

# Análisis de Factores

Stefano Molina Martínez

Octubre 2016

## Introducción

Existen ocasiones en las que se lleva a cabo un estudio estadístico en el que las variables deseadas son creadas por el sujeto que está llevando a cabo el estudio y no son cuantificables con variables observables o medibles de manera fácil. Éstas variables se llaman *latentes*, las cuales se identifican por no ser medibles pero se pueden relacionar a otras que sí pueden ser medidas, llamadas *variables manifiestas*. El análisis de factores es un método para desarrollar las relaciones que existen entre estos tipos de variables, las cuales se definen como combinaciones lineales de varias *variables manifiestas* para formar una *variable latente*, mejor conocida como *factor*. El modelo de factores se define de la siguiente manera:

$$x = \Lambda f + \theta + e$$

donde  $x^T = (x_1, x_2, \dots, x_N)$  es el vector de variables *manifiestas* que se relacionan con las *variables latentes* no observadas, o *factores*, denotados por  $f = (f_1, f_2, \dots, f_k)$  de dimension  $k < N$  y tan pequeño como sea posible, mientras que la matriz  $\Lambda$  contiene la información que relaciona a éstas dos variables, conocida como la matriz de loadings, los cuales representan el porcentaje de la varianza explicada por ese factor para la variable  $x_i$ . El vector  $e \in \mathbb{R}^N$  representa al error generado por el modelo y generalmente posee una varianza pequeña.

# Modelo factorial

Sea el modelo de análisis factorial

$$x = \Lambda f + \theta + e$$

A las suposiciones antes mencionadas se agregan otras con respecto a  $f$ :

- $f \perp e$ , *i.e.*  $f$  y  $e$  son independientes
- $f \sim N(0, I_k)$
- $e \sim N(0, D_\phi)$
- $\Lambda^T D_\psi^{-1} \Lambda \equiv D_J$ ,  $D_J \in \mathbb{R}^{k \times k}$ ,  $D_J = \text{diag}(J_1, J_2, \dots, J_k)$  con  $J_1 > J_2 > \dots > J_k$

La necesidad de tener  $k < N$  recae en que lo que se busca con el modelo es simplificar la estructura de los datos observados, de manera que su representación debe ser sencilla y consistente.

Press(1982), sugiere asumir las distribuciones de  $f$  y  $e$  como normales para poder llevar a cabo la estimación por máxima verosimilitud de los parámetros  $\theta$ ,  $\Lambda$  y  $D_\psi$ , también sugiere, en el caso de no existir normalidad, llevar a cabo una transformación que induzca normalidad a la variables. La matriz de varianza y covarianza para  $f$ , definida como  $I_k$  permite que los factores sean invariantes a cambios de escala, mientras que la no correlación entre factores puede cambiar acorde al tipo de modelo factorial que se trabaje.

Los errores  $e$  tienen la propiedad de ser mutuamente independientes y media cero para buscar la menor variación posible con respecto a la variable real, pero también se tiene la propiedad de heterocedasticidad que permite que las varianzas de dos errores no sean necesariamente las mismas.

El supuesto para  $D_J$  permite eliminar la infinidad de soluciones que admite el modelo inicial, donde cada transformación ortogonal posible genera una matriz  $\Lambda$  diferente. La matriz obtenida no siempre es la que permite llevar a cabo un mejor estudio, por lo que a partir de la solución impuesta se puede trabajar con rotaciones que se ajusten de mejor manera al modelo.

Dadas las suposiciones nombradas, se tienen las siguientes propiedades respecto a la varianza en el análisis factorial:

La varianza de la variable  $x_i$ ,  $\sigma_i^2$ , se define como

$$\sigma_i^2 = \sum_{j=1}^k \lambda_{ij}^2 + \psi_i^2$$

donde  $\lambda_{ij}$  determina el *loading* para el factor  $f_j$  con respecto a la variable  $x_i$  y  $\psi_i$  es la varianza del error correspondiente,  $e_i$ . La primera parte de la suma,  $h_i^2 = \sum_{j=1}^k \lambda_{ij}^2$ , representa la varianza que  $x_i$  comparte con las otras variables por medio de los factores, mientras que  $\psi_i$  es la varianza intrínseca de la variable y no la comparte con las demás.

La covarianza entre dos variables  $x_i$  y  $x_j$  con  $i \neq j$  se determina como

$$\sigma_{ij} = \sum_{l=1}^k \lambda_{il} \lambda_{lj}$$

Dadas éstas propiedades de la varianza, se obtiene que la matriz de varianzas y covarianzas de  $x$ ,  $\Sigma$ , se puede representar como la suma de un producto de matrices y una matriz

$$\Sigma = \Lambda \Lambda^T + \Psi, \quad \Psi = \text{diag}(\psi_i)$$

El primer problema respecto al modelo de factores se presenta en este punto ya que las varianzas y covarianzas de  $x$  no siempre son conocidas o fáciles de obtener, por lo que es necesario estimarlas a través de  $\hat{\Lambda}$  y  $\hat{\Psi}$ .

## Estimación de parámetros

Existen dos métodos principales de estimación para los parámetros del modelo de factores, uno está basado en el método de análisis de componentes principales y el otro en el método de máxima verosimilitud, siendo éste último el que más se utiliza por la teoría detrás de su cálculo, donde se determina el mejor estimador insesgado para los parámetros.

### Estimación por máxima verosimilitud

Como el método en estimación de parámetros en Inferencia Estadística, el método de máxima verosimilitud para estimar los parámetros en el análisis de factores es un método robusto y ampliamente aceptado, se supone una distribución normal multivariada para  $x$  tal que la función de verosimilitud,  $L$ , se representa como  $L = -\frac{1}{2}mF$  sumada a una función de las observaciones. La función  $F$  se define como

$$F = \ln|\Lambda\Lambda^T - \Psi| + \text{tr}(S|\Lambda\Lambda^T - \Psi|^{-1}) - \ln(|S|) - n$$

donde  $S = \hat{\Sigma} = \frac{1}{n} \sum_{j=1}^n (x_j - \hat{\theta})(x_j - \hat{\theta})^T$ ,  $\hat{\theta} = \bar{x}$  y  $n$  la dimensión del vector  $x$ . La estimación de los parámetros se encuentra maximizando  $L$  con respecto a cada parámetro ( $\Lambda$  y  $\Psi$ ).

Una ventaja de la estimación por máxima verosimilitud es que posee un modelo asociado que permite estimar el número de factores que representan de manera adecuada a las variables estudiadas. El método es iterativo, opera sobre una prueba de hipótesis y se describe a

continuación.

Sea  $K = m + 1 - \frac{1}{6}(2q + 5) - \frac{2}{3}k$ ,

**Para**  $k=1,2,\dots,n-1$

Probar la hipótesis  $H_k$  con la estadística de prueba  $U = K \min(F)$  para los valores de  $\Lambda$  y  $\Psi$  de máxima verosimilitud

**Si**  $U$  es significativa

Alto

**en otro caso**

$k=k+1$

**Fin si**

**Fin para**

Cuando se acepta la hipótesis  $H_k$ , se tiene que  $U$  tiene de manera asintótica una distribución Ji cuadrada con  $\frac{1}{2}(q - k)^2 - \frac{1}{2}(q + k)$  grados de libertad.

## Estimación de factores

Como se menciona anteriormente, los factores son variables que poseen la característica de no ser fácilmente medibles ni observables, por lo que su estudio se tiene que llevar a cabo sujeto a variables relacionadas que sí se pueden medir y ocupar para su estudio. La estimación de los factores se lleva a cabo de una forma similar, se tiene que suponer a la distribución de  $f$  como una función de  $x$ .

Dado que  $f$  tiene una distribución normal, se puede probar que  $f$  dado  $x$  se distribuye  $f|x \sim N(\Lambda^T \Sigma^{-1} x, (\Lambda^T \Psi^{-1} \Lambda + I_n)^{-1})$ , por lo que conociendo  $\hat{\Sigma} = S = \frac{1}{n} \sum_{j=1}^N (x_j - x)(x_j - x)^T = \frac{1}{n} X X^T - \hat{\theta} \hat{\theta}^T$ , se puede obtener el valor esperado de  $f$ , el cual se define como su estimador.

$$\hat{f} \approx \hat{\Lambda}^T (S + \hat{\theta} \hat{\theta}^T)^{-1} x$$

donde se asume  $x$ , centrada.

## Estimación por métodos Bayesianos

La estimación por métodos bayesianos adquiere interés cuando se buscan los valores de los parámetros sin la ambigüedad que puede generar la posibilidad de tener una infinidad de soluciones del método de máxima verosimilitud.

Tomando en cuenta que  $x$  se distribuye normal, se define  $A \sim W(\Sigma, k, n^*)$ , con  $W$  la distribución Wishart con  $p$  grados de libertad y  $n = N - 1$ , tal que

$$f(A) \propto \frac{|A|^{(n-p-1)/2}}{|\Sigma|^{n/2}} \exp\left(-\frac{1}{2} \text{tr} \Sigma^{-1} A\right)$$

Por lo que la verosimilitud dados los parámetros,  $p(A|\Lambda, \psi_1, \psi_2, \dots, \psi_k)$ , se obtiene sustituyendo  $\Sigma = \Lambda \Lambda^T + D_\psi$  en la densidad de  $A$ .

Sea  $w_0$  el vector con información acerca de  $q$  elementos de  $\Lambda$ , tal que  $w_0 \sim N(\Phi, G)$  con  $G > 0$ ,  $\Phi$  conocidas (se acostumbra  $\Phi = 0$  y  $G$  diagonal con valores pequeños). Si se define  $w_1$  como el vector con los valores restantes de  $\Lambda$  con distribución no informativa o constante y una a priori para  $(\psi_1, \psi_2, \dots, \psi_k)$  de la forma  $p(\psi_1, \psi_2, \dots, \psi_k) \propto |D_\psi|^{-v}$ ,  $v < 1$ , se tiene que la a priori conjunta para  $w_0$ ,  $w_1$  y  $\psi_1, \psi_2, \dots, \psi_k$ , es

$$\begin{aligned} p(\Lambda, D_\psi) &= p(w_0, w_1, \psi_1, \psi_2, \dots, \psi_k) \\ &\propto \frac{\exp\{-\frac{1}{2}(w_0 - \Phi)^T G (w_0 - \Phi)\}}{|D_\psi|^v} \end{aligned}$$

Por lo que la distribución posterior de  $\Lambda$  y  $D_\psi$  dada  $A$ , se obtiene con el producto de la verosimilitud y la a priori, resultando

$$\begin{aligned} p(\Lambda, D_\psi | A) &= p(w_0, w_1, \psi_1, \psi_2, \dots, \psi_k) p(A | \Lambda, \psi_1, \psi_2, \dots, \psi_k) \\ &\propto \frac{p(w_0)}{|D_\psi|^v |\Lambda \Lambda^T + D_\psi|^{n/2}} \exp\left\{-\frac{1}{2} \text{tr} A (\Lambda \Lambda^T + D_\psi)^{-1}\right\} \end{aligned}$$

la cual se integra respecto a  $(\psi_1, \psi_2, \dots, \psi_k)$  para encontrar la distribución posterior de  $\Lambda$ .