

# EL MÉTODO KERNEL DE EQUATING Y SU CONTRAPARTE BAYESIANA NO PARAMÉTRICA: UN ESTUDIO DE COMPARACIÓN BAJO EL DISEÑO DE GRUPOS EQUIVALENTES

Constanza Rojo Alfaro <sup>1</sup>, Jorge Gonzalez Burgos <sup>2</sup>

<sup>1</sup> *Dpto de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, mcrojo@mat.puc.cl*

<sup>2</sup> *Dpto de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, jgonzale@mat.puc.cl*

## RESUMEN

Equating es una familia de métodos y modelos estadísticos utilizados para ajustar los puntajes de dos versiones de un test de tal forma de hacerlos comparables. Los principales enfoques están basados ya sea en puntajes observados (i.e., número de respuestas correctas) o puntajes IRT (i.e., habilidad estimada mediante un modelo de teoría de respuesta al ítem). Un método particular de equating basado en puntaje observado es el denominado método Kernel. En este método se estiman paramétricamente las funciones de distribución acumuladas de los puntajes de ambas pruebas, y se utilizan para construir la denominada función de equating, por medio de la cual se realiza el ajuste de los puntajes de ambos tests. Recientemente, métodos Bayesianos no-paramétricos se han propuesto para la estimación de las funciones de distribución de puntajes, como una alternativa al método paramétrico. En este trabajo estudiamos el método Bayesiano no-paramétrico de equating. Se propone el uso de polinomios de Bernstein para el modelamiento de las funciones de distribución de puntajes acumulada. Utilizando datos reales analizados ampliamente en la literatura se realiza una comparación entre la metodología propuesta y el método tradicional de Kernel.

*PALABRAS CLAVES:* R Software, kernel equating, métodos Bayesianos no-paramétricos.

## 1. Introducción

Las técnicas de equating constituyen una familia de métodos y modelos estadísticos que son utilizados para ajustar, por diferencias en dificultad, los puntajes de dos versiones de una prueba de tal forma de hacerlos comparables. Mediante distintos esquemas de muestreo, es posible corregir el efecto de diferencias intrínsecas en habilidad de los grupos que se comparan. Una vez corregidos estos efectos, el propósito del equating es obtener puntajes comparables en ambos grupos. En este trabajo, nos enfocaremos en el diseño de grupos equivalentes (GE) (von Davier et al., 2004, Cap. 2; Kolen & Brennan, 2004, Sección 1.4). Denotemos por  $X$  e  $Y$  a los puntajes que provienen de las pruebas  $\mathcal{X}$  e  $\mathcal{Y}$ , respectivamente. Sean además  $F_X(x)$  y  $F_Y(y)$  las funciones de distribución acumulada de  $X$  e  $Y$ , respectivamente. Para transformar los puntajes  $X$  en la escala de  $Y$ , se define la función de equating

$$e_Y(x) = F_Y^{-1}(F_X(x)) \quad (1)$$

En este trabajo consideramos puntajes observados para pruebas con preguntas puntuadas como correctas o incorrectas<sup>1</sup>. De esta forma, el puntaje obtenido es definido como el número total de respuestas correctas en el test. Un evidente problema con la función de equating (1) es que las distribuciones de los puntajes son discretas, lo cual impide encontrar la inversa de las funciones  $F_X$  y/o  $F_Y$ . Una solución a este problema es continuizar las funciones de distribución discretas  $F_X$  y  $F_Y$  para poder utilizar de manera correcta la función de equating definida en (1).

### 1.1. El método Kernel de equating (KE)

En KE, se continuizan las funciones de distribución de puntajes acumulada  $F_X$  y  $F_Y$  siguiendo la siguiente idea. La variable aleatoria  $X$ , originalmente discreta, se transforma como  $X(h_X) = X + h_X V$  en donde  $V$  es una v.a. continua, de manera que la nueva variable  $X(h_X)$  es también continua. El parámetro  $h_X$  controla el grado de suavizamiento en la continuización. Se demuestra en von Davier et al. (2004) que si  $V \sim N(0,1)$  y la variable continuizada se define como  $X(h_X) = a_X(X + h_X V) + (1 - a_X)\mu_X$ , con  $a_X^2 = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_V^2 h_X^2}$ , entonces el kernel Gaussiano (Silverman, 1986) de  $F_X(x)$ , definido como

$$F_{h_X}(x) = \sum_j r_j \Phi(R_{jX}(x)), \quad (2)$$

---

<sup>1</sup>Los métodos de equating basados en la teoría de respuesta al ítem (IRT) consideran la habilidad estimada del individuo como puntaje.

en donde  $R_{jX}(x) = \frac{x - a_X x_j - (1 - a_X) \mu_X}{a_X h_X}$ , es exactamente la función de distribución acumulada de  $X(h_X)$ . La conversión de puntajes desde la escala de  $\mathcal{X}$  a la escala de  $\mathcal{Y}$  se basa finalmente en  $\hat{e}_Y(x) = G_{h_Y}^{-1}(F_{h_X}(x; \hat{r}); \hat{s})$  en donde  $r_j = P(X = x_j)$ ,  $s_j = P(Y = y_j)$  tal que  $F_X(x) = P(X \leq x) = \sum_{j, x_j \leq x} r_j$  y  $F_Y(y) = P(Y \leq y) = \sum_{k, y_k \leq y} s_k$ . Los valores  $\hat{r}$  y  $\hat{s}$  son estimados utilizando, típicamente, modelos log-lineales. El error estándar de equating (SEE) es una medida utilizada para evaluar la incertidumbre en la estimación de  $\hat{e}_Y(x)$ .

## 1.2. Método Bayesiano no-paramétrico de equating

Nuevas metodologías para hacer equating basadas en métodos bayesianos no-paramétricos han sido introducidas y estudiadas por Karabatsos y Walker (2009). Bajo este enfoque, la inferencia sobre la función de equating  $\hat{e}_Y(x)$ , se realiza de manera no-paramétrica. Contrario a como se hace en el método KE tradicional, este método asigna prioris sobre las funciones  $F_X(x)$  y  $F_Y(y)$ , que combinadas con los datos inducen posterioris continuas que pueden ser utilizadas para obtener la función (1).

En su trabajo, Karabatsos y Walker utilizan polinomios de Bernstein para realizar la inferencia sobre la distribución a posteriori del modelo bivariado  $(F_X(x), F_Y(y))$ . La estimación de este modelo está basada en una extensión del algoritmo Gibbs sampling descrito por Petrone (1999). Un punto relevante de la propuesta es que mediante la utilización de un proceso de Dirichlet bivariado se induce una estructura de correlación entre  $F_X(x)$  y  $F_Y(y)$ .

Lo que se propone en éste trabajo es comparar el equating obtenido considerando la estructura de correlación y omitiéndola, utilizando métodos Bayesianos no-paramétricos. El omitir la estructura de correlación entre las funciones de distribución acumulada de puntajes permitirá comparar el método Bayesiano no-paramétrico con el más tradicional KE.

## 2. Metodología

Una priori basada en el proceso Dirichlet (introducido por Ferguson, 1973) permite generar funciones de distribuciones aleatorias que pueden ser representadas por un parámetro finito dimensional.

El proceso Dirichlet puede ser descrito mediante la siguiente representación (Sethuraman, 1994): Si  $V_1, V_2, \dots$  son iid  $Beta(1, m)$  y  $\theta_1, \theta_2, \dots$  son iid  $G_0$ , entonces  $G \stackrel{c.s.}{=} \sum_{j=1}^{\infty} W_j \delta_{\theta_j}$ , en donde  $W_1 = V_1$  y para  $j \geq 2$   $W_j = V_j \prod_{i=1}^{j-1} (1 - V_i)$ .

Esta priori es denotada por  $DP(m, G_0)$  y por construcción cualquier distribución  $F$  generada de una priori de proceso Dirichlet es discreta con probabilidad 1.

Para modelar  $F_X$  y  $F_Y$  como distribuciones continuas, es posible utilizar polinomios de Bernstein (Petroni, 1999). Si se quiere modelar la dependencia entre  $F_X$  y  $F_Y$ , es posible considerar polinomios de Bernstein bivariados cuya dependencia es inducida mediante un proceso Dirichlet bivariado (ver más adelante).

Sea  $G : [0, 1] \rightarrow R$  una función (no necesariamente una distribución de probabilidad). El polinomio de Bernstein de grado  $p$  asociado está dado por

$$B(x|p, G) = \sum_{k=0}^p G\left(\frac{k}{p}\right) \binom{p}{k} x^k (1-x)^{p-k}.$$

Si  $G$  es la función de distribución acumulada (CDF) de probabilidad en el intervalo unitario, entonces la expresión anterior también es una CDF en  $[0, 1]$  y representa una mezcla de distribuciones Beta. Si  $G(0) = 0$ , su densidad está dada por

$$f(x; G, p) = \sum_{k=1}^p \omega_{k,p} \beta(x|k, p-k+1),$$

en donde  $\omega_{k,p} = G\left(\frac{k}{p}\right) - G\left(\frac{k-1}{p}\right)$ ,  $k = 1, \dots, p$  y  $\beta(\cdot|a, b)$  representa la densidad de la distribución Beta con parámetros  $a$  y  $b$ .

## 2.1. Estimación para el modelo sin estructura de correlación

Debido a que los polinomios de Bernstein generan distribuciones de probabilidad continuas en el intervalo unitario, se realizó primeramente una transformación de los puntajes observados (originalmente discretos) para llevarlos al intervalo  $[0, 1]$ , considerando  $x' = \frac{x - \min\{x\} + \epsilon}{\max\{x\} - \min\{x\} + 2\epsilon}$ .

De manera similar se transformaron los datos provenientes del test  $Y$ .

El algoritmo utilizado para la estimación se basa en la introducción de una variable auxiliar  $u_i$  para cada punto de los datos  $x_i$  tal que  $u_1, \dots, u_n | p, G$  son iid con respecto a  $G$  ( $i = 1, \dots, n$ ), y  $x_1, \dots, x_n | p, G, u_1, \dots, u_n$  son independientes con densidad conjunta dada por

$$\prod_{i=1}^n \beta(x_i | \theta(u_i, p), p - \theta(u_i, p) + 1),$$

en donde  $\theta(u_i, p) = \sum_{k=1}^p k \mathbf{1}(\mathbf{u}_i \in \mathbf{A}_{k,p})$  con  $A_{k,p} = \left[\frac{k-1}{p}, \frac{k}{p}\right]$  para  $k = 1, \dots, p$ . Luego para la inferencia de la distribución a posteriori, el Gibbs sampling procede generando desde la distribución condicional completa a posteriori de  $G$ ,  $p$  y  $u_i$  ( $i = 1, \dots, n$ ) para un cierto número de iteraciones. Para un  $p$  y  $u_i$  dados, la condicional a posteriori completa de  $\omega_p =$

$(\omega_{1,p}, \dots, \omega_{p,p})$  es  $Dirichlet(\omega_p | \alpha_{1,p}, \dots, \alpha_{p,p})$ , con

$$\alpha_{k,p} = mG_0(A_{k,p}) + n\hat{F}_u(A_{k,p}), \quad k = 1, \dots, p,$$

en donde  $G_0$  es la distribución de base del proceso Dirichlet para  $G$ , y  $\hat{F}_u$  es la distribución empírica de las variables latentes.

Para un  $u_i$  dado, la distribución a posteriori completa de  $p$  es proporcional a

$$\pi(p) \prod_{i=1}^n \beta(x_i | \theta(u_i, p), p - \theta(u_i, p) + 1).$$

Al considerar el muestreo de los  $u_i$  se consideran casos para los datos  $X$  e  $Y$  y es directo muestrear desde la distribución condicional a posteriori de los  $u_i$  (para más detalles, ver Petrone, 1999).

## 2.2. Estimación de modelos bivariados para correlación

La idea es la siguiente: Tomar  $G_X \sim DP(m, G_0)$  y para algún  $r \in \{0, 1, 2, \dots\}$  fijo, tomar  $z_1, \dots, z_r$  independientes e idénticamente distribuidos desde  $G_X$ .

Luego tomar  $G_Y \sim DP\left(m + r, \frac{mG_0 + r\hat{F}_r}{m+r}\right)$ , donde  $\hat{F}_r$  es la distribución empírica de  $\{z_1, \dots, z_r\}$ .

Se puede demostrar que la distribución marginal de  $G_Y$  es  $DP(m, G_0)$  y es posible tener las marginales de diferentes procesos Dirichlet. Sin embargo, se asume que las prioris de dos distribuciones aleatorias son la misma. Luego, para un intervalo  $A$ , la correlación entre  $G_X(A)$  y  $G_Y(A)$  está dada por  $Corr(G_X(A), G_Y(A)) = \frac{r}{m+r}$ .

De ésta forma se tiene una interpretación para la priori de  $r$  y así es posible considerar la dependencia entre las dos pruebas.

Después de iterar el algoritmo, y una vez que se alcanza convergencia aproximada de las cadenas de Markov, se dispone de muestras a posteriori de  $F_X$  y  $F_Y$ . Estas son resumidas ya sea tomando la media o la mediana a posteriori y se utilizan para estimar la función de equating definida en la ecuación (1).

## 3. Resultados

Se aplicaron las metodologías descritas anteriormente en el software R a un conjunto de datos ampliamente analizado en la literatura (ver, por ejemplo, von Davier et al., 2004; von Davier, 2011). Un total de 1.453 examinados rindieron el test  $X$  mientras que 1.455 realizaron el test

Y en el dominio de matemáticas. Cada test tiene 20 ítems y sus puntajes corresponden al número de respuestas correctas.

Se utilizó una distribución Beta como distribución base  $G_0$  y se compararon los del equating introduciendo una estructura de correlación y sin introducir correlación. Resultados preliminares muestran una amplia diferencia de los métodos KE y BNP, cuando se considera la correlación entre  $X$  e  $Y$ . En comparación a los métodos más tradicionales (e.g., KE), el método Bayesiano no-paramétrico garantiza consistencia en las estimaciones de las distribuciones marginales  $F_X$  y  $F_Y$  lo que implica consistencia en las estimaciones de la función de equating  $e_Y(\cdot)$ . Dada la flexibilidad del modelo inducido por la priori bivariada de Bernstein es posible modelar la dependencia entre  $G_X$  y  $G_Y$  y así inducir una estructura de correlación entre  $F_X$  y  $F_Y$  (gracias a las propiedades del proceso Dirichlet bivariado), lo cual no es posible hacer mediante los métodos usualmente utilizados para hacer equating. Ésta dependencia parece razonable considerando que, en la práctica, los dos tests están diseñados para medir las mismas habilidades de los examinados y los resultados muestran evidencia en contra de el supuesto de independencia. A pesar que los métodos BNP requieren mayor esfuerzo computacional, este garantiza mejores resultados además de considerar supuestos más realistas sobre las distribuciones de los puntajes.

#### 4. Referencias

- Karabatsos, G. y S. Walker (2009a). A Bayesian nonparametric approach to test equating. *Psychometrika* 74 (2), 211-232.
- Petrone, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics* 26 (3), 373-393.
- Von Davier, A., P. Holland, y D. Theayer (2004). *The kernel method of test equating*. Springer Verlag.
- Walker, S.G. and Muliere, P. (2003). A bivariate Dirichlet process. *Statistics and Probability Letters* 64, 1-7.