= Mezclas de Normales en media-varianza =

(01)

P: David Castillo

x - vector p-dimensional $(x = (x_1, \dots, x_p)')$

μ - vector p-dimensional $(\mu = (\mu_1, \dots, \mu_p)')$

γ - vector p-dimensional $(\gamma = (\gamma_1, \dots, \gamma_p)')$

Σ - matriz de covarianzas $(p \times p)$.

λ - escalar positivo.

$\lambda \longmapsto \mu + \lambda\gamma$ es un mapeo de $\mathbb{R}_+$ a $\mathbb{R}^p$

con un vector de localización μ y una dirección dada por el vector γ, en magnitud λ.

$\lambda \longmapsto \lambda\Sigma$ es un mapeo de $\mathbb{R}_+$ a $\mathbb{R}^{p \times p}$ (simétrico)

es una expansión (o contracción, para $\lambda < 1$) de Σ. en todas direcciones.

Modelo Parte 1

$$X \mid \lambda, \mu, \gamma, \Sigma \sim N_p(x \mid \mu + \lambda\gamma, \underbrace{\lambda\Sigma}_{\text{varianza}})$$

$$f(x \mid \lambda, \mu, \gamma, \Sigma) \propto |\Sigma|^{-1/2} \lambda^{-p/2} \exp\left\{ (x-\mu)'\Sigma^{-1}\gamma \right\} \times \exp\left\{ -\frac{1}{2\lambda} (x-\mu)'\Sigma^{-1}(x-\mu) \right\} \times \exp\left\{ \frac{1}{2/\lambda} \gamma'\Sigma^{-1}\gamma \right\}.$$

supongamos que $\lambda \sim G(\cdot)$

$$\Rightarrow p(x \mid \mu, \gamma, \Sigma) = \underbrace{\int_{\mathbb{R}_+} N_p(x \mid \mu + \lambda\gamma, \lambda\Sigma) \cdot g(\lambda)\, d\lambda}_{\text{mezcla en media - varianza}}$$

Caso particular.

Si $\lambda \sim$ Generalized-Inverse-Gaussian

$\Rightarrow$ $p(x|\mu,\gamma,\Sigma)$ $\leftarrow$ Distribución hiperbólica generalizada.

(McNeil, pp. 78-85)

Caso alternativo

$$\underbrace{\lambda \sim \text{Gaussiana-Inversa (Inverse-gaussian)}}$$

Ambos casos, $p(x|\mu,\gamma,\Sigma)$ tiene curvas de nivel NO elípticas. con distribuciones marginales asimétricas, para los argumentos de x.

—— o —— o —— o ——

Identificabilidad (p. 79, McNeil)

los parámetros Σ y γ inducen la misma distribución que $k\Sigma$ y $k\gamma$ para todo $k>0$.

$\Rightarrow$ restringir $|\Sigma|$ $\leftarrow$ determinante, sea de un valor particular.

Particularmente $|\Sigma|=1$, a partir de la matriz de correlaciones.

—— o —— o —— o ——

Modelo Parte 1

$$p(x|\mu,\gamma,\Sigma) = \int_{\mathbb{R}_+} N_p(x|\mu+\lambda\gamma,\lambda\Sigma)\,\text{Ga-Inv}(\lambda|\,d\lambda$$

con Ga-Inv.(λ) $\leftarrow$ Gaussiana-Inversa, con función de densidad:

$$f(\lambda|\alpha,\beta) = \left(\frac{\beta}{2\pi\lambda^3}\right)^{1/2}\exp\left\{-\frac{\beta(\lambda-\alpha)^2}{2\alpha^2\lambda}\right\}\mathbb{1}_{(0,\infty)}(\lambda)$$

con $\alpha>0$ $\leftarrow$ parámetro de localización $\quad \mathbb{E}(\lambda|\alpha,\beta)=\alpha$

$\beta>0$ $\leftarrow$ parámetro de forma $\quad var(\lambda|\alpha,\beta)=\frac{\alpha^3}{\beta}$.

Modelo Parte 1

Parámetros:

1) μ - vector $(p\times 1)$ $\in \mathbb{R}^p$
2) γ - vector $(p\times 1)$ $\in \mathbb{R}^p$
3) $\Sigma \leftarrow$ matriz $(p\times p)$ $\leftarrow$ correlación con $|\Sigma|=1$ $\in \mathbb{R}^{p\times p}$.
4) α — media (localización) $\in \mathbb{R}_+$
5) β — forma $\in \mathbb{R}_+$.

— o —

Modelo jerárquico

I.— $x_j \mid \lambda_j, \mu, \gamma, \Sigma, \alpha, \beta \overset{ind}{\sim} N_p(x_j \mid \mu + \lambda_j\gamma, \lambda_j\Sigma)$ (observable)

$j=1,\dots,n$

$\lambda_j \overset{ind}{\sim} \text{GaussInv}(\lambda_j \mid \alpha, \beta)$ (latente)

II.— distrib. iniciales sobre parámetros.

sobre obs.— (μ,γ,Σ)

sobre latente (α,β)

$$\pi(\underbrace{\mu,\gamma,\Sigma}_{s/obs}, \underbrace{\alpha,\beta}_{s/lat.}) =$$

$$= \pi(\mu,\gamma,\Sigma)\cdot\pi(\alpha,\beta)$$

$$= \pi(\mu\mid\Sigma)\cdot\pi(\gamma\mid\Sigma)\cdot\pi(\Sigma)\cdot\pi(\alpha)\cdot\pi(\beta).$$

Sobre ~~los~~ latentes:

$$\pi(\alpha) = \text{Gamma}(\alpha \mid a_\alpha, b_\alpha)$$
$$\pi(\beta) = \text{Gamma}(\beta \mid a_\beta, b_\beta).$$

$a_\alpha, b_\alpha > 0$

$\underbrace{a_\beta, b_\beta > 0}_{\text{Hiperparámetros}}$

Sobre observables:

$$\pi(\mu \mid \Sigma) = N_p(\mu \mid m_\mu, s_0 \Sigma)$$
$$\pi(\gamma \mid \Sigma) = N_p(\gamma \mid m_\gamma, s_0 \Sigma)$$

m_μ, s_0

$\underbrace{m_\gamma, s_0}_{\text{Hiperparámetros}}$

$$\pi(\Sigma) = \text{Wishart}(\Sigma \mid \nu, S)$$

$\nu > p \leftarrow$ grados de libertad

$\underbrace{S \in \mathbb{R}^{p \times p} \text{ positivo definida}}_{\text{Hiperparámetros.}}$

Verosimilitud:

$$p(\underbrace{x_1, \dots, x_n}_{\text{obs}} \mid \mu, \gamma, \Sigma, \alpha, \beta) =$$
$$= \prod_{j=1}^{n} p(x_j \mid \mu, \gamma, \Sigma, \alpha, \beta)$$
$$= \prod_{j=1}^{n} \int_{\mathbb{R}_+} N_p(x_j \mid \mu + \lambda\gamma, \lambda\Sigma) \cdot \text{G-Inv}(\lambda \mid \alpha, \beta)\, d\lambda$$

Verosimilitud extendida:

$$p(\underbrace{x_1, \dots, x_n}_{\text{obs}}, \underbrace{\lambda_1, \dots, \lambda_n}_{\text{latentes}} \mid \mu, \gamma, \Sigma, \alpha, \beta) =$$
$$= \prod_{j=1}^{n} N_p(x_j \mid \mu + \lambda_j\gamma, \lambda_j\Sigma) \cdot \text{Ga-Inv}(\lambda_j \mid \alpha, \beta)$$

Distribución final sobre ~~[illegible]~~ latentes, y parámetros,

$$\pi(\mu,\gamma,\Sigma,\alpha,\beta,\lambda_1,\dots,\lambda_n \mid x_1,\dots,x_n) =$$

$$= \left\{ \prod_{j=1}^{n} N_p(x_j \mid \mu+\lambda_j\gamma, \lambda_j\Sigma) \cdot \text{Ga-Inv}(\lambda_j \mid \alpha,\beta) \right\}$$

$$\times\, Ga(\alpha \mid a_\alpha, b_\alpha) \cdot Ga(\beta \mid a_\beta, b_\beta)$$

$$\times\, N_p(\mu \mid m_\mu, s_0\Sigma)$$

$$\times\, N_p(\gamma \mid m_\gamma, s_0\Sigma)$$

$$\cdot\, Wi(\Sigma \mid \tau, S) \quad \dots\dots \quad (*)$$

Hiperparámetros:

•) a_α, $b_\alpha > 0$

••) $a_\beta, b_\beta > 0$

•••) m_μ y m_γ, ambos $\in \mathbb{R}^p$

••••) $s_0 > 0$

—) $\tau > 0$

—•) $S \in \mathbb{R}^{p\times p}$ positiva definida y simétrica.

Fijados / calibrados por nosotros.

(*) no tiene solución analítica.

→ Necesitamos emplear el Gibbs sampler (entre otros).

= Algoritmo - Gibbs sampler =

Parámetros del algoritmo:

$M \leftarrow$ extensión de la cadena de Markov (# simulaciones)

<u>Fijar repositorios</u> para

α ~~[illegible]~~ $\leftarrow$ vector $M \times 1$

β ~~[illegible]~~ $\leftarrow$ 〃 〃

μ, γ $\leftarrow$ arreglo de $M \times p$.

$\Sigma \leftarrow$ arreglo de $M \times p \times p$.

λ_j's $\leftarrow$ arreglo de $M \times n$

<u>Fijar hiperparámetros</u>:

$a_\alpha, b_\alpha \leftarrow$ dos escalares positivo

$a_\beta, b_\beta \leftarrow$ 〃 〃 positivo

$m_\mu, m_\gamma \leftarrow$ dos vectores de $p \times 1$

$\tau \leftarrow$ un escalar positivo

$S_0 \leftarrow$ 〃 〃 〃

$S \leftarrow$ matriz $(p \times p)$ positiva definida y simétrica.

<u>Fijar valores iniciales</u>

$\alpha^{(0)}, \beta^{(0)}, \mu^{(0)}, \Sigma^{(0)}, \gamma^{(0)}$ } parámetros

$\lambda_1^{(0)}, \ldots, \lambda_n^{(0)}$ } latentes.

<u>Recursión</u>:

para $m = 1, \ldots, M$

a) $\alpha^{(m)} | \ldots \sim \pi(\alpha | \ldots) \propto$

$$\propto \prod_{j=1}^{n} \text{Ga-Inv}(\lambda_j^{(m-1)} | \alpha, \beta^{(m-1)}) \cdot \text{Ga}(\alpha | a_\alpha, b_\alpha).$$

b)

$$\beta^{(m)} | \ldots \sim \pi(\beta | \ldots)$$
$$\propto \prod_{j=1}^{n} \text{Ga-Inv}\left(\lambda_j^{(m-1)} \middle| \alpha^{(m)}, \beta\right) \cdot$$
$$\cdot \, \text{Ga}(\beta | a_\beta, b_\beta).$$

c)

$$\mu^{(m)} | \ldots \sim \pi(\mu | \ldots)$$
$$\propto \prod_{j=1}^{n} N_p\left(x_j \middle| \mu + \lambda_j^{(m-1)} \gamma^{(m-1)}, \lambda_j^{(m-1)} \Sigma^{(m-1)}\right)$$
$$\cdot \, N_p\left(\mu | m_\mu, s_0 \Sigma^{(m-1)}\right)$$
$$\propto \prod_{j=1}^{n} N_p\left(x_j - \lambda_j^{(m-1)} \gamma^{(m-1)} \middle| \mu, \lambda_j^{(m-1)} \Sigma^{(m-1)}\right)$$
$$\cdot \, N_p\left(\mu | m_\mu, s_0 \Sigma^{(m-1)}\right)$$

d)

$$\gamma^{(m)} | \ldots \sim \pi(\gamma | \ldots)$$
$$\propto \prod_{j=1}^{n} N_p\left(x_j \middle| \mu^{(m)} + \lambda_j^{(m-1)} \gamma, \lambda_j^{(m-1)} \Sigma^{(m-1)}\right)$$
$$\cdot \, N_p\left(\gamma | m_\gamma, s_0 \Sigma^{(m-1)}\right)$$
$$\propto \prod_{j=1}^{n} N_p\left(\frac{x_j - \mu^{(m)}}{\lambda^{(m-1)}} \middle| \gamma, \Sigma^{(m-1)}\right)$$
$$\cdot \, N_p\left(\gamma | m_\gamma, s_0 \Sigma^{(m-1)}\right)$$

e)

$$\Sigma^{(m)} | \ldots \sim \pi(\Sigma | \ldots)$$
$$\propto \prod_{j=1}^{n} N_p\left(x_j \middle| \mu^{(m)} + \lambda_j^{(m-1)} \gamma^{(m)}, \lambda_j^{(m-1)} \Sigma\right)$$
$$\cdot \, Wi(\Sigma | z, S)$$

$$\propto \prod_{j=1}^{n} N_p\left(x_j / \lambda^{(m-1)^{1/2}} \;\middle|\; \frac{\mu^{(m)} + \lambda_j^{(m-1)} \gamma^{(m)}}{\lambda_j^{(m-1)^{1/2}}}, \Sigma \right)$$

$$\cdot \; Wi\,(\Sigma \mid \eta, S)$$

f) $\pi(\lambda_1, \ldots, \lambda_n \mid \ldots) \propto$ (no se puede calcular analíticamente; empleamos un MH.

para $j = 1, \ldots, n$

$$\lambda_j^{(m)} \mid \ldots \propto \text{Gam-Inv}\left(\lambda_j \mid \alpha^{(m)}, \beta^{(m)}\right) \cdot N_p\left(x_j \mid \mu^{(m)} + \lambda_j \gamma^{(m)}, \lambda_j \Sigma\right)$$

— · — · —

<u>Final</u> (Salida)

En cada iteración m $(m = 1, \ldots, M)$

almacenamos
$\alpha^{(m)}$
$\beta^{(m)}$
$\mu^{(m)}$
$\gamma^{(m)}$
$\Sigma^{(m)}$
$\lambda_1^{(m)}, \ldots, \lambda_n^{(m)}$

en los repositorios correspondientes. . . .