

# Chapter 8

## Using Exponential Families for Equating

Shelby J. Haberman

### 8.1 Introduction

In common equipercentile equating methods such as the percentile-rank method or kernel equating (von Davier, Holland, & Thayer, 2004b), sample distributions of test scores are approximated by continuous distributions with positive density functions on intervals that include all possible scores. The use of continuous distributions with positive densities facilitates the equating process, for such distributions have continuous and strictly increasing distribution functions on intervals of interest, so conversion functions can be constructed based on the principles of equipercentile equating. When the density functions are also continuous, as is the case in kernel equating, the further gain is achieved that the conversion functions are differentiable. This gain permits derivation of normal approximations for the distribution of the conversion function, so estimated asymptotic standard deviations (EASDs) can be derived.

An obvious challenge with any approach to equipercentile equating is the accuracy of an approximation of a discrete distribution by a continuous distribution. The percentile-rank approach, even with log-linear smoothing, provides an approximating continuous distribution with the same expectation as the original sample distribution but with a different variance. The kernel method provides an approximating continuous distribution with the same mean and variance as the original sample distribution, but higher order moments do not normally coincide.

With continuous exponential families, continuous distributions with positive and continuous density functions are obtained with a selected collection of moments that are consistent with the corresponding sample moments for the test scores. For example, one can specify that the first four moments of a distribution from a continuous exponential family are equal to the first four moments from a sample distribution.

---

S.J. Haberman  
Educational Testing Service, Rosedale Rd. Princeton, New Jersey 08541, USA  
e-mail: shaberman@ets.org

For simplicity, two equating designs are considered, a design for randomly equivalent groups and a design for single groups. In each design, Forms 1 and 2 are compared. Raw scores on Form 1 are real numbers from  $c_1$  to  $d_1$ , and raw scores on Form 2 are real numbers from  $c_2$  to  $d_2$ . For  $j$  equal 1 or 2, let  $X_j$  be a random variable that represents the score on form  $j$  of a randomly selected population member, so that  $X_j$  has values from  $c_j$  to  $d_j > c_j$ . It is not necessary for the  $X_j$  to have integer values or to be discrete, but many typical applications do involve raw scores that are integers. To avoid cases of no interest, it is assumed that  $X_j$  has a positive variance  $\sigma^2(X_j)$  for each form  $j$ .

The designs under study differ in terms of data collection. In the design for randomly equivalent groups, two independent random samples are drawn. For  $j$  equal 1 or 2, sample  $j$  has size  $n_j$ . The observations  $X_{ij}$ ,  $1 \leq i \leq n_j$ , are independent and identically distributed with the same distribution as  $X_j$ . In the design for a single group, one sample of size  $n_1 = n_2 = n$  is drawn with observations  $\mathbf{X}_i = (X_{i1}, X_{i2})$  with the same distribution as  $\mathbf{X} = (X_1, X_2)$ .

For either sampling approach, many of the basic elements of equating are the same. For any real random variable  $Y$ , let  $F(Y)$  denote the distribution function of  $Y$ , so that  $F(x, Y)$ ,  $x$  real, is the probability that  $Y \leq x$ . Let the quantile function  $Q(Y)$  be defined for  $p$  in  $(0,1)$  so that  $Q(p, Y)$  is the smallest  $x$  such that  $F(x, Y) \geq p$ . The functions  $F(X_j)$  and  $Q(X_j)$  are nondecreasing; however, they are not continuous in typical equating problems in which the raw scores are integers or fractions and thus are not readily employed in equating. In addition, even if  $F(X_j)$  and  $Q(X_j)$  are continuous, the sample functions  $\bar{F}(X_j)$  and  $\bar{Q}(X_j)$  are not. Here  $\bar{F}(x, X_j)$  is the fraction of  $X_{ij} \leq x$ ,  $1 \leq i \leq n_j$ , and  $\bar{Q}(p, X_j)$ ,  $0 < p < 1$ , is the smallest  $x$  such that  $\bar{F}(x, X_j) \geq p$ .

Instead of  $F(X_j)$  and  $Q(X_j)$ ,  $j$  equal 1 or 2, equipercentile equating uses continuous random variables  $A_j$  such that each  $A_j$  has a positive density  $f(A_j)$  on an open interval  $B_j$  that includes  $[c_j, d_j]$ , and the distribution function  $F(A_j)$  of  $A_j$  approximates the distribution function  $F(X_j)$ . For  $x$  in  $B_j$ , the density of  $A_j$  has value  $f(x, A_j)$ . Because the distribution function  $F(A_j)$  is continuous and strictly increasing, the quantile function  $Q(A_j)$  of  $A_j$  satisfies  $F(Q(p, A_j), A_j) = p$  for  $p$  in  $(0,1)$ , so that  $Q(A_j)$  is the strictly increasing continuous inverse of the restriction of  $F(A_j)$  to  $B_j$ . The equating function  $e(A_1, A_2)$  for conversion of a score on Form 1 to a score on Form 2 is the composite function  $Q(F(A_1), A_2)$ , so that, for  $x$  in  $B_1$ ,  $e(A_1, A_2)$  has value  $e(x, A_1, A_2) = Q(F(x, A_1), A_2)$  in  $B_2$ . Clearly,  $e(A_1, A_2)$  is strictly increasing and continuous. The conversion function  $e(A_2, A_1) = Q(F(A_2), A_1)$  from Form 2 to Form 1 may be defined so that  $e(A_2, A_1)$  is a function from  $B_2$  to  $B_1$ . The functions  $e(A_1, A_2)$  and  $e(A_2, A_1)$  are inverses of each other, so that  $e(e(x, A_1, A_2), A_2, A_1) = x$  for  $x$  in  $B_1$  and  $e(e(x, A_2, A_1), A_1, A_2) = x$  for  $x$  in  $B_2$ . If  $f(A_j)$  is continuous on  $B_j$  for each form  $j$ , then application of standard results from calculus show that the restriction of the distribution function  $F(A_j)$  to  $B_j$  is continuously differentiable with derivative  $f(x, A_j)$  at  $x$  in  $B_j$ , the quantile function  $Q(A_j)$  is continuously differentiable on  $(0,1)$  with derivative  $1/f(Q(p, A_j), A_j)$  at  $p$  in  $(0,1)$ , the conversion function  $e(A_1; A_2)$  is continuously differentiable with derivative  $e'(x, A_1, A_2) = f(x, A_1)/f(e(x, A_1, A_2), A_2)$  at  $x$  in  $B_1$ , and the conversion function

$e(A_2, A_1)$  is continuously differentiable with derivative  $e'(x, A_2, A_1) = f(x, A_2) / f(e, (x, A_1, A_2)A_2)$  at  $x$  in  $B_2$ .

In Section 8.2, continuous exponential families are considered for equivalent groups, and in Section 8.3, continuous exponential families are considered for single groups. The treatment of continuous exponential families in equating is closely related to Wang (2008, this volume); however, the discussion in this chapter differs from Wang in terms of numerical methods, model evaluation, and the generality of models for single groups.

## 8.2 Continuous Exponential Families for Randomly Equivalent Groups

To define a general continuous exponential family, consider a bounded real interval  $C$  with at least two points. Let  $K$  be a positive integer, and let  $\mathbf{u}$  be a bounded  $K$ -dimensional integrable function on  $C$ . In most applications in this chapter, given  $C$  and  $K$ ,  $\mathbf{u}$  is  $\mathbf{v}(K, C)$ , where  $\mathbf{v}(K, C)$  has coordinates  $v_k(C)$ ,  $1 \leq k \leq K$ ;  $v_k(C)$ ,  $k \geq 0$ , is a polynomial of degree  $k$  on  $C$  with values  $v_k(x, C)$  for  $x$  in  $C$ ; and, for a uniformly distributed random variable  $U_C$  on the interval  $C$ , the  $v_k(C)$  satisfy the orthogonality constraints

$$E(v_i(U_C, C)v_k(U_C, C)) = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases}$$

for integers  $i$  and  $k$ ,  $1 \leq i \leq k \leq K$ . Computation of the  $v_k(C)$  is discussed in the Appendix. The convention is used in the definition of  $v_k(x, C)$  that the coefficient of  $x^k$  is positive.

The definition of a continuous exponential family is simplified by use of standard vector inner products. For  $K$ -dimensional vectors  $\mathbf{y}$  and  $\mathbf{z}$  with respective coordinates  $y_k$  and  $z_k$ ,  $1 \leq k \leq K$ , let  $\mathbf{y}'\mathbf{z}$  be  $\sum_{k=1}^K y_k z_k$ . To any  $K$ -dimensional vector  $\mathbf{y}$  corresponds a random variable  $Y(\mathbf{y}, \mathbf{u})$  with values in  $C$  with a density function  $f(Y(\mathbf{y}, \mathbf{u}))$  equal to

$$g(\mathbf{y}, \mathbf{u}) = \gamma(\mathbf{y}, \mathbf{u})\exp(\mathbf{y}'\mathbf{u}), \tag{8.1}$$

where

$$1/\gamma(\mathbf{y}, \mathbf{u}) = \int_C \exp(\mathbf{y}'\mathbf{u})$$

(Gilula & Haberman, 2000) and  $g(\mathbf{y}, \mathbf{u})$  has value  $g(z, \mathbf{y}, \mathbf{u})$  at  $z$  in  $C$ . The family of distributions with densities  $g(\mathbf{y}, \mathbf{u})$  for  $K$ -dimensional vectors  $\mathbf{y}$  is the continuous exponential family of distributions defined by  $\mathbf{u}$ . A fundamental characteristic

of these distributions is that they have positive density functions on  $C$ . These density functions are continuous if  $\mathbf{u}$  is continuous. In all cases, if  $0_K$  denotes the  $K$ -dimensional vector with all coordinates 0, then  $Y(0_K, \mathbf{u})$  has the same uniform distribution on  $C$  as does  $U_C$ . If  $\mathbf{u} = \mathbf{v}_2(C)$  and  $y_2 < 0$ , then  $Y(\mathbf{y}, \mathbf{u})$  is distributed as a normal random variable  $X$ , conditional on  $X$  being in  $C$ . To ensure that all distributions in the continuous exponential family are distinct, it is assumed that the covariance matrix of  $\mathbf{u}(U_C)$  is positive definite. Under this condition, the covariance matrix  $\mathbf{V}(\mathbf{y}, \mathbf{u})$  of  $\mathbf{u}(Y(\mathbf{y}, \mathbf{u}))$  is positive definite for each  $\mathbf{y}$ . As a consequence of the fundamental theorem of algebra, this condition on the covariance matrix of  $\mathbf{u}(U_C)$  holds in the polynomial case of  $\mathbf{u} = \mathbf{v}(K, C)$ .

For continuous exponential families, distribution functions are easily constructed and are strictly increasing and continuous on  $C$ . Let the indicator function  $\chi_C(x)$  be the real function on  $C$  such that  $\chi_C(x)$  has value  $\chi_C(z, x) = 1$  for  $z \leq x$  and value  $\chi_C(z, x) = 0$  for  $z > x$ . Then the restriction of the distribution function  $F(Y(\mathbf{y}, \mathbf{u}))$  of  $Y(\mathbf{y}, \mathbf{u})$  to  $C$  is  $G(\mathbf{y}, \mathbf{u})$ , where, for  $x$  in  $C$ ,  $G(\mathbf{y}, \mathbf{u})$  has value

$$G(x, \mathbf{y}, \mathbf{u}) = \int_C \chi_C(x) g(\mathbf{y}, \mathbf{u}).$$

As in Gilula and Haberman (2000), the distribution of a random variable  $Z$  with values in  $C$  may be approximated by a distribution in the continuous exponential family of distributions generated by  $\mathbf{u}$ . The quality of the approximation provided by the distribution with density  $g(\mathbf{y}, \mathbf{u})$  is assessed by the expected log penalty

$$H(Z, \mathbf{y}, \mathbf{u}) = -E(\log g(Z, \mathbf{y}, \mathbf{u})) = -\log \gamma(\mathbf{y}, \mathbf{u}) + \mathbf{y}'E(\mathbf{u}(Z)). \quad (8.2)$$

The smaller the value of  $H(Z, \mathbf{y}, \mathbf{u})$ , the better is the approximation.

Several rationales can be considered for use of the expected logarithmic penalty  $H(Z, \mathbf{y}, \mathbf{u})$ , according to Gilula and Haberman (2000). Consider a probabilistic prediction of  $Z$  by use of a positive density function  $h$  on  $C$ . If  $Z = z$ , then let a log penalty of  $-\log h(z)$  be assigned. If  $-\log f(Z)$  has a finite expectation, then the expected log penalty is  $H(h) = E(-\log h(Z))$ . If  $Z$  is continuous and has positive density  $f$  and if the expectation of  $-\log f(Z)$  is finite, then  $I(Z) = H(f) \leq H(h)$ , so that the optimal probability prediction is obtained with the actual density of  $Z$ . In addition,  $H(h) = I(Z)$  only if  $f$  is a density function of  $Z$ . This feature in which the penalty is determined by the value of the density at the observed value of  $Z$  and the expected penalty is minimized by selection of the density  $f$  of  $Z$  is only encountered if the penalty from use of the density function  $h$  is of the form  $a - b \log h(z)$  for  $Z = z$  for some real constants  $a$  and  $b > 0$ .

This rationale is not applicable if  $Z$  is discrete. In general, if  $Z$  is discrete, then the smallest possible expected log penalty  $E(-\log h(Z))$  is  $-\infty$ , for, given any real  $c > 0$ ,  $h$  may be defined so that  $h(Z) = c$  with probability 1 and the expected log penalty is  $-\log c$ . The constant  $c$  may be arbitrarily large, so the expected log penalty may be arbitrarily small. Nonetheless, the criterion  $E(-\log h(Z))$  cannot be made arbitrarily small if adequate constraints are imposed on  $h$ . In this section, the

requirement that the density function used for prediction of  $Z$  is  $g(\mathbf{y}, \mathbf{u})$  from Equation 8.1 suffices to ensure existence of a finite infimum  $I(Z, \mathbf{u})$  of the expected log penalty  $H(Z, \mathbf{y}, \mathbf{u})$  from Equation 8.2 over all  $\mathbf{y}$ .

Provided that  $\text{Cov}(\mathbf{u}(Z))$  is positive definite, a unique  $K$ -dimensional vector  $\boldsymbol{\theta}(Z, \mathbf{u})$  with coordinates  $\theta_k(Z, \mathbf{u})$ ,  $1 \leq k \leq K$ , exists such that  $H(Z, \boldsymbol{\theta}(Z, \mathbf{u}))$  is equal to the infimum  $I(Z, \mathbf{u})$  of  $H(Z, \mathbf{y}, \mathbf{u})$  for  $K$ -dimensional vectors  $\mathbf{y}$ . Let  $Y_*(Z, \mathbf{u}) = Y(\boldsymbol{\theta}(Z, \mathbf{u}), \mathbf{u})$ . Then  $\boldsymbol{\theta}(Z, \mathbf{u})$  is the unique solution of the equation

$$E(\mathbf{u}(Y_*(Z, \mathbf{u}))) = E(\mathbf{u}(Z)).$$

The left-hand side of the equation is readily expressed in terms of an integral. If

$$\boldsymbol{\mu}(\mathbf{y}, \mathbf{u}) = \int_C \mathbf{u}g(\mathbf{y}, \mathbf{u}),$$

then

$$E(\mathbf{u}(Y(\mathbf{y}, \mathbf{u}))) = \boldsymbol{\mu}(\mathbf{y}, \mathbf{u}).$$

Thus,

$$\boldsymbol{\mu}(\boldsymbol{\theta}(Z, \mathbf{u}), \mathbf{u}) = E(\mathbf{u}(Z)).$$

In the polynomial case of  $\mathbf{u} = \mathbf{v}(K, C)$ , the fundamental theorem of calculus implies that  $\text{Cov}(\mathbf{u}(Z))$  is positive definite, unless a finite set  $C_0$  with no more than  $K$  points exists such that  $P(Z \in C_0) = 1$ . In addition, the moment constraints  $E([Y_*(Z, \mathbf{v}(K, C))]^k) = E(Z^k)$  hold for  $1 \leq k \leq K$ .

The value of  $\boldsymbol{\theta}(Z, \mathbf{u})$  may be found by the Newton-Raphson algorithm. Let

$$\mathbf{V}(\mathbf{y}, \mathbf{u}) = \int_C [\mathbf{u} - \boldsymbol{\mu}(\mathbf{y}, \mathbf{u})][\mathbf{u} - \boldsymbol{\mu}(\mathbf{y}, \mathbf{u})]'g(\mathbf{y}, \mathbf{u}).$$

Given an initial approximation  $\boldsymbol{\theta}_0$ , the algorithm at step  $t \geq 0$  yields a new approximation

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + [\mathbf{V}(\boldsymbol{\theta}_t, \mathbf{u})]^{-1}[E(\mathbf{u}(Z)) - \boldsymbol{\mu}(\boldsymbol{\theta}_t, \mathbf{u})] \quad (8.3)$$

of  $\boldsymbol{\theta}(Z, \mathbf{u})$ . Normally,  $\boldsymbol{\theta}_t$  converges to  $\boldsymbol{\theta}(Z, \mathbf{u})$  as  $t$  increases. The selection of  $\boldsymbol{\theta}_0 = \mathbf{0}_K$  is normally acceptable. When  $\mathbf{u}$  is infinitely differentiable with bounded derivatives of all orders on  $C$ , use of Gauss-Legendre integration facilitates numerical work (Abramowitz & Stegun, 1965, p. 887).

### 8.2.1 Estimation of Parameters

Estimation of  $\boldsymbol{\theta}(Z, \mathbf{u})$  is straightforward. Let  $U_n$  be a uniformly distributed random variable on the integers 1 to  $n$ . For any  $n$ -dimensional vector  $\mathbf{z}$  of real numbers, let

$c(\mathbf{z})$  be the real random variable such that  $c(\mathbf{z})$  has value  $z_i$  if  $U_n = i$ ,  $1 \leq i \leq n$ . Thus  $E(\mathbf{u}(c(\mathbf{z})))$  is the sample average  $n^{-1} \sum_{i=1}^n \mathbf{u}(z_i)$ . Let  $Z_i$ ,  $1 \leq i \leq n$ , be independent and identically distributed random variables with common distribution  $Z$ , and let  $\mathbf{Z}$  be the  $n$ -dimensional vector with coordinates  $Z_i$ ,  $1 \leq i \leq n$ . Then  $\boldsymbol{\theta}(Z, \mathbf{u})$  has estimate  $\boldsymbol{\theta}(c(\mathbf{Z}), \mathbf{u})$  whenever the sample covariance matrix of the  $\mathbf{u}(Z_i)$ ,  $1 \leq i \leq n$ , is positive definite. The estimate  $\boldsymbol{\theta}(c(\mathbf{Z}), \mathbf{u})$  converges to  $\boldsymbol{\theta}(Z, \mathbf{u})$  with probability 1 as  $n$  approaches  $\infty$ . If  $\mathbf{V}_*(Z, \mathbf{u})$  is  $\mathbf{V}(\boldsymbol{\theta}(Z, \mathbf{u}), \mathbf{u})$ , then  $n^{1/2}[\boldsymbol{\theta}(c(\mathbf{Z}), \mathbf{u}) - \boldsymbol{\theta}(Z, \mathbf{u})]$  converges in distribution to a multivariate normal random variable with zero mean and with covariance matrix

$$\Sigma(Z, \mathbf{u}) = [\mathbf{V}_*(Z, \mathbf{u})]^{-1} \text{Cov}(\mathbf{u}(Z)) [\mathbf{V}_*(Z, \mathbf{u})]^{-1}. \quad (8.4)$$

The estimate  $\Sigma(c(\mathbf{Z}), \mathbf{u})$  converges to  $\Sigma(Z, \mathbf{u})$  with probability 1 as the sample size  $n$  increases. For any  $K$ -dimensional vector  $\mathbf{d}$  that is not equal to  $0_K$ , approximate confidence intervals for  $\mathbf{d}'\boldsymbol{\theta}(Z, \mathbf{u})$  may be based on the observation that

$$\frac{\mathbf{d}'\boldsymbol{\theta}(c(\mathbf{Z}), \mathbf{u}) - \mathbf{d}'\boldsymbol{\theta}(Z, \mathbf{u})}{[\mathbf{d}'\Sigma(c(\mathbf{Z}), \mathbf{u})\mathbf{d}/n]^{1/2}}$$

converges in distribution to a standard normal random variable. The denominator  $[\mathbf{d}'\Sigma(c(\mathbf{Z}), \mathbf{u})\mathbf{d}/n]^{1/2}$  may be termed the EASD of  $\mathbf{d}'\boldsymbol{\theta}(c(\mathbf{Z}), \mathbf{u})$ .

The estimate  $I(c(\mathbf{Z}), \mathbf{u})$  converges to  $I(Z, \mathbf{u})$  with probability 1, and  $n^{1/2}[I(c(\mathbf{Z}), \mathbf{u}) - I(Z, \mathbf{u})]$  converges in distribution to a normal random variable with mean 0 and variance  $\sigma^2(I, Z, \mathbf{u})$  equal to the variance of  $\boldsymbol{\theta}(Z, \mathbf{u})/\text{Cov}(\mathbf{u}(Z))/\boldsymbol{\theta}(Z, \mathbf{u})$  of  $\log g(Z, \boldsymbol{\theta}(Z, \mathbf{u}), \mathbf{u})$ . As the sample size  $n$  increases,  $\sigma^2(I, c(\mathbf{Z}), \mathbf{u})$  converges to  $\sigma^2(I, Z, \mathbf{u})$  with probability 1, so that the EASD of  $I(c(\mathbf{Z}), \mathbf{u})$  is  $[\sigma^2(I, c(\mathbf{Z}), \mathbf{u})/n]^{1/2}$ .

The estimated distribution function  $F_*(x, c(\mathbf{Z}), \mathbf{u}) = F(x, Y_*(c(\mathbf{Z}), \mathbf{u}), \mathbf{u})$  converges with probability 1 to  $F_*(x, Z, \mathbf{u}) = F(x, Y_*(Z, \mathbf{u}))$  for each  $x$  in  $C$ , and the estimated quantile function  $Q_*(p, c(\mathbf{Z}), \mathbf{u}) = Q(p, Y_*(c(\mathbf{Z}), \mathbf{u}))$  converges with probability 1 to  $Q_*(p, Z, \mathbf{u}) = Q(p, Y_*(Z, \mathbf{u}))$  for  $0 < p < 1$ . The scaled difference  $n^{1/2}[F_*(x, c(\mathbf{Z}), \mathbf{u}) - F_*(x, Z, \mathbf{u})]$  converges in distribution to a normal random variable with mean 0 and variance

$$\sigma^2(F_*, x, Z, \mathbf{u}) = [\mathbf{T}_*(x, Z, \mathbf{u})]^t \Sigma(Z, \mathbf{u}) [\mathbf{T}_*(x, Z, \mathbf{u})],$$

where

$$\mathbf{T}_*(x, Z, \mathbf{u}) = \mathbf{T}(x, \boldsymbol{\theta}(Z, \mathbf{u}), \mathbf{u})$$

and

$$\mathbf{T}(x, \mathbf{y}, \mathbf{u}) = \int_C \chi_C(x) [\mathbf{u} - \mu(\mathbf{y}, \mathbf{u})] g(\mathbf{y}, \mathbf{u}).$$

The estimate  $\sigma^2(F_*, x, c(\mathbf{Z}), \mathbf{u})$  converges with probability 1 to  $\sigma^2(F_*, x, Z, \mathbf{u})$ . Let  $g_*(Z, \mathbf{u}) = g(\boldsymbol{\theta}(Z, \mathbf{u}), \mathbf{u})$  have value  $g_*(x, Z, \mathbf{u})$  at  $x$  in  $C$ . If  $\mathbf{u}$  is continuous at  $Q_*(p, Z, \mathbf{u})$ , then  $n^{1/2}[Q_*(p, c(\mathbf{Z}), \mathbf{u}) - Q_*(p, Z, \mathbf{u})]$  converges in distribution to a normal random variable with mean 0 and variance

$$\sigma^2(Q_*, p, Z, \mathbf{u}) = \frac{\sigma^2(F_*, Q_*(p, Z, \mathbf{u}), Z, \mathbf{u})}{[g_*(Q_*(p, Z, \mathbf{u}), Z, \mathbf{u})]^2}.$$

The estimate  $\sigma^2(Q_*, p, c(\mathbf{Z}), \mathbf{u})$  converges with probability 1 to  $\sigma^2(Q_*, p, Z, \mathbf{u})$  as  $n$  increases. Note that if  $\mathbf{u}$  is  $\mathbf{v}(K, C)$ , then the continuity requirement always holds.

### 8.2.2 Equating Functions for Continuous Exponential Families

In the equating application, for  $j$  equals 1 or 2, let  $K_j$  be a positive integer, and let  $\mathbf{u}_j$  be a bounded,  $K_j$ -dimensional, integrable function on  $B_j$ . Let  $\text{Cov}(\mathbf{u}_j(X_j))$  be positive definite. Then one may consider the conversion function

$$e_*(X_1, \mathbf{u}_1, X_2, \mathbf{u}_2) = e(Y_*(X_1, \mathbf{u}_1), Y_*(X_2, \mathbf{u}_2))$$

for conversion from Form 1 to Form 2 and the conversion function

$$e_*(X_2, \mathbf{u}_2, X_1, \mathbf{u}_1) = e(Y_*(X_2, \mathbf{u}_2), Y_*(X_1, \mathbf{u}_1))$$

for conversion from Form 2 to Form 1. For  $x$  in  $B_1$ , let  $e_*(X_1, \mathbf{u}_1, X_2, \mathbf{u}_2)$  have value  $e_*(x, X_1, \mathbf{u}_1, X_2, \mathbf{u}_2)$ . For  $x$  in  $B_2$ , let  $e_*(X_2, \mathbf{u}_2, X_1, \mathbf{u}_1)$  have value  $e_*(x, X_2, \mathbf{u}_2, X_1, \mathbf{u}_1)$ .

Given the available random sample data  $X_{ij}$ ,  $1 \leq i \leq n_j$ ,  $1 \leq j \leq 2$ , estimation of the conversion functions is straightforward. Let  $\mathbf{X}_j$  be the  $n_j$ -dimensional vector with coordinates  $X_{ij}$ ,  $1 \leq i \leq n_j$ . Then  $e_*(x, c(\mathbf{X}_1), \mathbf{u}_1, c(\mathbf{X}_2), \mathbf{u}_2)$  converges with probability 1 to  $e_*(x, X_1, \mathbf{u}_1, X_2, \mathbf{u}_2)$  for  $x$  in  $B_1$ , and  $e_*(x, c(\mathbf{X}_2), \mathbf{u}_2, c(\mathbf{X}_1), \mathbf{u}_1)$  converges with probability 1 to  $e_*(x, X_2, \mathbf{u}_2, X_1, \mathbf{u}_1)$  for  $x$  in  $B_2$  as  $n_1$  and  $n_2$  approach  $\infty$ . If  $\mathbf{u}$  is continuous at  $e_*(x, X_1, \mathbf{u}_1, X_2, \mathbf{u}_2)$  for an  $x$  in  $B_1$ , then

$$\frac{e_*(x, c(\mathbf{X}_1), \mathbf{u}_1, c(\mathbf{X}_2), X_2, \mathbf{u}_2) - e_*(x, X_1, \mathbf{u}_1, X_2, \mathbf{u}_2)}{\sigma(e_*, x, X_1, \mathbf{u}_1, X_2, \mathbf{u}_2, n_1, n_2)}$$

converges in distribution to a standard normal random variable as  $n_1$  and  $n_2$  become large, as in Equation 8.4, where

$$\begin{aligned} \sigma^2(e_*, x, X_1, \mathbf{u}_1, X_2, \mathbf{u}_2, n_1, n_2) &= \frac{\sigma^2(F_*, x, X_1, \mathbf{u}_1)}{n_1 [g(e_*(x, X_1, \mathbf{u}_1, X_2, \mathbf{u}_2))]^2} \\ &+ \frac{\sigma^2(Q_*, F_*(x, X_1, \mathbf{u}_1), X_2, \mathbf{u}_2)}{n_2}. \end{aligned} \quad (8.5)$$

In addition, the ratio

$$\frac{\sigma^2(e_*, x, c(\mathbf{X}_1), \mathbf{u}_1, c(\mathbf{X}_2), \mathbf{u}_2, n_1, n_2)}{\sigma^2(e_*, x, X_1, \mathbf{u}_1, X_2, \mathbf{u}_2, n_1, n_2)} \quad (8.6)$$

converges to 1 with probability 1, so that the EASD of  $e_*(x, c(\mathbf{X}_1), \mathbf{u}_1, c(\mathbf{X}_2), \mathbf{u}_2)$  is  $\sigma(e_*, x, c(\mathbf{X}_1), \mathbf{u}_1, c(\mathbf{X}_2), \mathbf{u}_2, n_1, n_2)$ . Similar results apply to conversion from Form 2 to Form 1. Note that continuity requirements always hold in the polynomial case with  $\mathbf{u}_j = \mathbf{v}(K_j, B_j)$ .

### 8.2.3 Example

Table 7.1 of von Davier et al. (2004b) provided two distributions of test scores that are integers from  $c_j = 0$  to  $d_j = 20$ . To illustrate results, the intervals  $B_j = (-0.5, 20.5)$  are employed. Results in terms of estimated expected log penalties are summarized in Table 8.1. These tables suggest that gains over the quadratic case ( $K_j = 2$ ) are very modest for both  $X_1$  and  $X_2$ .

Equating results are provided in Table 8.2 for conversions from Form 1 to Form 2. Given Table 8.1, the quadratic model is considered for both  $X_1$  and  $X_2$ , so that  $K_1 = K_2 = 2$ . Two comparisons are provided with familiar equating procedures. In the case of kernel equating, comparable quadratic log-linear models were used. The bandwidths employed by Version 2.1 of the LOGLIN/KE program (Chen, Yan, Han, & von Davier, 2006) were employed. The percentile-rank computations correspond to the use of the tangent rule of integration in Wang (2008) for the quadratic continuous exponential families. In terms of continuous exponential families, the percentile-rank results also can be produced if  $\mathbf{v}(x, 2, B_j)$  is replaced by the rounded approximation  $\mathbf{v}(\text{rnd}(x), 2, B_j)$ , where  $\text{rnd}(x)$  is the nearest integer to  $x$ . The convention to adopt for the definition of  $\text{rnd}(x)$  for values such as  $x = 1.5$  has no material effect on the analysis; however, the discontinuity of  $\text{rnd}(x)$  for such values does imply that asymptotic normality approximations are not entirely satisfactory. As a consequence, they are not provided. In this example, the three conversions are very similar for all possible values of  $X_1$ . For the two methods for which EASDs are available, results are rather similar. The results for the continuous exponential family are relatively best at the extremes of the distribution.

**Table 8.1** Estimated Expected Log Penalties for Variables  $X_1$  and  $X_2$  for Polynomial Models

Variable	Degree	Estimate	EASD
$X_1$	2	2.747	0.015
$X_1$	3	2.747	0.015
$X_1$	4	2.745	0.015
$X_2$	2	2.773	0.014
$X_2$	3	2.772	0.014
$X_2$	4	2.771	0.014

Note: EASD = estimated asymptotic standard deviation.

**Table 8.2** Comparison of Conversions From Form 1 to Form 2

Value	Continuous exponential		Kernel		Percentile-rank estimate
	Estimate	EASD	Estimate	EASD	
0	0.091	0.110	-0.061	0.194	0.095
1	1.215	0.209	1.234	0.235	1.179
2	2.304	0.239	2.343	0.253	2.255
3	3.377	0.240	3.413	0.253	3.325
4	4.442	0.230	4.473	0.242	4.392
5	5.504	0.214	5.529	0.225	5.458
6	6.564	0.198	6.582	0.207	6.522
7	7.621	0.182	7.634	0.189	7.585
8	8.677	0.169	8.685	0.174	8.647
9	9.732	0.159	9.734	0.162	9.706
10	10.784	0.155	10.781	0.155	10.761
11	11.834	0.155	11.825	0.153	11.823
12	12.880	0.160	12.865	0.156	12.859
13	13.919	0.168	13.900	0.163	13.898
14	14.950	0.177	14.925	0.172	14.928
15	15.966	0.184	15.936	0.179	15.947
16	16.959	0.187	16.925	0.182	16.949
17	17.912	0.179	17.879	0.178	17.927
18	18.802	0.156	18.799	0.164	18.871
19	19.592	0.109	19.723	0.145	19.760
20	20.240	0.040	20.818	0.119	20.380

Note: EASD = estimated asymptotic standard deviation.

### 8.3 Continuous Exponential Families for a Single Group

In the case of a single group, the joint distribution of  $\mathbf{X} = (X_1, X_2)$  is approximated by use of a bivariate continuous exponential family. The definition of a bivariate continuous exponential family is similar to that for a univariate continuous exponential family. However, in bivariate exponential families, a nonempty, bounded, convex open set  $C$  of the plane is used such that each value of  $\mathbf{X}$  is in a closed subset  $D$  of  $C$ .

As in the univariate case, let  $K$  be a positive integer, and let  $\mathbf{u}$  be a bounded  $K$ -dimensional integrable function on  $C$ . In many cases,  $\mathbf{u}$  is defined by use of bivariate polynomials. To any  $K$ -dimensional vector  $\mathbf{y}$  corresponds a two-dimensional random variable  $\mathbf{Y}(\mathbf{y}, \mathbf{u}) = (Y_1(\mathbf{y}, \mathbf{u}), Y_2(\mathbf{y}, \mathbf{u}))$  with values in  $C$  with a density function  $f(\mathbf{Y}(\mathbf{y}, \mathbf{u}))$  equal to

$$g(\mathbf{y}, \mathbf{u}) = \gamma(\mathbf{y}, \mathbf{u}) \exp(\mathbf{y}'\mathbf{u}),$$

where from Equation 8.1

$$1/\gamma(\mathbf{y}, \mathbf{u}) = \int_C \exp(\mathbf{y}'\mathbf{u})$$

and  $g(\mathbf{y}, \mathbf{u})$  has value  $g(\mathbf{z}, \mathbf{y}, \mathbf{u})$  at  $\mathbf{z}$  in  $C$ . As in the univariate case, the family of distributions with densities  $g(\mathbf{y}, \mathbf{u})$  for  $K$ -dimensional vectors  $\mathbf{y}$  is the continuous exponential family of distributions defined by  $\mathbf{u}$ . These density functions are always positive on  $C$ , and they are continuous if  $\mathbf{u}$  is continuous. If  $\mathbf{U}_C$  is a random vector with a uniform distribution on  $C$ , then problems of parameter identification may be avoided by the requirement that  $\mathbf{u}(\mathbf{U}_C)$  have a positive-definite covariance matrix. As in the univariate case,  $\mathbf{Y}(0_K, \mathbf{u})$  has the same distribution as  $\mathbf{U}_C$ .

In equating applications, marginal distributions are important. For  $j$  equal 1 or 2, let  $C_j$  be the open set that consists of real  $x$  such that  $x = x_j$  for some  $(x_1, x_2)$  in  $C$ . For  $x$  in  $C_j$ , let  $\chi_{jC}(x)$  be the set of  $(x_1, x_2)$  in  $C_1$  with  $x_j \leq x$ . Then the restriction  $G_j(\mathbf{y}, \mathbf{u})$  of the distribution function  $F(Y_j(\mathbf{y}, \mathbf{u}))$  to  $C_j$  has value

$$G_j(x, \mathbf{y}, \mathbf{u}) = \int_C \chi_{jC}(x) g(\mathbf{y}, \mathbf{u})$$

at  $x$  in  $C_j$ . The function  $G_j(\mathbf{y}, \mathbf{u})$  is continuous and strictly increasing on  $C_j$ , and  $Y_j(\mathbf{y}, \mathbf{u})$  has density function  $g_j(x, \mathbf{y}, \mathbf{u})$ . For  $z_1$  in  $C_1$ ,

$$g_1(z_1, \mathbf{y}, \mathbf{u}) = \int_{C_2(z_1)} g(\mathbf{z}, \mathbf{y}, \mathbf{u}) dz_2,$$

where  $C_2(z_1) = \{z_2 \in C_2 : (z_1, z_2) \in C_2\}$ . Similarly,

$$g_2(z_2, \mathbf{y}, \mathbf{u}) = \int_{C_1(z_2)} g(\mathbf{z}, \mathbf{y}, \mathbf{u}) dz_1,$$

where  $C_1(z_2) = \{z_1 \in C_1 : (z_1, z_2) \in C_1\}$ . The function  $G_j(\mathbf{y}, \mathbf{u})$  is continuously differentiable if  $\mathbf{u}$  is continuous.

As in the univariate case, the distribution of the random vector  $\mathbf{X}$  with values in  $C$  may be approximated by a distribution in the continuous exponential family of distributions generated by  $\mathbf{u}$ . The quality of the approximation provided by the distribution with density  $g(\mathbf{y}, \mathbf{u})$  is assessed by the expected log penalty from Equation 8.2

$$H(\mathbf{X}, \mathbf{y}, \mathbf{u}) = -E(\log g(\mathbf{X}, \mathbf{y}, \mathbf{u})) = -\log \gamma(\mathbf{y}, \mathbf{u}) + \mathbf{y}'E(\mathbf{u}(\mathbf{Z})).$$

The smaller the value of  $H(\mathbf{X}, \mathbf{y}, \mathbf{u})$ , the better is the approximation.

Provided that  $\text{Cov}(\mathbf{u}(\mathbf{X}))$  is positive definite, a unique  $K$ -dimensional vector  $\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$  with coordinates  $\theta_k(\mathbf{X}, \mathbf{u})$ ,  $1 \leq k \leq K$ , exists such that  $H(\mathbf{X}, \boldsymbol{\theta}(\mathbf{X}, \mathbf{u}))$  is equal to the infimum  $I(\mathbf{X}, \mathbf{u})$  of  $H(\mathbf{X}, \mathbf{y}, \mathbf{u})$  for  $K$ -dimensional vectors  $\mathbf{y}$ . Let  $\mathbf{Y}_*(\mathbf{X}, \mathbf{u}) = (Y_{*1}(\mathbf{X}, \mathbf{u}), Y_{*2}(\mathbf{X}, \mathbf{u})) = \mathbf{Y}(\boldsymbol{\theta}(\mathbf{X}, \mathbf{u}), \mathbf{u})$ . Then  $\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$  is the unique solution of the equation

$$E(\mathbf{u}(\mathbf{Y}_*(\mathbf{X}, \mathbf{u}))) = E(\mathbf{u}(\mathbf{X})).$$

If

$$\boldsymbol{\mu}(\mathbf{y}, \mathbf{u}) = \int_C \mathbf{u}g(\mathbf{y}, \mathbf{u}),$$

then

$$E(\mathbf{u}(\mathbf{Y}(\mathbf{y}, \mathbf{u}))) = \boldsymbol{\mu}(\mathbf{y}, \mathbf{u}),$$

and

$$\boldsymbol{\mu}(\boldsymbol{\theta}(\mathbf{X}, \mathbf{u}), \mathbf{u}) = E(\mathbf{u}(\mathbf{X})).$$

The value of  $\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$  may be found by the Newton-Raphson algorithm. Note that the positive-definite covariance matrix of  $\mathbf{Y}(\mathbf{y}, \mathbf{u})$  is

$$\mathbf{V}(\mathbf{y}, \mathbf{u}) = \int_C [\mathbf{u} - \boldsymbol{\mu}(\mathbf{y}, \mathbf{u})][\mathbf{u} - \boldsymbol{\mu}(\mathbf{y}, \mathbf{u})]^\prime g(\mathbf{y}, \mathbf{u}).$$

Given an initial approximation  $\boldsymbol{\theta}_0$ , the algorithm at step  $t \geq 0$  yields a new approximation

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + [\mathbf{V}(\boldsymbol{\theta}_t, \mathbf{u})]^{-1}[E(\mathbf{u}(\mathbf{X})) - \boldsymbol{\mu}(\boldsymbol{\theta}_t, \mathbf{u})]$$

of  $\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$  as in Equation 8.3. Normally,  $\boldsymbol{\theta}_t$  converges to  $\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$  as  $t$  increases. As in the univariate case, the selection of  $\boldsymbol{\theta}_0 = 0_K$  is normally acceptable.

### 8.3.1 Estimation of Parameters

Estimation of  $\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$  is quite similar to the corresponding estimation in the univariate case. For any  $n$ -by-2 real matrix  $\mathbf{z}$  with elements  $z_{ij}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq 2$ , let  $\mathbf{c}(\mathbf{z})$  be the two-dimensional random variable such that  $\mathbf{c}(\mathbf{z})$  has value  $(z_{i1}, z_{i2})$  if  $U_n = i$ ,  $1 \leq i \leq n$ . Let  $\mathbf{Z}$  be the  $n$ -by-2 matrix with elements  $X_{ij}$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq 2$ . Then  $\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$  has estimate  $\boldsymbol{\theta}(\mathbf{c}(\mathbf{Z}), \mathbf{u})$  whenever the sample covariance matrix of the  $\mathbf{u}(\mathbf{X}_i)$ ,  $1 \leq i \leq n$ , is positive definite. The estimate  $\boldsymbol{\theta}(\mathbf{c}(\mathbf{Z}), \mathbf{u})$  converges to  $\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$  with probability 1 as  $n$  approaches  $\infty$ . If  $\mathbf{V}_*(\mathbf{X}, \mathbf{u})$  is  $\mathbf{V}(\boldsymbol{\theta}(\mathbf{X}, \mathbf{u}), \mathbf{u})$ , then  $n^{1/2}[\boldsymbol{\theta}(\mathbf{c}(\mathbf{Z}), \mathbf{u}) - \boldsymbol{\theta}(\mathbf{X}, \mathbf{u})]$  converges in distribution to a multivariate normal random variable with zero mean and with covariance matrix

$$\Sigma(\mathbf{X}, \mathbf{u}) = [\mathbf{V}_*(\mathbf{X}, \mathbf{u})]^{-1} \text{Cov}(\mathbf{u}(\mathbf{X})) [\mathbf{V}_*(\mathbf{X}, \mathbf{u})]^{-1},$$

as in Equation 8.4.

The estimate  $\Sigma(\mathbf{c}(\mathbf{Z}), \mathbf{u})$  converges to  $\Sigma(\mathbf{X}, \mathbf{u})$  with probability 1 as the sample size  $n$  increases. For any  $K$ -dimensional vector  $\mathbf{d}$  that is not equal to  $0_K$ ,

approximate confidence intervals for  $\mathbf{d}'\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$  may be based on the observation that

$$\frac{\mathbf{d}'\boldsymbol{\theta}(\mathbf{c}(\mathbf{Z}), \mathbf{u})\mathbf{d} - \mathbf{d}'\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})}{[\mathbf{d}'\boldsymbol{\Sigma}(\mathbf{c}(\mathbf{Z}), \mathbf{u})\mathbf{d}/n]^{1/2}}$$

converges in distribution to a standard normal random variable. The denominator  $[\mathbf{d}'\boldsymbol{\Sigma}(\mathbf{c}(\mathbf{Z}), \mathbf{u})\mathbf{d}/n]^{1/2}$  may be termed the EASD of  $\mathbf{d}'\boldsymbol{\theta}(\mathbf{c}(\mathbf{Z}), \mathbf{u})$ .

The estimate  $I(\mathbf{c}(\mathbf{Z}), \mathbf{u})$  converges to  $I(\mathbf{X}, \mathbf{u})$  with probability 1, and  $n^{1/2}[I(\mathbf{c}(\mathbf{Z}), \mathbf{u}) - I(\mathbf{X}, \mathbf{u})]$  converges in distribution to a normal random variable with mean 0 and variance  $\sigma^2(I, \mathbf{X}, \mathbf{u})$  equal to the variance of  $\log g(\mathbf{X}, \boldsymbol{\theta}(\mathbf{X}, \mathbf{u}), \mathbf{u})$ . This variance is  $[\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})]'\text{COV}(\mathbf{u}(\mathbf{X}))\boldsymbol{\theta}(\mathbf{X}, \mathbf{u})$ . As the sample size  $n$  increases,  $\sigma^2(I, \mathbf{c}(\mathbf{Z}), \mathbf{u})$  converges to  $\sigma^2(I, \mathbf{X}, \mathbf{u})$  with probability 1, so that the EASD of  $I(\mathbf{c}(\mathbf{Z}), \mathbf{u})$  is  $[\sigma^2(I, \mathbf{c}(\mathbf{Z}), \mathbf{u})/n]^{1/2}$ .

For  $x$  in  $C_j$  and  $j$  equal 1 or 2,  $F_{j*}(x, \mathbf{c}(\mathbf{Z}), \mathbf{u}) = F(x, Y_{j*}(\mathbf{c}(\mathbf{Z}), \mathbf{u}), \mathbf{u})$  converges to  $F_{j*}(x, \mathbf{X}, \mathbf{u})$  with probability 1 for each  $x$  in  $C$ . Thus, the estimated quantile function  $Q_{*j}(p, \mathbf{c}(\mathbf{Z}), \mathbf{u}) = Q(p, Y_{*j}(\mathbf{c}(\mathbf{Z}), \mathbf{u}))$  converges with probability 1 to  $Q_{*j}(p, \mathbf{X}, \mathbf{u})$  for  $0 < p < 1$ . The scaled difference  $n^{1/2}[F_{*j}(x, \mathbf{c}(\mathbf{Z}), \mathbf{u}) - F_{*j}(x, \mathbf{X}, \mathbf{u})]$  converges in distribution to a normal random variable with mean 0 and variance

$$\sigma^2(F_{*j}, x, \mathbf{X}, \mathbf{u}) = [\mathbf{T}_{*j}(x, \mathbf{X}, \mathbf{u})]'\boldsymbol{\Sigma}(\mathbf{X}, \mathbf{u})[\mathbf{T}_{*j}(x, \mathbf{X}, \mathbf{u})],$$

where

$$\mathbf{T}_{*j}(x, Z, \mathbf{u}) = \mathbf{T}_j(x, \boldsymbol{\theta}(\mathbf{X}, \mathbf{u}), \mathbf{u})$$

and

$$\mathbf{T}_j(x, \mathbf{y}, \mathbf{u}) = \int_C \chi_{jC}(x)[\mathbf{u} - \mu(\mathbf{y}, \mathbf{u})]g(\mathbf{y}, \mathbf{u}).$$

The estimate  $\sigma^2(F_{*j}, x, \mathbf{c}(\mathbf{Z}), \mathbf{u})$  converges with probability 1 to  $\sigma^2(F_{*j}, x, \mathbf{X}, \mathbf{u})$ . Let  $g_{*j}(\mathbf{X}, \mathbf{u}) = g(\boldsymbol{\theta}(\mathbf{X}, \mathbf{u}), \mathbf{u})$  have value  $g_{*j}(x, \mathbf{X}, \mathbf{u})$  at  $x$  in  $C_j$ . If  $\mathbf{u}$  is continuous, then  $n^{1/2}[Q_{*j}(p, \mathbf{c}(\mathbf{Z}), \mathbf{u}) - Q_{*j}(p, \mathbf{X}, \mathbf{u})]$  converges in distribution to a normal random variable with mean 0 and variance

$$\sigma^2(Q_{*j}, p, \mathbf{X}, \mathbf{u}) = \frac{\sigma^2(F_{*j}, Q_{*j}(p, \mathbf{X}, \mathbf{u}), \mathbf{X}, \mathbf{u})}{[g_{*j}(Q_{*j}(p, \mathbf{X}, \mathbf{u}), \mathbf{X}, \mathbf{u})]^2}.$$

The estimate  $\sigma^2(Q_{*j}, p, \mathbf{c}(\mathbf{Z}), \mathbf{u})$  converges with probability 1 to  $\sigma^2(Q_{*j}, p, \mathbf{X}, \mathbf{u})$  as  $n$  increases.

### 8.3.1.1 Conversion Functions

The conversion function  $e_{*1}(\mathbf{X}, \mathbf{u}) = e(Y_{*1}(\mathbf{X}, \mathbf{u}), Y_{*2}(\mathbf{X}, \mathbf{u}))$  may be used for conversion from Form 1 to Form 2. The conversion function  $e_{*2}(\mathbf{X}, \mathbf{u}) = e(Y_{*2}(\mathbf{X}, \mathbf{u}), Y_{*1}(\mathbf{X}, \mathbf{u}))$  may be used for conversion from Form 2 to Form 1. For  $x$  in  $C_j$ , let  $e_{*j}(\mathbf{X}, \mathbf{u})$  have value  $e_{*j}(x, \mathbf{X}, \mathbf{u})$ . Then  $e_{*j}(x, \mathbf{c}(\mathbf{Z}), \mathbf{u})$  converges with probability 1 to  $e_{*j}(x, \mathbf{Z}, \mathbf{u})$ . If  $\mathbf{u}$  is continuous and  $h = 3 - j$ , so that  $h = 1$  if  $j = 2$  and  $h = 2$  if  $j = 1$ , then  $n^{1/2}[e_{*j}(x, \mathbf{c}(\mathbf{Z}), \mathbf{u}) - e_{*j}(x, \mathbf{X}, \mathbf{u})]$  converges in distribution to a normal random variable with mean 0 and variance

$$\sigma^2(e_{*j}, x, \mathbf{X}, \mathbf{u}) = [\mathbf{T}_{dj}(x, \mathbf{X}, \mathbf{u})]' \Sigma(\mathbf{X}, \mathbf{u}) \mathbf{T}_{dj}(x, \mathbf{X}, \mathbf{u}),$$

as in Equation 8.4 and 8.5, where

$$\mathbf{T}_{dj}(x, \mathbf{X}, \mathbf{u}) = \mathbf{T}_{*j}(x, \mathbf{X}, \mathbf{u}) - \mathbf{T}_{*h}(e_{*j}(x, \mathbf{X}, \mathbf{u}), \mathbf{X}, \mathbf{u}).$$

The estimate  $\sigma^2(e_{*j}, x, \mathbf{c}(\mathbf{Z}), \mathbf{u})$  converges with probability 1 to  $\sigma^2(e_{*j}, x, \mathbf{X}, \mathbf{u})$ .

### 8.3.1.2 Polynomials

In the simplest case,  $C$  is the Cartesian product  $B_1 \times B_2$ , so that  $C$  consists of all pairs  $(x_1, x_2)$  such that each  $x_j$  is in  $B_j$ . One common case has  $K_j \geq 2$  and  $\mathbf{u} = \mathbf{v}(K_1, K_2, B_1, B_2)$ , where, for  $\mathbf{x} = (x_1, x_2)$  in  $C$ , coordinate  $k$  of  $\mathbf{v}(K_1, K_2, B_1, B_2)$  has value  $v_k(\mathbf{x}, K_1, K_2, B_1, B_2) = v_k(x_1, K_1, B_1)$  for  $1 \leq k \leq K_1$ , coordinate  $K_1 + k$  has value  $v_k(x_2, K_2, B_2)$  for  $1 \leq k \leq K_2$ , and coordinate  $k = K_1 + K_2 + 1$  is  $v_1(x_1, K_1, B_1)v_2(x_2, K_2, B_2)$ . For this definition of  $\mathbf{u}$ ,  $\mathbf{u}$  is continuous, so that all normal approximations apply. For the marginal variable  $Y_{*j}(\mathbf{X}, \mathbf{u})$ , the first  $K_j$  moments are the same as the corresponding moments of  $X_j$ . In addition, the covariance of  $Y_{*1}(\mathbf{X}, \mathbf{u})$  and  $Y_{*2}(\mathbf{X}, \mathbf{u})$  is the same as the covariance of  $X_1$  and  $X_2$ . If  $K_1 = K_2 = 2$  and if  $\theta_2(\mathbf{X}, \mathbf{u})$  and  $\theta_4(\mathbf{X}, \mathbf{u})$  are negative, then  $\mathbf{X}$  is distributed as the conditional distribution of a bivariate normal vector given that the vector is in  $C$ . Other choices are possible. For example, Wang (2008) considered a case with  $K = K_1K_2$ ,  $C = B_1 \times B_2$ , and with each coordinate of  $\mathbf{u}(\mathbf{x})$  a product  $v(x_1, k_1, B_1)v(x_2, k_2, B_2)$  for  $k_j$  from 1 to  $K_j$ . If  $\mathbf{u}$  is the vector with the first  $K_1 + K_2$  coordinates of  $\mathbf{v}(K_1, K_2, B_1, B_2)$ , then it is readily seen that  $e_{*j}(x, \mathbf{X}, \mathbf{u})$  is the same as the conversion function  $e_*(x, X_j, \mathbf{v}(K_j, B_j), X_h, \mathbf{v}(K_h, B_h))$  from the case of equivalent groups, although the use of single groups typically leads to a different normal approximation for  $e_{*j}(x, \mathbf{c}(\mathbf{Z}), \mathbf{u})$  than the normal approximation for  $e_*(x, c(\mathbf{X}_j), \mathbf{v}(K_j, B_j), c(\mathbf{X}_h), \mathbf{v}(K_h, B_h))$ .

### 8.3.2 Example

Table 8.2 of von Davier et al. (2004b) provided an example of a single-group design with  $c_j = 0$  and  $d_j = 20$  for  $1 \leq j \leq 2$ . To illustrate results, let

$B_1 = B_2 = (-0.5, 20.5)$ ,  $C = B_1 \times B_2$ , and  $\mathbf{u} = \mathbf{v}(K, K, B_1, B_2)$ ,  $2 \leq K \leq 4$ . Results in terms of estimated expected log penalties are summarized in Table 8.3. These results suggest that gains beyond the quadratic case are quite small, although the quartic case differs from the cubic case more than the cubic case differs from the quadratic case.

Not surprisingly, the three choices of  $K$  lead to rather similar conversion functions. Consider Table 8.4 for the case of conversion of Form 1 to Form 2. A bit more variability in results exists for very high or very low values, although estimated asymptotic standard deviations are more variable than are estimated conversions. Note that results are also similar to those for kernel equating (von Davier et al., 2004b, Ch. 8) shown in Table 8.5. These results employ a log-linear model for the joint distribution of the scores, which is comparable to the model defined by  $K = 3$  for a continuous exponential family. The log-linear fit preserves the initial three marginal moments for each score distribution as well as the covariance of the two scores. As a consequence, the marginal distributions produced by the kernel method have the same means and variances as do the corresponding distributions of  $X_{i1}$  and

**Table 8.3** Estimated Expected Log Penalties

$K$	Estimate	EASD
2	4.969	0.022
3	4.968	0.022
4	4.960	0.022

*Note:* EASD = estimated asymptotic standard deviation.

**Table 8.4** Comparison of Conversions From Form 1 to Form 2

Value	$K = 2$		$K = 3$		$K = 4$	
	Estimate	EASD	Estimate	EASD	Estimate	EASD
0	0.111	0.077	-0.040	0.113	0.404	0.262
1	1.168	0.128	0.927	0.204	1.404	0.264
2	2.144	0.135	1.917	0.208	2.269	0.221
3	3.091	0.130	2.910	0.182	3.121	0.176
4	4.028	0.120	3.899	0.151	3.987	0.140
5	4.959	0.108	4.881	0.122	4.874	0.117
6	5.889	0.097	5.854	0.101	5.785	0.105
7	6.819	0.086	6.819	0.087	6.721	0.100
8	7.748	0.076	7.775	0.080	7.679	0.095
9	8.677	0.069	8.722	0.078	8.653	0.090
10	9.606	0.065	9.661	0.078	9.634	0.086
11	10.536	0.064	10.591	0.077	10.611	0.085
12	11.465	0.066	11.512	0.077	11.574	0.088
13	12.394	0.073	12.425	0.077	12.514	0.092
14	13.324	0.081	13.331	0.081	13.427	0.094
15	14.256	0.091	14.231	0.090	14.310	0.096
16	15.193	0.102	15.128	0.105	15.166	0.101
17	16.141	0.113	16.033	0.126	16.003	0.116
18	17.119	0.121	16.967	0.149	16.838	0.142
19	18.173	0.123	17.985	0.167	17.716	0.174
20	19.495	0.094	19.304	0.150	18.842	0.198

*Note:* EASD = estimated asymptotic standard deviation.

**Table 8.5** Conversions From Form 1 to Form 2 by Kernel Equating

Value	Estimate	EASD
0	-0.002	0.162
1	0.999	0.221
2	1.981	0.221
3	2.956	0.193
4	3.926	0.159
5	4.890	0.128
6	5.850	0.104
7	6.805	0.089
8	7.756	0.080
9	8.702	0.078
10	9.643	0.077
11	10.580	0.077
12	11.512	0.077
13	12.439	0.078
14	13.362	0.083
15	14.283	0.095
16	15.206	0.115
17	16.140	0.140
18	17.105	0.167
19	18.155	0.185
20	19.411	0.158

*Note:* EASD = estimated asymptotic standard deviation.

$X_{i2}$ . However, the kernel methods yields the distribution of a continuous random variable for the first form with a skewness coefficient that is 0.987 times the original skewness coefficient for  $X_{i1}$  and a distribution of a continuous random variable for the second form with a skewness coefficient that is 0.983 times the original skewness coefficient for  $X_{i2}$ .

### 8.4 Conclusions

Equating via continuous exponential families can be regarded as a viable competitor to kernel equating and to the percentile-rank approach. Continuous exponential families lead to simpler procedures and more thorough moment agreement, for fewer steps are involved in equating by continuous exponential families, due to elimination of kernel smoothing. In addition, equating by continuous exponential families does not require selection of bandwidths.

One example does not produce an operational method, and kernel equating is rapidly approaching operational use, so it is important to consider some required steps. Although equivalent-groups designs and single-group designs are used in testing programs, a large fraction of equating designs are more complex. Nonetheless, these designs typically can be explored by repeated application of single-group or equivalent-groups designs. For example, the single-group design provides the basis for more complex linking designs with anchor tests (von Davier et al., 2004b,

Ch. 9). No reason exists to expect that continuous exponential families cannot be applied to any standard equating situation to which kernel equating has been applied.

It is certainly appropriate to consider a variety of applications to data, and some work on quality of large-sample approximations is appropriate when smaller sample sizes are contemplated. Although this gain is not apparent in the examples studied, a possible gain from continuous exponential families is that application to assessments with unevenly spaced scores or very large numbers of possible scores is completely straightforward. Thus, direct conversion from a raw score on one form to an unrounded scale score on a second form involves no difficulties. In addition, in tests with formula scoring, no need exists to round raw scores to integers during equating.

## Chapter 8 Appendix

### 8.A.1 Computation of Orthogonal Polynomials

Computation of the orthogonal polynomials  $v_k(C)$ ,  $k \geq 0$ , is rather straightforward given standard properties of Legendre polynomials (Abramowitz & Stegun, 1965, Ch. 8, 22). The Legendre polynomial of degree 0 is  $P_0(x) = 1$ ; the Legendre polynomial of degree 1 is  $P_1(x) = x$ ; and the Legendre polynomial  $P_{k+1}(x)$  of degree  $k + 1$ ,  $k \geq 1$ , is determined by the recurrence relationship

$$P_{k+1}(x) = (k + 1)^{-1}[(2k + 1)xP_k(x) - kP_{k-1}(x)], \quad (8.A.1)$$

so that  $P_2(x) = (3x^2 - 1)/2$ . For nonnegative integers  $i$  and  $k$ , the integral  $\int_{-1}^1 P_i P_k$  is 0 for  $i \neq k$  and  $1/(2k + 1)$  for  $i = k$ . Use of elementary rules of integration shows that one may let

$$v_k(x, C) = (2k + 1)^{1/2} P_k([2x - \inf(C) - \sup(C)]/[\sup(C) - \inf(C)]).$$

**Author Note:** Any opinions expressed in this chapter are those of the author and not necessarily of Educational Testing Service.