Taller de estimación de variables latentes

Eric Magar
Guillermo Rosas
y el amable apoyo de Ernesto Barrios

ITAM – Junio 2011

Temas

Introducción

- Variables latentes
- Inferencia bayesiana
- R, RStudio y Jags (Barrios)

Regresión lineal: El efecto de arrastre de ejecutivos

- Estimación clásica
- Estimación bayesiana

IRT I: Introducción, teoría espacial del voto, CDHCU

IRT II: Modelos jerárquicos, dinámicos, multidimensionales

Modelos con variables latentes: análisis factorial bayesiano

Temas

Introducción
- Variables latentes
- Inferencia bayesiana
- R, RStudio y Jags (Barrios)

Regresión lineal: El efecto de arrastre de ejecutivos
- Estimación clásica
- Estimación bayesiana

IRT I: Introducción, teoría espacial del voto, CDHCU

IRT II: Modelos jerárquicos, dinámicos, multidimensionales

Modelos con variables latentes: análisis factorial bayesiano

Variables latentes y su potencial gráfico

1. A menudo interesa medición de variables que no pueden ser observadas directamente
2. Deben ser inferidas de indicadores que sí son observables
3. Un *modelo de variable latente* conecta (2) a (1)

Ejemplos

- Sicometría: habilidad, personalidad autoritaria
- Mercadotecnia: demanda de productos, *herd behavior*
- Economía: confianza del consumidor, felicidad
- CPol: ideología, preferencias en regulación

La estructura de los datos

Poole (2005)

Un procedimiento capaz de capturar la estructura subyacente de tus datos (votos) permite

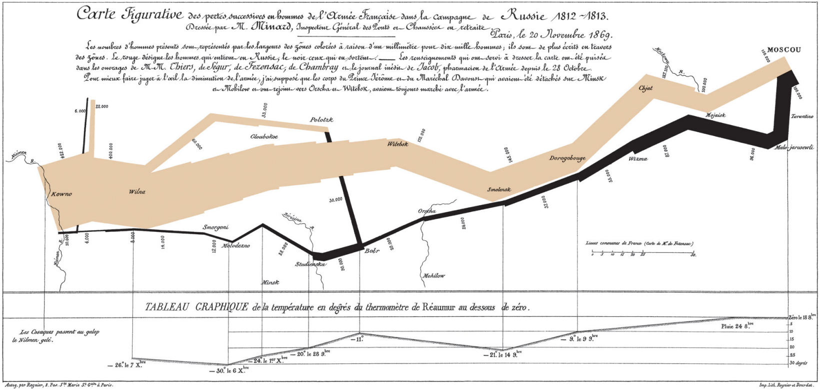
1. tomar una serie de números
2. transformarlos en gráficas simples
3. que comunican significado

Resumes información compleja con dibujos

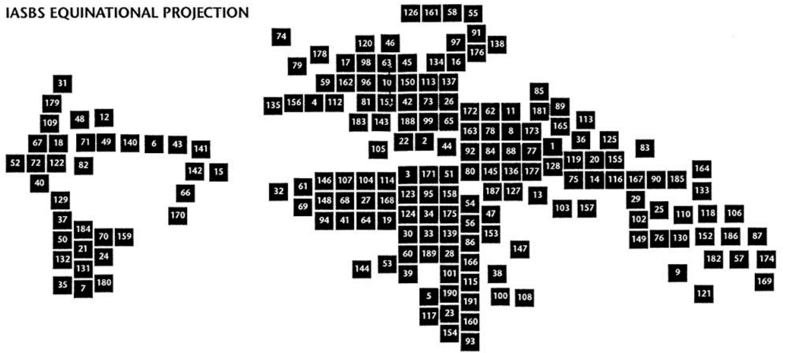
Aproximación gráfica

Un bon croquis vaut mieux qu'un long discours
—(atribuido a) Napoleón Bonaparte

Un bon croquis vaut mieux qu'un long discours
—(atribuido a) Napoleón Bonaparte



Aproximación gráfica

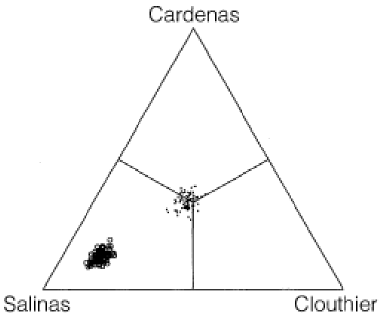


1 afghanistan
2 albania
3 algeria
4 andorra
5 angola
6 antigua + barbuda
7 argentina
8 armenia
9 australia
10 austria
11 azerbaijan
12 bahamas
13 bahrain
14 bangladesh
15 barbados
16 belarus
17 belgium
18 belize
19 benin
20 bhutan
21 bolivia
22 bosnia-herzegovina
23 botswana
24 brazil
25 brunei
26 bulgaria
27 burkina faso
28 burundi
29 cambodia
30 cameroon
31 canada
32 cape verde
33 central african republic
34 chad
35 chile
36 china
37 colombia
38 comoros
39 congo
40 costa rica
41 côte d'ivoire
42 croatia
43 cuba
44 cyprus
45 czech republic
46 denmark
47 djibouti
48 dominica
49 dominican republic
50 ecuador
51 egypt
52 el salvador
53 equatorial guinea
54 eritrea
55 estonia
56 ethiopia
57 fiji
58 finland
59 france
60 gabon
61 gambia
62 georgia
63 germany
64 ghana
65 greece
66 grenada
67 guatemala
68 guinea
69 guinea-bissau
70 guyana
71 haiti
72 honduras
73 hungary
74 iceland
75 india
76 indonesia
77 iran
78 iraq
79 ireland
80 israel
81 italy
82 jamaica
83 japan
84 jordan
85 kazakhstan
86 kenya
87 kiribati
88 kuwait
89 kyrgyzstan
90 laos
91 latvia
92 lebanon
93 lesotho
94 liberia
95 libya
96 liechtenstein
97 lithuania
98 luxembourg
99 macedonia
100 madagascar
101 malawi
102 malaysia
103 maldives
104 mali
105 malta
106 marshall islands
107 mauritania
108 mauritius
109 mexico
110 micronesia
111 moldova
112 monaco
113 mongolia
114 morocco
115 mozambique
116 myanmar
117 namibia
118 nauru
119 nepal
120 netherlands
121 new zealand
122 nicaragua
123 niger
124 nigeria
125 north korea
126 norway
127 oman
128 pakistan
129 panama
130 papua new guinea
131 paraguay
132 peru
133 philippines
134 poland
135 portugal
136 qatar
137 romania
138 russia
139 rwanda
140 st. kitts + nevis
141 st. lucia
142 st. vincent + the grenadines
143 san marino
144 sao tome + principe
145 saudi arabia
146 senegal
147 seychelles
148 sierra leone
149 singapore
150 slovakia
151 slovenia
152 solomon islands
153 somalia
154 south africa
155 south korea
156 spain
157 sri lanka
158 sudan
159 suriname
160 swaziland
161 sweden
162 switzerland
163 syria
164 taiwan
165 tajikistan
166 tanzania
167 thailand
168 togo
169 tonga
170 trinidad + tobago
171 tunisia
172 turkey
173 turkmenistan
174 tuvalu
175 uganda
176 ukraine
177 united arab emirate
178 united kingdom
179 united states
180 uruguay
181 uzbekistan
182 vanuatu
183 vatican city
184 venezuela
185 viet nam
186 western samoa
187 yemen
188 yugoslavia
189 zaire
190 zambia
191 zimbabwe

Globehead!
Journal of extreme geography
Vol. 1 Thing 1 1994

Aproximación gráfica

FIGURE 3 Simulated Electoral Outcomes



Coordinates in this ternary diagram are predicted fractions of the vote received by each of the three candidates. Each point is an election outcome drawn randomly from a world in which all voters believe Salinas' PRI party is strengthing (for the "o"'s in the bottom left) or weakening (for the "·"'s in the middle), with other variables held constant at their means.

Simulaciones con el logit multinomial de Domínguez y McCann (King et al. 2000).

Aproximación gráfica

Poole recurre a la "basic-space theory of ideology" para resumir/interpretar votaciones nominales con dibujos

1. Teoría espacial del voto (cf. Downs)
2. Sistema de creencias restrinje la dimensionalidad (cf. Converse)

Escalamiento multidimensional

1. Métodos agnósticos: factorial, componentes principales, clasificación óptima, ...
2. Microfundamentos dan la estructura: item response theory, utilidad euclidiana, ...

Ventajas de la estadística bayesiana

1. La inferencia bayesiana es simple y directa: Depende de enunciados *a posteriori* que reflejan nuestras creencias acerca de parámetros, hipótesis, modelos o datos faltantes
2. La inferencia bayesiana permite responder preguntas relevantes: ¿qué tan plausible es la hipótesis H_0 en vista de los datos que tenemos?
3. La inferencia bayesiana permite el análisis de eventos no repetibles
4. La inferencia bayesiana facilita el análisis de relaciones causa-efecto en situaciones de heterogeneidad causal (modelos jerárquicos)

Si Ω es un conjunto de eventos y $\Pr(A)$ es una función que asigna un número real a cada evento $A \subset \Omega$, $\Pr(A)$ es una medida de probabilidad si:

1. $\Pr(A) \geq 0, \quad \forall A \subset \Omega$
2. $\Pr(\Omega) = 1$
3. Si $A \subset \Omega$ y $B \subset \Omega$ son conjuntos inconexos, $\Pr(A \cup B) = \Pr(A) + \Pr(B)$

¿Qué es probabilidad?

Visiones alternativas

- Visión "frecuentista" (clásica): La probabilidad es la frecuencia relativa de un evento en el largo plazo, y es una propiedad del objeto que se estudia
- Visión "subjetivista" (bayesiana): La probabilidad es una expresión del grado de certeza que uno tiene acerca de la veracidad de cierta hipótesis

Teorema de Bayes (regla de probabilidad inversa)

Si H es una hipótesis, y E representa evidencia, la respuesta a la pregunta sobre la plausibilidad de la hipótesis H dada la evidencia E es:

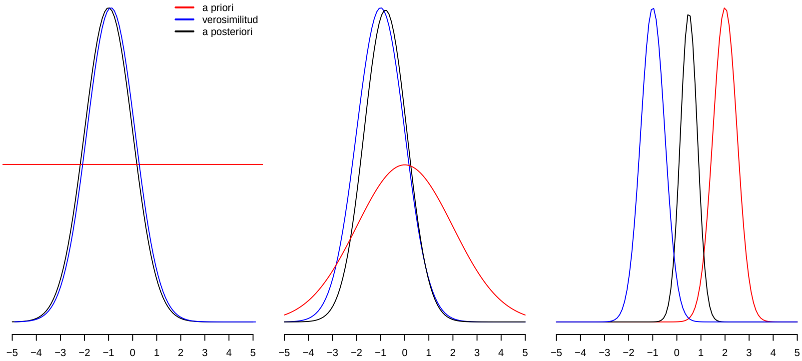
$$\underbrace{\Pr(H|E)}_{\text{a posteriori}} = \frac{\Pr(E \cap H)}{\Pr(E)} = \frac{\overbrace{\Pr(E|H)}^{\text{verosimilitud}}\ \overbrace{\Pr(H)}^{\text{a priori}}}{\Pr(E)}$$

De manera más general,

$$\Pr(\psi|\mathbf{y}) \propto \Pr(\mathbf{y}|\psi)\Pr(\psi),$$

es decir, la distribución *a posteriori* es proporcional al producto de la distribución *a priori* y la función de verosimilitud

densidad a priori $\times$ verosimilitud $=$ densidad a posteriori



Creencias a priori, creencias a posteriori, y probabilidad

$$\text{a priori} \rightarrow \text{datos} \rightarrow \text{a posteriori}$$

$$p(\psi) \rightarrow y \rightarrow p(\psi|y)$$

Inferencia bayesiana: El enunciado condicional tiene que ver con la probabilidad de que el parámetro esté en cierto rango, dado que observamos ciertos datos

Inferencia frecuentista: El enunciado condicional tiene que ver con la probabilidad de observar ciertos datos, dado que suponemos la veracidad de la hipótesis nula

Chabelo es bayesiano, los cuates de Chabelo no

- El niño abre la puerta A:

$$\Pr(A) = \Pr(B) = \Pr(C) = \frac{1}{3}$$

- Chabelo abre la puerta C:

$$\Pr(A) = \Pr(B) = \frac{1}{2}$$

- El niño se queda con la puerta A

- Información disponible:

$$\begin{aligned} \Pr(A) = \Pr(B) = \Pr(C) &= 1/3 \\ \Pr(\text{Ch abra } C|A) &= 1/2 \\ \Pr(\text{Ch abra } C|B) &= 1 \\ \Pr(\text{Ch abra } C|C) &= 0 \end{aligned}$$

- Probabilidad de que Chabelo abra C:

$$\begin{aligned} \Pr(\text{Ch abra } C) = &\Pr(A)\Pr(\text{Ch abra } C|A) \\ &+ \Pr(B)\Pr(\text{Ch abra } C|B) \\ &+ \Pr(C)\Pr(\text{Ch abra } C|C) \end{aligned}$$

- Teorema de Bayes:

$$\Pr(A|\text{Ch abra } C) = \frac{\Pr(A)\Pr(\text{Ch abra } C|A)}{\Pr(\text{Ch abra } C)}$$

$$\Pr(B|\text{Ch abra } C) = \frac{\Pr(B)\Pr(\text{Ch abra } C|B)}{\Pr(\text{Ch abra } C)}$$

¿Qué es R?

- Es una colección de herramientas para manipular datos, llevar a cabo cómputos y preparar gráficas
- Numerosas técnicas estadísticas han sido desarrolladas:
  - algunas pre-empaquetadas en la instalación de R
  - otras (las más) como paquetes extensibles `http://www.youtube.com/watch?v=ZO-93iULKBI`
- Código de fuente abierta: Encyclopaedia Britannica *v.* Wikipedia
- Muchísima ayuda en línea `http://www.burns-stat.com/`

Algunas complicaciones

- Consola abre con pantalla blanca y ‘>’
  turn-off para tantas generaciones que han crecido con mouse
- R crea objetos que no necesariamente tienen que ser rectangulares. Salir de la camisa de fuerza desconcierta
- Todo es *case sensitive* (funciones y nombres de objeto)
- En vez de sacar output copioso como SPSS, el análisis en R va guardando los resultados como objetos. El usuario los interroga para sacar conclusiones
- R suele operarse a partir de código (tipo *do file* de Stata) que se prepara en un editor de texto — usaremos RStudio para esto

Regresión lineal Bayesiana

COMANDOS EN R

Temas

Introducción
- Variables latentes
- Inferencia bayesiana
- R, RStudio y Jags (Barrios)

Regresión lineal: El efecto de arrastre de ejecutivos
- Estimación clásica
- Estimación bayesiana

IRT I: Introducción, teoría espacial del voto, CDHCU

IRT II: Modelos jerárquicos, dinámicos, multidimensionales

Modelos con variables latentes: análisis factorial bayesiano

Análisis de elecciones legislativas

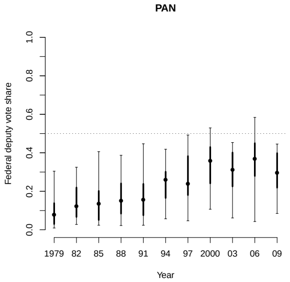
- El desenlace de una elección resulta de la combinación de dos clases de fuerzas (Converse 1966)
    1. largo plazo: encapsuladas en PID, cambio lento, "voto normal"
    2. corto plazo: calidad de candidatos, swings nacionales, shock efímero al voto normal
- Arrastre presidencial es una de las fuerzas de corto plazo
- EE.UU.: 1980s $= \frac{1}{3}$; 1940s $= \frac{1}{2}$; 1890s $= \frac{9}{10}$ (Campbell 1991)
- ¿Elección concurrente de gobernador también produce un efecto de arrastre? (Jones 1997, Samuels 2000)

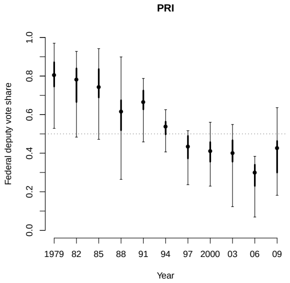
Un modelo de arrastre

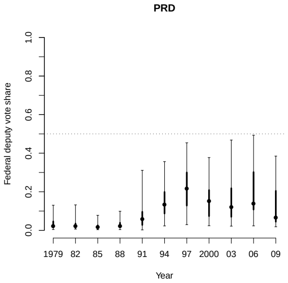
$$
\begin{aligned}
Dvot &= \beta_0 + \beta_1 RecentDvot + \beta_2 GovOnlyConcurs \\
&\quad + \beta_3 PresOnlyConcurs + \beta_4 Gov\&PresConcur \\
&\quad + \color{red}{\beta_5} Gvote|GovOnlyConcurs + \beta_6 Gvote|Gov\&PresConcur \\
&\quad + \color{red}{\beta_7} Pvote|PresOnlyConcurs + \beta_8 Pvote|Gov\&PresConcur \\
&\quad + \text{controles} + \text{error}
\end{aligned}
$$

- Elecciones de diputados federales 1997–2009
- Unidad de análisis: estados
- Panel 32×5

Elecciones de diputados federales por estado







Regresión en Stata

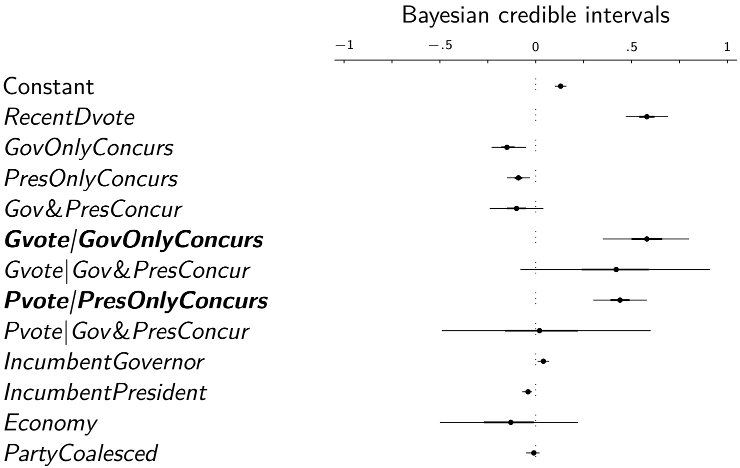
Resultados OLS

Variable	OLS $\hat{\beta}$	OLS p	OLS-PCSE $\hat{\beta}$	OLS-PCSE p
Part A. PAN				
Constant	.135	.000	.135	.000
RecentDvote	.586	.000	.586	.000
GovOnlyConcurs	−.151	.000	−.151	.005
PresOnlyConcurs	−.097	.000	−.097	.054
Gov&PresConcur	−.106	.000	−.106	.102
Gvote|GovOnlyConcurs	.584	.000	.584	.000
Gvote|Gov&PresConcur	.444	.000	.444	.043
Pvote|PresOnlyConcurs	.446	.000	.446	.002
Pvote|Gov&PresConcur	.018	.848	.018	.956
IncumbentGovernor	.045	.030	.045	.021
IncumbentPresident	−.047	.002	−.047	.004
Economy	−.145	.451	−.145	.465
PartyCoalesced	.016	.319	.016	.416
N	160		160	
R^2	.82		.82	

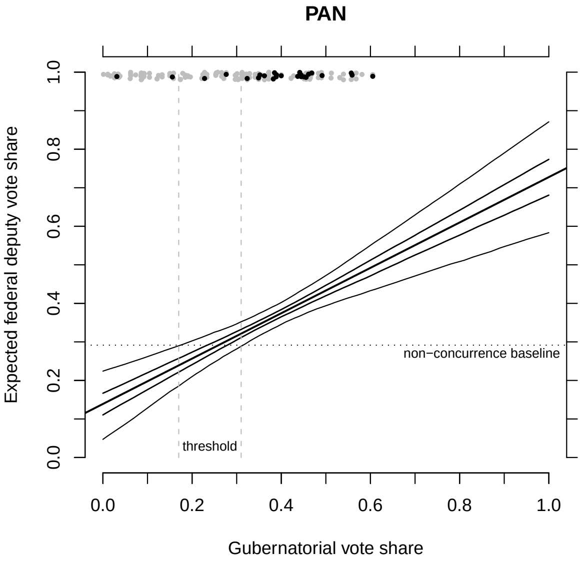
Regresión en R: función `lm`

Regresión bayesiana en R: función `MCMCregress`

Resultados MCMC



El efecto de arrastre neto de gobernadores



¿Resultados idénticos?

Coeficiente de `RecentDvote`

1. Versión frecuentista:

$$\hat{\beta}_{ML} \approx 0.59, \quad \mathrm{se}(\hat{\beta}) \approx 0.06, \quad H_0 = 0.45$$

$$z = \frac{\hat{\beta}_{ML} - H_0}{\mathrm{se}(\hat{\beta})} \approx 2.33$$

La frecuencia con que uno observaría $\hat{\beta}_{ML} \geq 0.59$ bajo H_0 en el largo plazo es 0.009.

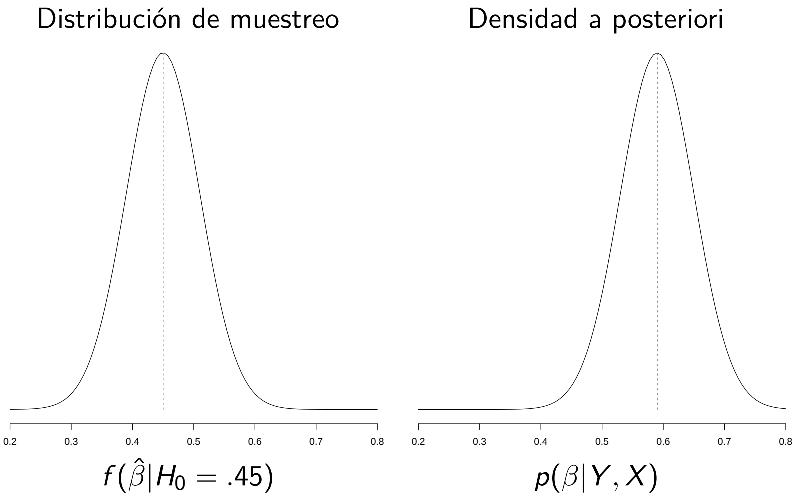
2. Versión bayesiana:

$$p(\beta) \sim \mathcal{N}(0, 100^2); \quad p(\beta|Y, X) \sim \mathcal{N}(0.59, 0.06^2)$$

$$\begin{aligned}
\Pr(\beta \leq 0.45|Y, X) &= \int_{-\infty}^{0.45} p(\beta|Y, X) d\beta \\
&= \int_{-\infty}^{0.45} \phi\left(\frac{0.59 - .45}{0.06}\right) d\beta = .009
\end{aligned}$$

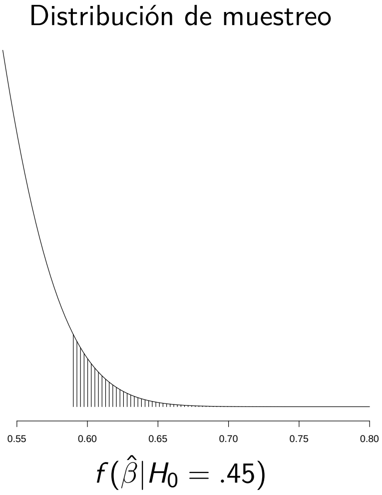
¿Resultados idénticos?

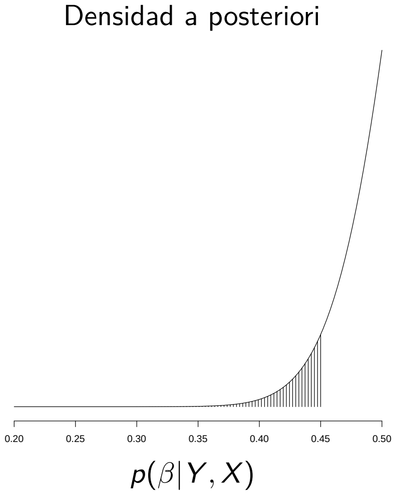
Coeficiente de `RecentDvote`



¿Resultados idénticos?

Coeficiente de `RecentDvote`





¿Resultados idénticos?

Coeficiente de `RecentDvote`

$$H_1 : \beta > .45$$
$$H_0 : \beta \leq .45$$

El factor bayesiano que resume la cantidad de evidencia a favor de H_1 es:

$$\begin{aligned} \mathrm{B}_{10} &= \left\{ \frac{p(H_1|y)}{p(H_0|y)} \right\} \Big/ \left\{ \frac{p(H_1)}{p(H_0)} \right\} \\ &= \left\{ \frac{1 - .009}{.009} \right\} \Big/ \left\{ \frac{.5}{.5} \right\} \\ &= 110.11 \end{aligned}$$

¿Cómo podemos estimar la probabilidad de que el coeficiente de `RecentDVote` sea negativo?
¿Cómo podemos estimar la probabilidad de que el coeficiente de `RecentDVote` se encuentre entre 0.3 y 0.45?

La tentación de la probabilidad inversa en el frecuentismo

1. Modus tollens

$$P \rightarrow Q; \neg Q; \vdash \neg P$$

2. Modus tollens probabilístico?
   - Si un individuo es mexicano, es *improbable* que sea miembro de la Cámara de Diputados
   - Este individuo es miembro de la Cámara
   - Es *improbable* que este individuo sea mexicano

3. Inverse probability problem

$$\underbrace{\Pr(H_0|D)}_{\text{mexicano}|\text{diputado}} \propto \Pr(H_0) \times \underbrace{\Pr(D|H_0)}_{\text{diputado}|\text{mexicano}}$$

Regresión lineal: densidades

- Modelo de regresión lineal:

$$y_i \sim \mathcal{N}(\mu_i,\ \sigma^2)$$

$$\mu_i = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k = \mathbf{x}_i \boldsymbol{\beta}$$

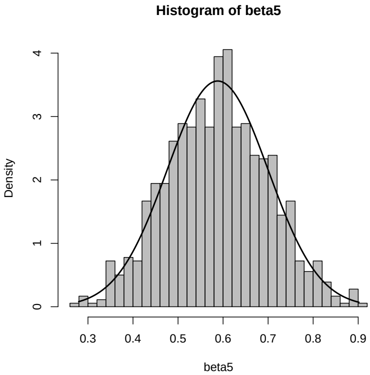
- La inferencia deseada es sobre $\boldsymbol{\beta} = [\beta_0 \ldots \beta_k]$ y σ^2
- Densidades a priori:

$$\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\mathbf{b_0}, \sigma^2 \mathbf{B_0})$$

$$\sigma^2 \sim \text{Gamma-inversa}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

Densidades conjugadas

- Cuando las densidades a priori y a posteriori son de la misma familia, se dice que la densidad a priori y la verosimilitud están *conjugadas*
- Si uno es capaz de proveer un a priori no tendrá dificultad en interpretar el a posteriori conjugado
- $\beta_5 \sim \mathcal{N}(0, 100^2)$ — a priori no informativo



Densidades conjugadas

Sea $y_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ (donde $i = 1, \ldots, n$), con σ^2 conocida y $\mathbf{y} = (y_1, \ldots, y_n)\prime$. Si $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$ es la distribución a priori de μ, la distribución a posteriori es:

$$\mu|\mathbf{y} \sim \mathcal{N}\left(\frac{\mu_0 \sigma_0^{-2} + \bar{y}\frac{n}{\sigma^2}}{\sigma_0^{-2} + \frac{n}{\sigma^2}}, \left(\sigma_0^{-2} + \frac{n}{\sigma^2}\right)^{-1}\right)$$

Si la distribución a priori y la verosimilitud son conjugadas, la densidad a posteriori es una mezcla ponderada de ambas densidades, donde el ponderador es la precisión respectiva de ambas densidades:

$$\mathrm{E}(\mu|\mathbf{y}) = \bar{y} + \omega(\mu_0 - \bar{y})$$

$$\omega = \frac{\sigma_0^{-2}}{\sigma_0^{-2} + \frac{n}{\sigma^2}}$$

- Distribución a posteriori

$$\beta|\sigma^2, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\mathbf{b_1}, \sigma^2\mathbf{B_1})$$

$$\sigma^2 \sim \text{inverse Gamma}\left(\frac{\nu_1}{2}, \frac{\nu_1\sigma_1^2}{2}\right)$$

- Los parámetros a posteriori son un "compromiso" entre la información a priori y la información proporcionada por los datos:

$$\mathbf{b_1} = (\mathbf{B_0}^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{B_0}^{-1}\mathbf{b_0} + \mathbf{X}'\mathbf{X}\hat{\theta})$$
$$\mathbf{B_1} = (\mathbf{B_0}^{-1} + \mathbf{X}'\mathbf{X})^{-1}$$

MCMC: principio de Monte Carlo y cadenas de Markov

- Principio de Monte Carlo: Todo lo que querramos saber acerca de una variable ψ lo podemos averiguar tomando muchas muestras de la densidad $f(\psi)$ de ψ

- Cadenas de Markov: Una cadena de Markov es un proceso estocástico definido sobre cierto espacio paramétrico. Tiene la propiedad de ser un proceso *ergódico*: una cadena de Markov visita cada región del espacio paramétrico en proporción a la probabilidad de esa región dada una densidad de interés

A priori, verosimilitud, a posteriori

Se puede aproximar una densidad a posteriori sin recurrir al análisis matemático

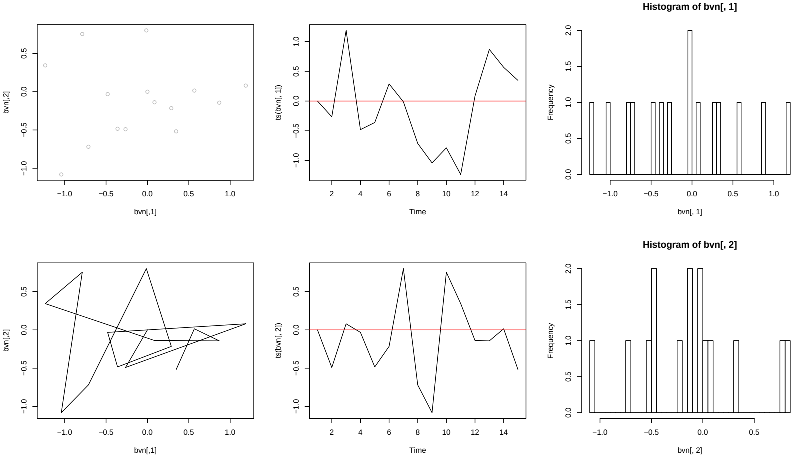
```
library(LearnBayes)

midpt = seq(0.05, 0.95, by = 0.1)
prior = c(1, 5.2, 8, 7.2, 4.6, 2.1, 0.7, 0.1, 0, 0)
prior = prior/sum(prior)

curve(histprior(x,midpt,prior), from=0, to=1,
  ylab="Prior density",ylim=c(0,round (max(prior),2)))

s = 11
f = 16

curve(dbeta(x,s,f), from=0, to=1, ylab="Data likelihood")

curve(histprior(x,midpt,prior) * dbeta(x,s,f),
  from=0, to=1, ylab="Posterior density")
```

Algoritmo de Gibbs

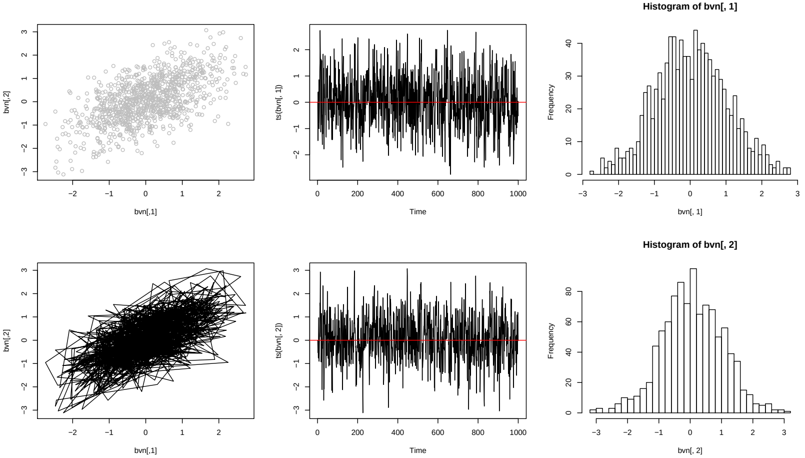
El algoritmo de muestreo de Gibbs (Gibbs sampler) obtiene muestras de las *densidades condicionales* de los parámetros para caracterizar la *densidad conjunta* de los parámetros:

```
gibbs<-function (n, rho)   {
        # Open plotting window
        plot (c(-3.5,3.5),c(-3.5,3.5), type="n", xlab="x", ylab="y")
        mat <- matrix(ncol = 2, nrow = n)
        # Assign starting values
        x <- 0
        y <- 0
        mat[1, ] <- c(x, y)
        points (xy.coords(x,y), pch="1", col="black")
        for (i in 2:n) {
                # Sample x2 from p(x|y1)
                x <- rnorm(1, rho * y, sqrt(1 - rho^2))
                # Sample y2 from p(y|x2)
                y <- rnorm(1, rho * x, sqrt(1 - rho^2))
                mat[i, ] <- c(x, y)
                points (xy.coords(x,y), pch=as.character(i),
                     col="grey")
        }
        mat
}
```

Algoritmo de Gibbs: Después de 15 muestras



Algoritmo de Gibbs: Después de 1,000 muestras



Algoritmo de Gibbs: Comparación con resultados analíticos

Jackman 2009, tabla 5.3

	Analítico	1000 muestras	50000 muestras
$E(\mu_1 \mid \mathbf{y})$	29.44	30.34	29.38
$E(\mu_2 \mid \mathbf{y})$	37.72	38.63	37.65
$V(\mu_1 \mid \mathbf{y})$	38.41	35.54	38.91
$V(\mu_2 \mid \mathbf{y})$	43.17	40.12	43.84
$C(\mu_1, \mu_2 \mid \mathbf{y})$	30.59	28.07	31.12

Temas

Introducción
- Variables latentes
- Inferencia bayesiana
- R, RStudio y Jags (Barrios)

Regresión lineal: El efecto de arrastre de ejecutivos
- Estimación clásica
- Estimación bayesiana

IRT I: Introducción, teoría espacial del voto, CDHCU

IRT II: Modelos jerárquicos, dinámicos, multidimensionales

Modelos con variables latentes: análisis factorial bayesiano

Item Response Theory

- Modelo estándar de éxito/fracaso en pruebas es el logístico de respuesta a items (sicometría, Rasch)
- J sujetos contestan prueba de aptitudes con K preguntas puede expresarse como (Gelman&Hill 2007:314)

$$\Pr(y_{jk} = 1) = \text{logit}^{-1}(\alpha_j - \psi_k)$$

  donde α_j es la habilidad del sujeto j
  y ψ_k la dificultad de la pregunta k
- Si preg. 4 tiene dificultad 1.4 y sujeto 13 habilidad 3.8, la prob. de que conteste correctamente el ítem 4 es $\text{logit}^{-1}(2.4) = .92$
- Trivial generalizarlo para estudiar el voto

La aproximación gráfica

- Mapas políticos de una legislatura son fáciles de hacer
- Cuatro ingredientes:
  1. Datos
  2. Técnica de escalamiento
  3. Modelo de variable latente
  4. Conocimiento sustantivo — política de la legislatura
- Sin #3 o sin #4, el mapa carece de significado

Elaboramos el modelo de variable latente
(ojo: reto real es #4)

Orígenes

El modelo de variable latente que usaremos conjuga avances en sicología, economía, ciencia política, ... (Poole 2005)

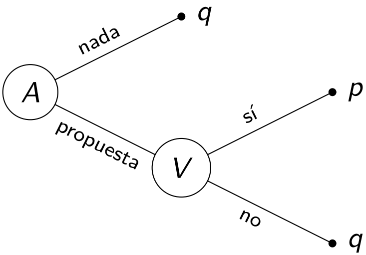
Desarrollo de la versión determinista

1. Hotelling (1929) competencia en pueblo lineal
2. Smithies (1941) demanda elástica
3. Black (1948, 1958) generaliza a comités
4. Downs (1957) generaliza a candidatos, populariza
5. Davis, Hinich, Ordeshook (1966, 1970) completan formalización para aplicaciones empíricas

Supuestos centrales

1. Preferencias y alternativas pueden representarse en un mismo espacio ($\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^n$)
2. Preferencias euclidianas: se opta por la alternativa más cercana al punto ideal ($\pm$ error)
   - $u(x) = ||i, x||$, i es el ideal, x una alternativa
   - opción 1: $u(x) = -|i - x|$
   - opción 2: $u(x) = -(i - x)^2$
   - opción 3: $u(x) = \phi(i - x, \sigma^2)$
3. El voto es sincero (sin *logrolls* ni voto estratégico)

Especificación unidimensional



Cálculo del voto:

- $u(p) > u(q)$ votas sí
- $u(p) \leq u(q)$ votas no

El diferencial de utilidad: $$d = u(p) - u(q) \overset{?}{\lesseqgtr} 0$$

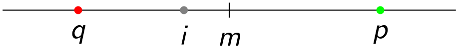
Especificación unidimensional

$p =$ propuesta, $q =$ status quo, $i =$ ideal del votante

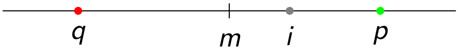
$$
\begin{aligned}
d &= u(p) - u(q) \\
&= -(i-p)^2 + (i-q)^2 \\
&= -(i^2 - 2ip + p^2) + i^2 - 2iq + q^2 \\
&= -2iq + q^2 + 2ip - p^2 \\
&= -2i(q-p) + q^2 - p^2 \\
&= -2i(q-p) + (q+p)(q-p) \\
&= -2(q-p)\left(i - \frac{(q+p)}{2}\right) \\
&= \delta(i-m)
\end{aligned}
$$

La centralidad del punto medio

Basta conocer m para deducir la conducta asociada con i



$$i - m < 0 \leftrightarrow i < m \rightarrow \text{votas no}$$



$$i - m > 0 \leftrightarrow i > m \rightarrow \text{votas sí}$$

Modelo estocástico

Propensión al voto del sujeto j para el ítem k:

$$v_{jk}{}^{*} = \delta_k(i_j - m_k) + \mathrm{error}_{jk}$$

- Regla de votación: $v_{jk} = \begin{cases} 1 & \text{si } v_{jk}{}^{*} > 0 \\ 0 & \text{si } v_{jk}{}^{*} \leq 0 \end{cases}$
- Signo de δ_k controla la polaridad
- Tamaño de $|\delta_k|$ estima el peso relativo del componente sistemático vs. el estocástico (*signal-to-noise ratio*)

Operacionalización

Si $\mathrm{error}_{jk} \overset{\mathrm{iid}}{\sim} \mathcal{N}(0, \sigma^2) \rightarrow v_{jk}{}^{*} \sim \mathcal{N}(\mu_{jk}, \sigma^2)$ donde $\mu_{jk} = \delta_k(i_j - m_k)$

Especificación bidimensional

$$\begin{aligned} d &= \|i,p\| - \|i,q\| \\ &= \sqrt{(x_i - x_p)^2 + (y_i - y_p)^2} - \sqrt{(x_i - x_q)^2 + (y_i - y_q)^2} \end{aligned}$$

Dado que cuando $a, b \geq 0 : a > b \leftrightarrow a^2 > b^2$, raíces se van

$$\begin{aligned} d &= (x_i - x_p)^2 + (y_i - y_p)^2 - [(x_i - x_q)^2 + (y_i - y_q)^2] \\ &= (x_i - x_p)^2 - (x_i - x_q)^2 + (y_i - y_p)^2 - (y_i - y_q)^2 \\ &= [(x_i - x_p) + (x_i - x_q)]\,[(x_i - x_p) - (x_i - x_q)] \\ &\quad + [(y_i - y_p) + (y_i - y_q)]\,[(y_i - y_p) - (y_i - y_q)] \\ &= (2x_i - x_q - x_p)(x_q - x_p) + (2y_i - y_q - y_p)(y_q - y_p) \\ &= 2x_i x_q - 2x_i x_p - {x_q}^2 + {x_p}^2 + 2y_i y_q - 2y_i y_p - {y_q}^2 + {y_p}^2 \end{aligned}$$

Dos posible parameterizaciones

1. Aditiva (Gelman&Hill 2007:319)

$$\begin{aligned} d &= 2x_i x_q - 2x_i x_p - {x_q}^2 + {x_p}^2 + 2y_i y_q - 2y_i y_p - {y_q}^2 + {y_p}^2 \\ &= -2(x_q - x_p)\left(x_i - \frac{(x_q + x_p)}{2}\right) - 2(y_q - y_p)\left(y_i - \frac{(y_q + y_p)}{2}\right) \\ &= \delta_x(x_i - m_x) + \delta_y(y_i - m_y) \end{aligned}$$

- Ésta es la que usaremos en la estimación
- Permite controlar polaridad de cada dimensión fácilmente

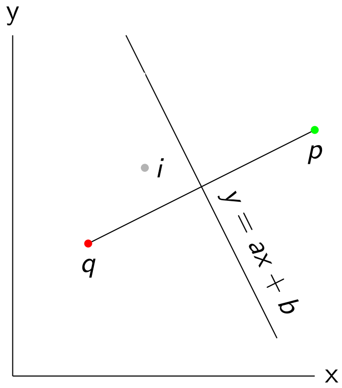
2. Bisectriz (*cutlines*)

Igualamos $d = 0$ para despejar y_v

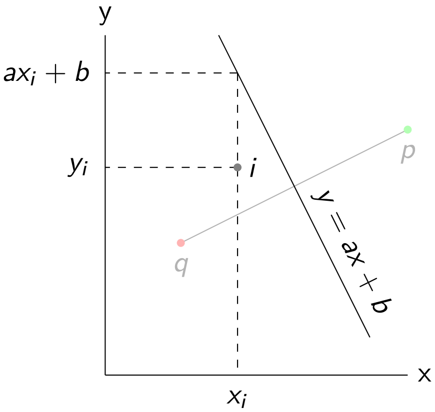
$$
\begin{aligned}
-2y_i y_q + 2y_i y_p &= 2x_i(x_q - x_p) - \left((x_q{}^2 - x_p{}^2) + (y_q{}^2 - y_p{}^2)\right) \\
y_i &= -\frac{x_q - x_p}{y_q - y_p} x_i + \frac{(x_q{}^2 - x_p{}^2) + (y_q{}^2 - y_p{}^2)}{2(y_q - y_p)} \\
&= \delta a x_i + \delta b \\
&= \delta(a x_i + b)
\end{aligned}
$$

- Ésta la usaremos en la ilustración

Cutlines



Cutlines



$y_i < ax_i + b \rightarrow$ vota no

Modelo reversible

La teoría espacial del voto puede aplicarse

1. fijando preferencias para explicar conducta (Downs 1957, Romer&Rosenthal 1978)
2. fijando conducta para inferir preferencias (Poole&Rosenthal 1985, Martin&Quinn 2002)

Haremos lo segundo:
inferir las preferencias políticas
de los diputados a partir de
sus votaciones nominales

"Data augmentation" en el modelo IRT

1. La clave es pensar en una variable latente que subyace cada respuesta dicotómica, de la siguiente manera:

$$y_{ij} = \begin{cases} 1 \text{ si } y^*_{ij} \geq 0 \\ 0 \text{ si } y^*_{ij} < 0 \end{cases}$$

2. La variable latente y^*_{ij} se puede interpretar como la propensión a votar *a favor* de una propuesta, y es una función de parámetros correspondientes a *individuos* y *propuestas*

$$y^*_{ij} = \beta_j(\theta_i - \alpha_j)$$

3. Alternativamente, podemos escribir

$$\Pr(y_{ij} = 1) = \Phi(y^*_{ij}),$$

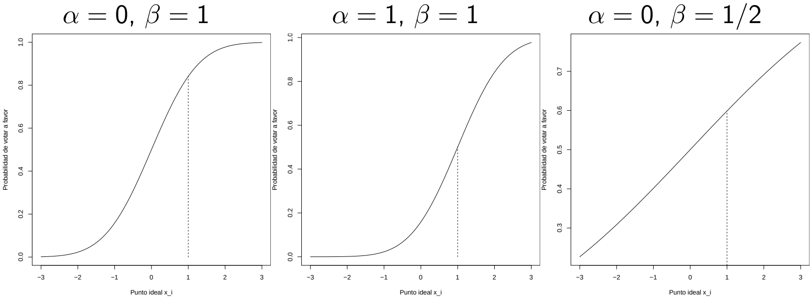
es decir, una función "probit"

Algoritmo de Gibbs para el modelo IRT

Interesa obtener una muestra de la densidad posterior conjunta de parámetros y variables latentes. El algoritmo repite tres pasos en cada turno:

1. $g(\mathbf{y}^*|\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{y})$ es una densidad normal truncada con media $\beta_j^{t-1}\theta_i^{t-1} - \alpha_j^{t-1}$ (data augmentation)
2. $g(\boldsymbol{\theta}|\mathbf{y}^*, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{y})$ es una densidad normal con media m_{x_i} y varianza ν_{x_i}
3. $g(\boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{\theta}, \mathbf{y}^*, \mathbf{y})$ es también una densidad normal con media m_j y varianza ν_j

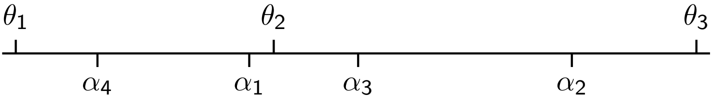
¿Para qué sirven los parámetros de discriminación y dificultad?



Múltiples rotaciones y escalamientos de los parámetros

El modelo básico $y_{ij}^* = \beta_j(\theta_i - \alpha_j)$ no está identificado:

- Escalas múltiples: El modelo no cambia si usamos $\theta_i' = k + \theta_i$ y $\alpha_j' = k + \alpha_j$
- Rotaciones múltiples: El modelo no cambia si usamos $\theta_i' = k \cdot \theta_i$ y $\beta_j' = \beta_j / k$ y $\alpha_j' = \alpha_j / k$



En resumen, hay un número infinito de soluciones para $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$

Densidades a priori para los parámetros del modelo IRT

Comúnmente se usan las densidades a prior de los parámetros θ, α y β para identificar el modelo IRT:

1. La escala del modelo se puede fijar con una densidad a priori sobre los puntos ideales:

$$\Pr(\theta) \sim \mathcal{N}(0, 1)$$

2. De manera similar, una densidad a priori de tipo

$$\Pr(\alpha) = \Pr(\beta) \sim \mathcal{N}(0, 2)$$

   permite estimar los parámetros de discriminación y dificultad en una escala similar

3. La polaridad del modelo se puede fijar "anclando" un par de puntos ideales ($\boldsymbol{\theta}$) o un par de parámetros de discriminación ($\boldsymbol{\beta}$)

```
posterior.1D <- MCMCirtKd(Roll.Call,
                dimensions=1,
                store.item=TRUE, store.ability=TRUE,
                burnin=15000, mcmc=30000, thin=10,
                seed=1971,
                item.constraints=list(
                      vote.left=list(2,"-"),
                      vote.right=list(2,"+")),
                b0=0, B0=.25)
```

¿Qué es BUGS/JAGS?

- Es una serie de modulos que permiten estimar modelos bayesianos mediante simulaciones MCMC
  - WinBUGS sólo funciona en plataforma Windows
  - JAGS se puede compilar en todo tipo de plataformas
- Código de fuente abierta
- Mucha ayuda para WinBUGS; relativamente poca para JAGS
- Los mensajes de error son muy poco informativos

Modelo unidimensional del Congreso mexicano

JAGS (Plummer)

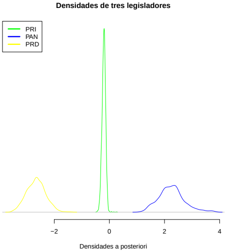
```
CJR.1D = "model  {
      for(i in 1:n.legs)
      {
           for(j in 1:n.item)
           {
                rc[i,j] ~ dbern(p[i,j]);
                #IRT MODEL
                probit(p[i,j]) <- mu[i,j];
                mu[i,j] <- beta[j]*theta[i] - alpha[j];
            }
       }
       # SAVE ONLY A HANDFUL OF PARAMETERS
       for (y in 1:50)
       {
            Beta[y]  <- beta[Y[y]];
            Alpha[y] <- alpha[Y[y]];
       }
```

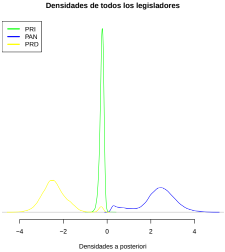
Modelo unidimensional del Congreso mexicano

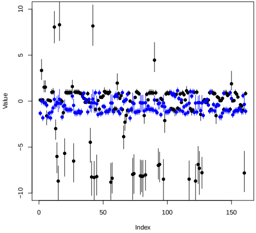
. . . continuación

```
    # PRIORS ON DISCRIMINATION PARAMETERS
    for(j in 1:(anchor[1]-1))
    {
       beta[j]  ~ dnorm( 0, 0.25);
    }
       beta[anchor[1]] ~ dnorm( 1, 100);
    for(j in (anchor[1]+1):(anchor[2]-1))
    {
       beta[j]  ~ dnorm( 0, 0.25);
    }
       beta[anchor[2]] ~ dnorm(-1, 100);
    for(j in (anchor[2]+1):n.item)
    {
       beta[j]  ~ dnorm( 0, 0.25);
    }
    # PRIORS ON DIFFICULTY PARAMETERS
    for(j in 1:n.item)
    {  alpha[j]  ~ dnorm( 0, 0.25);  }
    # PRIORS ON LEGISLATOR IDEAL POINTS
    for(i in 1:n.legs)
    {  theta[i]  ~  dnorm(0, 1);  }
}"
```

Modelo unidimensional del Congreso mexicano (MCMCpack)







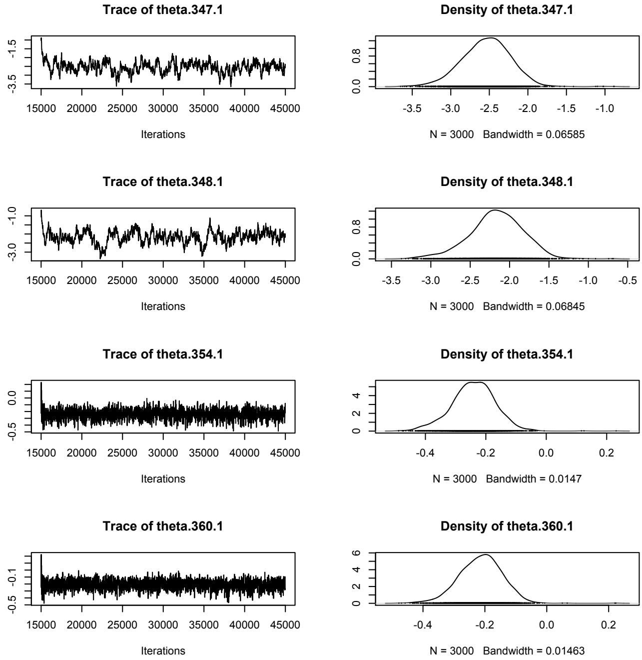
Tests de convergencia

1. Test de Geweke: Compara los promedios de ψ correspondientes a dos etapas de la simulación MCMC
2. Test de Heidelberger-Welch
3. Test de Raftery-Lewis
4. Cadenas múltiples (Gelman-Rubin):

$$\widehat{\text{var}}^{+}(\psi|y) = \frac{T-1}{T}W + \frac{1}{T}B$$

$$\sqrt{\widehat{R}} = \sqrt{\frac{\widehat{\text{var}}^{+}(\psi|y)}{W}}$$

En `JAGS`: `gelman.diag (post.2D.cjr$mcmc)`


Trace of theta.347.1
Iterations
Density of theta.347.1
N = 3000 Bandwidth = 0.06585
Trace of theta.348.1
Iterations
Density of theta.348.1
N = 3000 Bandwidth = 0.06845
Trace of theta.354.1
Iterations
Density of theta.354.1
N = 3000 Bandwidth = 0.0147
Trace of theta.360.1
Iterations
Density of theta.360.1
N = 3000 Bandwidth = 0.01463

Temas

Introducción
- Variables latentes
- Inferencia bayesiana
- R, RStudio y Jags (Barrios)

Regresión lineal: El efecto de arrastre de ejecutivos
- Estimación clásica
- Estimación bayesiana

IRT I: Introducción, teoría espacial del voto, CDHCU

IRT II: Modelos jerárquicos, dinámicos, multidimensionales

Modelos con variables latentes: análisis factorial bayesiano

Modelo bidimensional del Congreso mexicano

JAGS (Plummer)

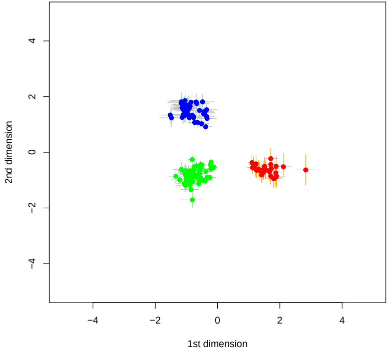
```
CJR.1D = "model  {
      for(i in 1:n.legs)
      {
           for(j in 1:n.item)
           {
                rc[i,j] ~ dbern(p[i,j]);
                #IRT MODEL
                probit(p[i,j]) <- mu[i,j];
                mu[i,j] <- beta[j]*theta[i,1] +     # <= Cambio
                           delta[j]*theta[i,2] -
                           alpha[j];
           }
       }
       # PRIORS ON DIFFICULTY PARAMETERS
       for(j in 1:n.item)
       {  alpha[j]  ~ dnorm( 0, 0.25);  }
       # PRIORS ON LEGISLATOR IDEAL POINTS
       for(i in 1:n.legs) {
          theta[i,1]  ~  dnorm(0, 1);        # <= Cambio
          theta[i,2]  ~  dnorm(0, 1);
       }
```

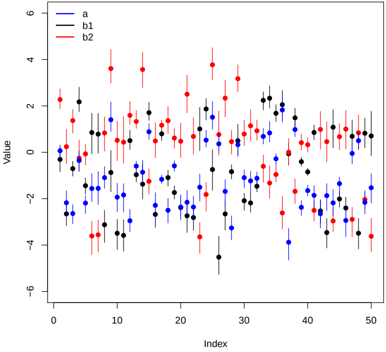
Modelo bidimensional del Congreso mexicano

. . . continuación

```
    # PRIORS ON DISCRIMINATION PARAMETERS
    for(j in 1:(anchor[1]-1))
    {
       beta[j]  ~ dnorm( 0, 0.25);
    }
       beta[anchor[1]] ~ dnorm( 1, 100);
    for(j in (anchor[1]+1):(anchor[2]-1))
    {
       beta[j]  ~ dnorm( 0, 0.25);
    }
       beta[anchor[2]] ~ dnorm(-1, 100);
    for(j in (anchor[2]+1):n.item)
    {
       beta[j]  ~ dnorm( 0, 0.25);
    }
    for(j in 1:(n.item-1))           # <= Cambio
    {
       delta[j] ~ dnorm( 0, 0.25);
    }
       delta[n.item] ~ dnorm( 1, 4);
}"
```

Modelo bidimensional del Congreso mexicano (MCMCpack)





Temas

Introducción
- Variables latentes
- Inferencia bayesiana
- R, RStudio y Jags (Barrios)

Regresión lineal: El efecto de arrastre de ejecutivos
- Estimación clásica
- Estimación bayesiana

IRT I: Introducción, teoría espacial del voto, CDHCU

IRT II: Modelos jerárquicos, dinámicos, multidimensionales

Modelos con variables latentes: análisis factorial bayesiano