimada de σ^2 para 1995



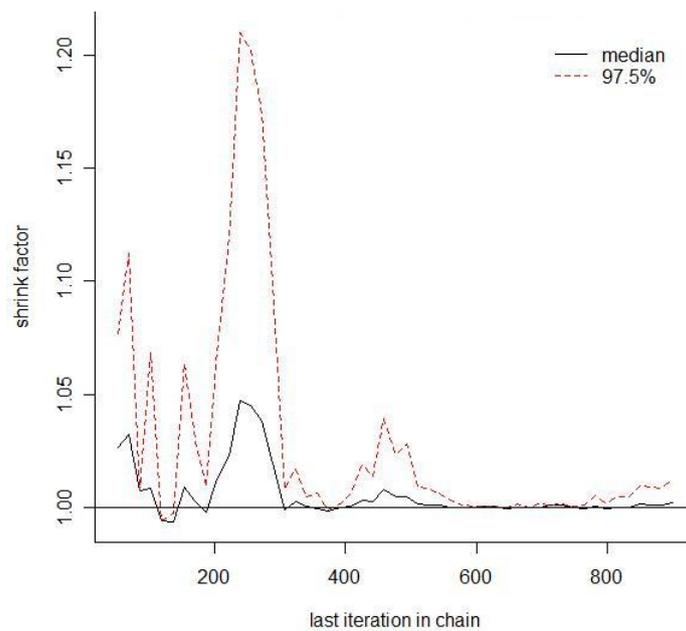
(b) Gráfica de la densidad estimada de σ^2 para el 2000

Figura A.3: Gráfica de la densidad estimada de σ^2 para 1995 y 2000

A. Apéndice



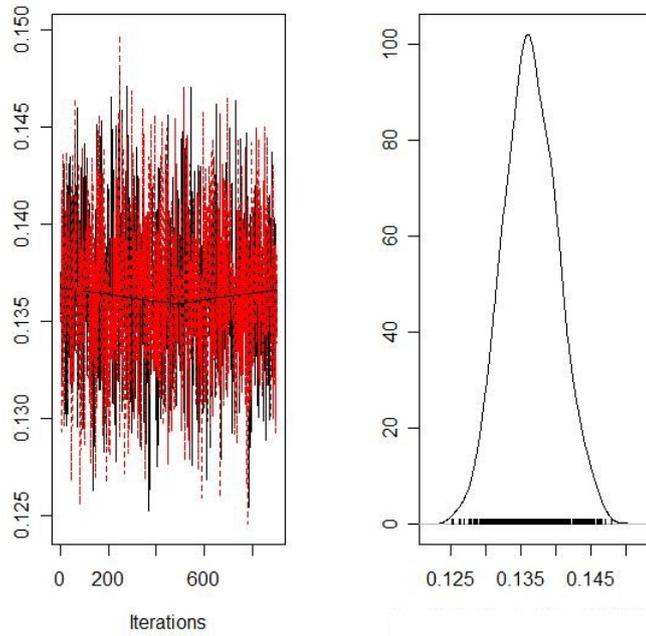
(a) Diagnóstico de convergencia de Gelman y Rubin para l_1 en 1995



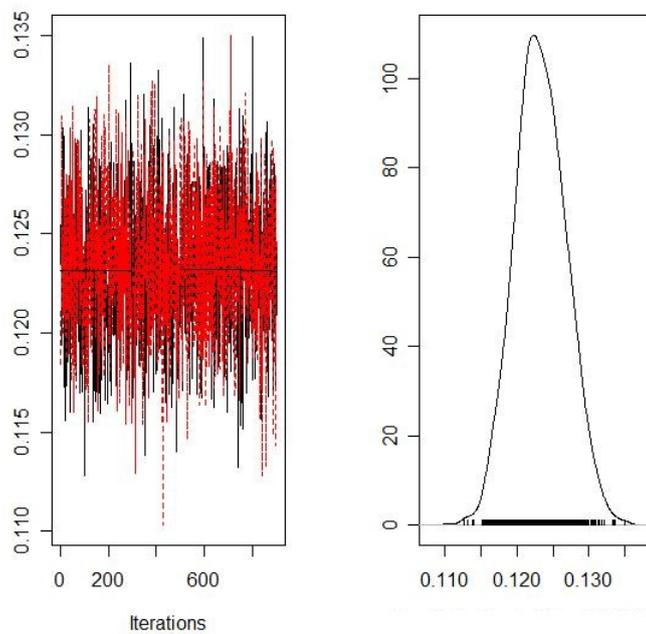
(b) Diagnóstico de convergencia de Gelman y Rubin para l_1 en el 2000

Figura A.4: Diagnóstico de convergencia de Gelman y Rubin para l_1 en 1995 y 2000

A. Apéndice



(a) Gráfica de la densidad estimada de l_1 para 1995



(b) Gráfica de la densidad estimada de l_1 para el 2000

Figura A.5: Gráfica de la densidad estimada de l_1 para 1995 y 2000

Apéndice

```
#####  
#Se crea todo el algoritmo como una función R solo se  
#requiere saber el número de iteraciones a realizar.  
#####  
resu=function(ite){  
#####  
# Inicialización:  
#obtener una muestra ( $q, sig^{-2}, l1^{-1}, \dots, lq^{-1}, tao$ ) de la a priori  
#####  
#Valores iniciales  
#Se acomoda la base de datos en una matrix, estos datos provienen  
#del URL que esta entre comillas.  
datos<-as.matrix(read.table("C:\\... \\datos1995.txt",header=TRUE,  
sep=" " ))  
S<-var(datos) #matriz de varianzas y covarianzas  
N<-nrow(datos) #número de datos  
di<-ncol(datos) #número de variables  
r<-3 #parámetro de las distribuciones Gamma de li y sig
```

B. Apéndice

```
alpha<-0.5 #parámetro de las distribuciones Gamma de tao
V<-sqrt(diag(S)) #desviación estándar de los datos
nu<-1.2/V #parametro de las distribuciones Gamma de tao
eg<-eigen(S) #eigenvalores de S
g<-eg$values
g

#####
#Valores de la muestra inicial
d1<-di-1
d2<-di-2
q<-round(runif(1,2,d1)) #número de CP a utilizar q<d
q
tao<-rgamma(1,alpha,nu) #parámetro de las distribuciones Gamma de li y
sig
sig<-rgamma(1,r,tao) #la varianza acumulada usando CP
li<-matrix(NA,q,1)
for (i in 1:q){
li[i]<-rgamma(1,r,tao) #CP
}

#####
#Se usa en el movimiento de nacimiento y muerte
b<-matrix(NA,d1,1) #b y d son las probabilidades de intento
d<-matrix(NA,d1,1) #de muerte y nacimiento respectivamente
b[1]<-1
d[1]<-0
b[d1]<-0
d[d1]<-1
```

B. Apéndice

```
for (i in 2:d2){
b[i]<-0.5
d[i]<-0.5 }

#####
#Aquí debe comenzar a repetir
#Actualiza los parámetros e hiperparámetros usando Gibbs sampling
ite<-100000
qf<-matrix(NA,ite,1)
lif<-matrix(NA,di,ite)
sigf<-matrix(NA,ite,1)

for (m in 1:ite){

if (q==1){
a<-(N/2)+r
z<-((N*g[1])/2)+tao
li[1]<-rgamma(1,a,z)
}

if (q!=1){
for (j in 1:q){
a<-(N/2)+r
z<-((N*g[j])/2)+tao
li[j]<-rgamma(1,a,z)

if (j==1){
while (li[2]<li[1]){
li[1]<-rgamma(1,a,z)
}
}
```

B. Apéndice

```
}

if (j!=1 & j!=q){
while(li[j]<li[j-1] & li[j+1]<li[j]){
li[j]<-rgamma(1,a,z)
}
}

if (j==q){
while (li[q]<li[q-1]){
li[q]<-rgamma(1,a,z)
}
}

}
}

sumg<-0
w<-q+1
for (j in w:di){
sumg<-sumg+g[j]
}

a<-((N*(di-q))/2)+r
z<-((N*sumg)/2)+tao
sig<-rgamma(1,a,z)
while(sig<li[q]){
sig<-rgamma(1,a,z)
}

suml<-0
```

B. Apéndice

```
for (j in 1:q){
suml<-suml+li[j]
}
a<-(q+1)*r+alpha
z<-suml+sig+nu
tao<-rgamma(1,a,z)

#####
#Se construye una variable aleatoria uniforme
u<-runif(1,0,1)

#####
#Si u<=bq, entonces decidimos nacimineto o muerte de la observación
if (u<=b[q]) {
lip<-rgamma(1,r,tao)
while(lip<li[q] & sig<lip){lip<-rgamma(1,r,tao)}
R<-(taor / gamma(r)) * lip(r-1) * exp(-1 * tao * lip)
a<-min(1,R)
v<-runif(1,0,1)
if (v<a) {
qm1<-q+1
li<-matrix(c(li,lip),qm1,1)
q<-nrow(li)}
}

if (q! = 1) {
if (u<=b[q]+d[q]){
R<-(taor / gamma(r)) * li[q](r-1) * exp(-tao * li[q])
a<-min(1,1/R)
v<-runif(1,0,1)
```

B. Apéndice

```
if (v<a) {q1<-q-1
lin<-matrix(NA,q1,1)
for (i in 1:q1){lin[i]<-li[i]}
li<-matrix(NA,q1,1)
li<-lin
q<-nrow(li)}
}
}

qf[m,1]<-q
sigf[m,1]<-sig

for (k in 1:di){
if (q<k){lif[k,m]<-NA} else {lif[k,m]<-li[k]}
}

}

qf<-1/qf[,1]
sf<-1/sigf[,1]
lf<-1/lif[1,] return(qf,sf,lf)
}

#####
#Se llama a la función para generar la primera cadena de resultados
#y se les asigna un nombre ya como una cadena de Markov

primera=resu(400)
lfmcmc<-mcmc(primera$lf[5:400])
sfmcmc<-mcmc(primera$sf[5:400])
qfmcmc<-mcmc(primera$qf[5:400])
```

B. Apéndice

```
#####  
#para el diagnóstico de convergencia debemos generar otra cadena  
#llamada lf1mcmc para después analizar con coda  
  
segunda=resu(400)  
lf1mcmc<-mcmc(segunda$lf[5:400])  
sf1mcmc<-mcmc(segunda$sf[5:400])  
qf1mcmc<-mcmc(segunda$qf[5:400])  
  
one<-mcmc.list(lf1mcmc,lf1mcmc)  
#se juntan ambas cadenas en una lista para poder realizar los diagnósticos  
summary(one, quantiles = c(0.025, 0.25, 0.5, 0.75, 0.975))  
#Resumen de estadísticas e intervalos creíbles  
plot(one, trace=TRUE, density=TRUE, smooth=TRUE,auto.layout=TRUE)  
#Gráfica de la densidad a posteriori  
raftery.diag(one, q=0.5, r=0.05, s=0.95, converge.eps=0.001)  
#Diagnóstico de Raftery y Lewis  
gelman.diag(one,confidence=0.95, transform=FALSE, autoburnin=TRUE)  
#Diagnóstico de Gelman y Rudin  
gelman.plot(one, bin.width=10, max.bins=50,confidence=0.95, transform=FALSE,  
auto.layout=TRUE)  
#Gráfica del diagnóstico de convergencia de Gelman y Rudin  
  
two<-mcmc.list(sfmcmc,sf1mcmc)  
summary(two, quantiles = c(0.025, 0.25, 0.5, 0.75, 0.975))  
#Resumen de estadísticas e intervalos creíbles  
plot(two, trace=TRUE, density=TRUE, smooth=TRUE,auto.layout=TRUE)  
#Gráfica de la densidad a posteriori  
raftery.diag(two, q=0.5, r=0.05, s=0.95, converge.eps=0.001)  
#Diagnóstico de Raftery y Lewis
```

B. Apéndice

```
gelman.diag(two,confidence=0.95, transform=FALSE, autoburnin=TRUE)
#Diagnóstico de Gelman y Rudin
gelman.plot(two, bin.width=10, max.bins=50,confidence=0.95, transform=FALSE,
auto.layout=TRUE)
#Gráfica del diagnóstico de convergencia de Gelman y Rudin

three<-mcmc.list(qfmcmc,qf1mcmc)
summary(three, quantiles = c(0.025, 0.25, 0.5, 0.75, 0.975))
#Resumen de estadísticas e intervalos creíbles
plot(three, trace=TRUE, density=TRUE, smooth=TRUE,auto.layout=TRUE)
#Gráfica de la densidad a posteriori
```