

# Chapter 10

## Equating Through Alternative Kernels

Yi-Hsuan Lee and Alina A. von Davier

### 10.1 Introduction

The need for test equating arises when two or more test forms measure the same construct and can yield different scores for the same examinee. The most common example involves multiple forms of a test within a testing program, as opposed to a single testing instrument. In a testing program, different test forms that are similar in content and format typically contain completely different test items. Consequently, the tests can vary in difficulty depending on the degree of control available in the test development process.

The goal of test equating is to allow the scores on different forms of the same test to be used and interpreted interchangeably. Test equating requires some type of control for differential examinee ability in the assessment of, and adjustment for, differential test difficulty; the differences in abilities are controlled by employing an appropriate data collection design.

Many observed-score equating methods are based on the equipercentile equating function, which requires that the initial, discrete score distribution functions have been continuized. Several important observed-score equating methods may be viewed as differing only in the way the continuization is achieved. The classical equipercentile equating method (i.e., the percentile-rank method) uses linear interpolation to make the discrete distribution piecewise linear and therefore continuous. The kernel equating (von Davier, Holland, & Thayer, 2004b) method uses Gaussian kernel (GK) smoothing to approximate the discrete histogram by a continuous density function.

A five-step process of kernel equating was introduced in von Davier et al. (2004b) for manipulation of raw data from any type of data collection design, either for common examinees (e.g., the equivalent-groups, single-group, and counterbalanced

---

Y.-H. Lee (✉) and A.A. von Davier  
Educational Testing Service, Rosedale Rd, Princeton, New Jersey 08541, USA  
e-mail: ylee@ets.org

designs) or for common items (e.g., the nonequivalent groups with anchor test, or NEAT, design). The five steps are (a) presmoothing of the initial, discrete score distributions using log-linear models (Holland & Thayer, 2000); (b) estimation of the marginal discrete score distributions by applying the design function, which is a mapping reflecting the data-collection design; (c) continuization of the marginal, discrete score distributions; (d) computation and diagnosis of the equating functions; and (e) evaluation of statistical accuracy in terms of standard error of equating (SEE) and standard error of equating difference (SEED). Description of the five-step process can be found in the introductory chapter of this volume.

Kernel smoothing is a means of nonparametric smoothing. It continuizes a discrete random variable  $X$  by adding to it a continuous and independent random variable  $V$  with a positive constant  $h_X$  controlling the degree of smoothness. Let  $X(h_X)$  denote the continuous approximation of  $X$ . Then

$$X(h_X) = X + h_X V. \quad (10.1)$$

The  $h_X$  is the so-called *bandwidth* and is free to select to achieve certain practical purpose. Kernel function refers to the density function of  $V$ . When kernel smoothing was first introduced to test equating by Holland and Thayer (1989),  $V$  was assumed to be a standard normal random variable. As a continuation, the conceptual framework of kernel equating was further established and introduced in von Davier et al. (2004b) with concentration on GK.

Equation 10.1 can be regarded as the central idea of kernel smoothing. In principle, it is feasible to substitute any continuous random variable for the one following a standard normal distribution. The choice of bandwidth  $h_X$  is often believed to be more crucial than the choice of kernel function in kernel regression (Wasserman, 2006); however, knowledge of ordinary kernel smoothing is not necessarily applicable in kernel equating without further justification. One example is the selection of bandwidth. In most applications of observed-score equating, the test scores  $X$  of each test are discrete variables. The selection of  $h_X$  involves a compromise between two features. First, the distribution function of  $X(h_X)$  has to be relatively smooth. Second,  $X(h_X)$  should be a close approximation of  $X$  at each possible score. The common expectation that the constant  $h_X$  approaches 0 as the sample size becomes large should not hold in kernel equating.

A by-product of smoothing is that  $X(h_X)$  will carry not only the characteristics of  $X$  but also the characteristics of  $V$ . Thus, there is cause of concern about the impact of the characteristics of  $V$  on each step of the equating process after continuization. To illustrate, two alternative kernel functions, the logistic kernel (LK) and the (continuous) uniform kernel (UK), will be employed along with the GK. In item-response theory, a logistic function is acknowledged to closely approximate the classical normal-ogive curve with mathematical convenience. It has a simple expression for the cumulative distribution function (CDF), avoiding the integration involved in the CDF of a normal distribution and resolves many theoretical and numerical problems in conjunction with the computation of CDF. The same convenience also advantages the process of kernel equating when deriving formulas

for SEE and SEED. When  $V$  follows a uniform distribution, Equation 10.1 leads to the product of linear interpolation. The inclusion of UK in the framework allows direct comparisons through SEE and SEED between equating results from kernel equating with a specific kernel function and those from the percentile-rank method.

LK shares much common ground with GK with respect to distributional characteristics, with the exception of heavier tails and sharper peak. UK has a finite range and can be viewed as a no-tail distribution. These characteristics can be quantified by moments or cumulants of various orders, and it appears natural to evaluate the continuous approximations of LK, UK, and GK through these measures to see how the distributional properties are inherited.

Some notations are needed before we proceed. Two tests are to be equated, test form  $X$  and test form  $Y$ , and a target population,  $T$ , on which this is to be done. Assume that  $T$  is fixed throughout this chapter. Let  $X$  be the score on test  $X$  and  $Y$  be the score on test  $Y$ , where  $X$  and  $Y$  are random variables. The possible scores of  $X$  and  $Y$  are  $x_i$ ,  $1 \leq i \leq I$ , and  $y_j$ ,  $1 \leq j \leq J$ , respectively. The corresponding score probabilities are  $\mathbf{r} = \{r_i\}_{1 \leq i \leq I}$  and  $\mathbf{s} = \{s_j\}_{1 \leq j \leq J}$  with  $r_i = P(X = x_i)$  and  $s_j = P(Y = y_j)$ . In the case of concern, assume  $x_i$ ,  $1 \leq i \leq I$ , to be consecutive integers; similarly for  $y_j$ ,  $1 \leq j \leq J$ . The CDFs for  $X$  and  $Y$  are  $F(x) = P(X \leq x)$  and  $G(y) = P(Y \leq y)$ . If  $F(x)$  and  $G(y)$  were continuous and strictly increasing, the equipercentile equating function for the conversion from test  $X$  to test  $Y$  would be defined as  $e_Y(x) = G^{-1}(F(x))$ , and the conversion from test  $Y$  to test  $X$  would be defined similarly as  $e_X(y) = F^{-1}(G(y))$ . In practice,  $F(x)$  and  $G(y)$  are made continuous before applying the equipercentile equating functions. Let  $F_{h_X}(x; \mathbf{r})$  and  $f_{h_X}(x; \mathbf{r})$  be the CDF and probability density function (PDF) of  $X(h_X)$ . Similarly, let  $Y(h_Y)$  denote the continuous approximation of  $Y$  with bandwidth  $h_Y$ , and let  $G_{h_Y}(y; \mathbf{s})$  and  $g_{h_Y}(y; \mathbf{s})$  be its CDF and PDF, respectively.

The LK and UK considered in this study are presented in Section 10.2, including details about the quantities needed in the equating process. Section 10.3 focuses on the step of continuization using LK and UK. Most of the results are applicable to generic kernel functions. In Section 10.4, LK, UK, and GK are applied to the equivalent-groups data given in Chapter 7 of von Davier et al. (2004b). Results are concluded in Section 10.5. The computation of SEE and SEED involves the first derivatives of  $F_{h_X}(x; \mathbf{r})$  and  $G_{h_Y}(y; \mathbf{s})$  with respect to  $\mathbf{r}$  and  $\mathbf{s}$ , respectively. In the Appendix, the formulas for the derivatives are generalizations to LK and UK.

## 10.2 Alternative Kernels

The name of logistic or uniform distribution can refer to a family of distributions with variation in the parameters for location and scale or for boundaries. Moments and cumulants are functions of these parameters, but standardized measures such as skewness and kurtosis are invariant in this respect. The main concern is how choices of  $V$  can affect the equating process. One relevant issue regards the impact

of employing two distributions that diverge only in the scales. To investigate this issue, two distributions were chosen from each family of distributions under consideration.

### 10.2.1 Logistic Kernel (LK)

Suppose  $V$  is a logistic random variable. Its PDF has the form

$$k(v) = \frac{\exp(-v/s)}{s(1 + \exp(-v/s))^2},$$

and its CDF is given by

$$K(v) = \frac{1}{1 + \exp(-v/s)},$$

where  $s$  is the scale parameter.  $V$  has mean 0 and variance  $\sigma_V^2 = \pi^2 s^2 / 3$ . Varying the scale parameter would expand or shrink the distribution. If  $s=1$ , the distribution is called the *standard logistic*, whose variance is  $\pi^2/3$ . The distribution can be rescaled to have mean 0 and identity variance by setting  $s = \sqrt{3}/\pi$ , which is called the *rescaled logistic* herein. In the rest of the chapter, *SLK* stands for the cases where standard logistic is used as the kernel function, and *RLK* stands for those with rescaled logistic kernel function.

The heavier tails and sharper peak of a logistic distribution lead to larger cumulants of even orders than do those of a normal distribution. When  $V$  follows a standard logistic distribution, for  $|t| < 1$ , the moment-generating function of  $V$  is given by

$$\begin{aligned} M_V(t) &= E(\exp(tV)) = \int_{-\infty}^{\infty} \exp(tv) \cdot \frac{\exp(-v)}{(1 + \exp(-v))^2} dv \\ &= \int_0^1 \xi^{-t} (1 - \xi)^t d\xi \\ &= B(1 - t, 1 + t) \\ &= \Gamma(1 - t)\Gamma(1 + t), \end{aligned}$$

where  $\xi = (1 + \exp(v))^{-1}$ ,  $B(\cdot, \cdot)$  is the beta function, and  $\Gamma(\cdot)$  is the gamma function (Balakrishnan, 1992). The cumulant-generating function of  $V$  is

$$\log M_V(t) = \log \Gamma(1 - t) + \log \Gamma(1 + t). \quad (10.2)$$

Let  $\Gamma^{(n)}(\cdot)$  be the  $n$ th derivative of  $\Gamma(\cdot)$  for any positive integer  $n$ . The cumulants of SLK may be derived from Equation 10.2 by differentiating with respect to  $t$  and setting  $t$  to 0. The next theorem gives the mathematical expressions of the cumulants for a general LK.

**Theorem 10.1.** *Define*

$$\psi(u) = \frac{d \log \Gamma(u)}{du} = \frac{\Gamma^{(1)}(u)}{\Gamma(u)},$$

and let  $\psi^{(n)}(\cdot)$  be the  $n$ th derivative of  $\psi(\cdot)$  for any positive integer  $n$ . Then the  $n$ th cumulant of a logistic random variable  $V$  with scale parameter  $s$  is found to be

$$\kappa_{n,V} = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 2s^n \psi^{(n-1)}(1) & \text{if } n \text{ is even} \end{cases}.$$

For any  $n \geq 1$  the value of  $\psi^{(n-1)}(1)$  is given by  $\psi^{(n-1)}(1) = (-1)^n (n-1)! \zeta(n)$  and  $\psi(1) = \Gamma^{(1)}(1) = -0.5772$ , where  $\zeta(\cdot)$  is the Riemann zeta function. These numbers were tabulated by Abramowitz and Stegun (1972), and the first six values of  $\zeta(n)$  are  $\zeta(1) = \infty$ ,  $\zeta(2) = \pi^2/6$ ,  $\zeta(3) \approx 1.2021$ ,  $\zeta(4) = \pi^4/90$ ,  $\zeta(5) \approx 1.0369$ , and  $\zeta(6) = \pi^6/945$ . Note that  $\zeta(n) > 0$  for even  $n \geq 2$ , so  $\kappa_{n,V} > 0$  for even  $n \geq 2$ .

### 10.2.2 Uniform Kernel (UK)

Suppose  $V$  is a uniform random variable with PDF

$$k(v) = \begin{cases} 1/(2b) & \text{for } -b \leq v \leq b \\ 0 & \text{otherwise} \end{cases},$$

where  $b$  is a positive real number, and CDF

$$K(v) = \begin{cases} 0 & \text{for } v < -b \\ (v+b)/2b & \text{for } -b \leq v < b \\ 1 & \text{for } v \geq b \end{cases}.$$

The  $V$  has mean 0 and variance  $b^2/3$ . The *standard uniform* distribution often refers to  $V$  with  $b = 1/2$ ; the variance is  $\sigma_V^2 = 1/12$ . When  $V$  is rescaled to have identity variance, the resulting distribution is called *rescaled uniform*. Standard uniform distribution and rescaled uniform distribution will be incorporated in the procedure of continuation; these methods will be denoted as *SUK* and *RUK*, respectively.

Following the previous notation,  $\kappa_{n,V}$  is the  $n$ th cumulant of  $V$ . Kupperman (1952) showed that all odd cumulants vanish and even cumulants are given by

$$\kappa_{n,V} = \frac{(2b)^n B_n}{n} \text{ for even number } n,$$

where  $B_n$  are Bernoulli numbers. The first 11 Bernoulli numbers are  $B_0 = 1$ ,  $B_1 = -1/2$ ,  $B_2 = 1/6$ ,  $B_4 = -1/30$ ,  $B_6 = 1/42$ ,  $B_8 = -1/30$ ,  $B_{10} = 5/66$ , and  $B_3 = B_5 = B_7 = B_9 = 0$ . Note that the even cumulants of UK have no definite sign.

### 10.3 Continuization With Alternative Kernels

Equation 10.1 illustrates the central idea of kernel smoothing, but  $X(h_X)$  can be defined in various ways for different purposes. In test equating, one desirable feature is to preserve moments of the discrete score distribution. Accordingly, in the kernel equating framework  $X(h_X)$  is defined to preserve the mean and variance of  $X$  by

$$X(h_X) = a_X(X + h_X V) + (1 - a_X)\mu_X, \quad (10.3)$$

where

$$a_X^2 = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_V^2 h_X^2}.$$

The continuous approximation for  $Y$  is analogously defined as

$$Y(h_Y) = a_Y(Y + h_Y V) + (1 - a_Y)\mu_Y,$$

where

$$a_Y^2 = \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_V^2 h_Y^2}.$$

Continuization of  $X$  and  $Y$  is based on the same formulation with a specific  $V$ . We will take  $X$  as an example and describe the properties relevant to  $X(h_X)$  for different kernels.

Theorems 10.2–10.5 below are generalizations of Theorems 4.1–4.3 in von Davier et al. (2004b) to LK and UK. Theorem 10.2 illustrates a few limiting properties of  $X(h_X)$  and  $a_X^2$  as  $h_X$  approaches 0 or infinity. Theorems 10.3 and 10.4 define the CDFs and PDFs of  $X(h_X)$  in which  $V$  is logistically and uniformly distributed, respectively, and Theorem 10.5 shows their forms as  $h_X$  approaches 0 or infinity.

**Theorem 10.2.** *The following statements hold:*

- (a)  $\lim_{h_X \rightarrow 0} a_X = 1$ ;
- (b)  $\lim_{h_X \rightarrow \infty} a_X = 0$ ;
- (c)  $\lim_{h_X \rightarrow \infty} h_X a_X = \sigma_X / \sigma_V$ ;
- (d)  $\lim_{h_X \rightarrow 0} X(h_X) = X$ ; and
- (e)  $\lim_{h_X \rightarrow \infty} X(h_X) = (\sigma_X / \sigma_V)V + \mu_X$ .

**Theorem 10.3.** Assume that  $V$  is logistically distributed with CDF and PDF defined in Section 10.2.1. Then CDF of  $X(h_X)$  is given by

$$F_{h_X}(x; \mathbf{r}) = \sum_i r_i K(R_{iX}(x))$$

with  $R_{iX}(x) = (x - a_X x_i - (1 - a_X)\mu_X)/(a_X h_X)$ . The corresponding PDF is

$$f_{h_X}(x; \mathbf{r}) = \frac{1}{a_X h_X} \sum_i r_i k(R_{iX}(x)).$$

**Theorem 10.4.** If  $V$  follows a uniform distribution with CDF and PDF given in Section 10.2.2, then the CDF of  $X(h_X)$  is

$$F_{h_X}(x; \mathbf{r}) = \sum_{i: R_{iX}(x) \geq b} r_i + \sum_{i: -b \leq R_{iX}(x) \leq b} \left\{ r_i \cdot \frac{R_{iX}(x) + b}{2b} \right\}, \quad (10.5)$$

where  $R_{iX}(x)$  is defined in Theorem 10.3. In addition, the PDF is

$$f_{h_X}(x; \mathbf{r}) = \frac{1}{a_X h_X} \sum_{i: -b \leq R_{iX}(x) \leq b} \frac{r_i}{2b}. \quad (10.6)$$

Note that linear interpolation as it is achieved in existing equating practice does not involve rescaling, which leads to a continuous distribution that does not preserve the variance of the discrete score distribution.

**Theorem 10.5.** The  $R_{iX}(x)$  defined in Theorem 10.3 has the following approximate forms when  $h_X$  approaches 0 and infinity:

- (a)  $R_{iX}(x) = \frac{x - x_i}{h_X} + o(h_X)$  as  $h_X \rightarrow 0$ , and  
 (b)  $R_{iX}(x) = \frac{x - \mu_X}{\sigma_X/\sigma_V} - \left( \frac{\sigma_X}{\sigma_V h_X} \right) \cdot \left( \frac{x - \mu_X}{\sigma_X/\sigma_V} \right) + o\left( \frac{\sigma_X}{\sigma_V h_X} \right)$  as  $h_X \rightarrow \infty$ .

### 10.3.1 Selection of Bandwidth

In the kernel equating framework, the optimal bandwidth minimizes a penalty function comprising two components. One is the least square term

$$\text{PEN}_1(h_X) = \sum_i (\hat{r}_i - f_{h_X}(x_i; \hat{\mathbf{r}}))^2,$$

where  $\hat{\mathbf{r}} = \{\hat{r}_i\}_{1 \leq i \leq I}$  are the fitted score probabilities of  $\mathbf{r}$  in the presmoothing step, and  $f_{h_X}(x_i; \hat{\mathbf{r}})$  is an estimate of  $f_{h_X}(x_i; \mathbf{r})$ . The other is a smoothness penalty term that avoids rapid fluctuations in the approximated density,

$$\text{PEN}_2(h_X) = \sum_i A_i(1 - B_i),$$

where

$$A_i = \begin{cases} 1 & \text{if } f_{h_X}^{(1)}(x; \hat{\mathbf{r}}) < 0 \text{ at } x = x_i - 0.25, \\ 0 & \text{otherwise} \end{cases},$$

$$B_i = \begin{cases} 0 & \text{if } f_{h_X}^{(1)}(x; \hat{\mathbf{r}}) > 0 \text{ at } x = x_i + 0.25, \\ 1 & \text{otherwise} \end{cases},$$

and  $f_{h_X}^{(1)}(x; \hat{\mathbf{r}})$  is the first derivative of  $f_{h_X}(x; \hat{\mathbf{r}})$ . Choices of  $h_X$  that allow a U-shaped  $f_{h_X}(x; \hat{\mathbf{r}})$  around the score point  $x_i$  would result in a penalty of 1. Combining  $\text{PEN}_1(h_X)$  and  $\text{PEN}_2(h_X)$  gives the complete penalty function

$$\text{PEN}(h_X) = \text{PEN}_1(h_X) + \text{PEN}_2(h_X), \quad (10.7)$$

which will keep the histogram with fitted score probabilities  $\hat{\mathbf{r}}$  and the continuized density  $f_{h_X}(x; \hat{\mathbf{r}})$  close to each other at each score point, while preventing  $f_{h_X}(x; \hat{\mathbf{r}})$  from having too many 0 derivatives.

For LK, we have

$$f_{h_X}^{(1)}(x; \mathbf{r}) = \frac{1}{s(a_X h_X)^2} \sum_i r_i k(R_{iX}(x)) [1 - 2K(R_{iX}(x))].$$

For UK,  $f_{h_X}(x; \mathbf{r})$  is piecewise constant and is differentiable at  $x=x_i$ ,  $1 \leq i \leq I$ . From Equation 10.6,  $f_{h_X}^{(1)}(x; \mathbf{r}) = 0$  for all  $x$  satisfying  $R_{iX}(x) \neq \pm b$ ,  $1 \leq i \leq I$ . Thus  $\text{PEN}_2(h_X) = 0$  with probability 1. The optimal bandwidth for UK should yield  $2bh_X$  close to 1, the distance between two consecutive possible scores.

### 10.3.2 Evaluation

It is common to compare distributions through moments or cumulants. Here we chose to examine cumulants, for each cumulant of a sum of independent random variables is the sum of the corresponding cumulants of the addends. A concise equation can be achieved to describe the relationship between cumulants of the discrete score distribution, kernel function, and the resulting continuous approximation. It allows not only numerical but theoretical comparisons between cumulants of  $X(h_X)$  for various kernels.



**Theorem 10.6.** Let  $\kappa_n(h_X)$  denote the  $n$ th cumulant of  $X(h_X)$ ,  $\kappa_{n,X}$  denote the  $n$ th cumulant of  $X$ , and  $\kappa_{n,V}$  denote the  $n$ th cumulant of  $V$ . Then for  $n \geq 3$ ,

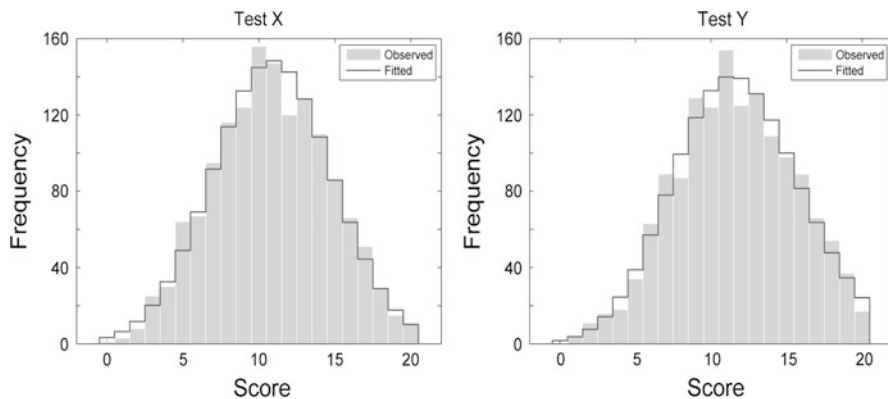
$$\kappa_n(h_X) = (a_X)^n (\kappa_{n,X} + (h_X)^n \kappa_{n,V}). \tag{10.8}$$

Because  $a_X \in (0, 1)$  and  $h_X > 0$ , a kernel function with  $\kappa_{n,V}$  having the same sign as  $\kappa_{n,X}$  leads to a closer approximation in terms of cumulants. Notice that Theorem 4.4 in von Davier et al. (2004b) is a special case of Theorem 10.6 because, for GK,  $\kappa_{n,V} = 0$  for any  $n \geq 3$ .

### 10.4 Application

The data used for illustration are results from two 20-item mathematics tests given in Chapter 7 of von Davier et al. (2004b). The tests, both number-right scored tests, were administered independently to two samples from a national population of examinees, which yields an equivalent-groups design. The two sample sizes are 1,453 and 1,455, respectively.

The raw data in an equivalent-groups design are often summarized as two sets of univariate frequencies. Figure 10.1 shows the histograms of the observed frequencies. Two univariate log-linear models were fitted independently to each set of frequencies. The moments preserved in the final models were the first two and three for  $X$  and  $Y$ , respectively. That is, the mean and variance of  $X$  and the mean, variance, and skewness of  $Y$  were preserved. The model fit was examined through the likelihood ratio test and the Freeman-Tukey residuals; the results showed no evidence of lack of fit. The fitted frequencies for test  $X$  and test  $Y$  are sketched in Figure 10.1 as well. The  $\hat{r}$ , the fitted score probabilities of  $X$ , are the ratios between the fitted frequencies and the sample size. The  $\hat{s}$  are attainable by the same means.



**Fig. 10.1** Observed frequencies and fitted frequencies for test  $X$  and test  $Y$

**Table 10.1** *Optimal Bandwidths for  $X(h_X)$  and  $Y(h_Y)$*

	Logistic kernel		Uniform kernel		Gaussian kernel
	Standard	Rescaled	Standard	Rescaled	
$h_X$	0.5117	0.9280	1.0029	0.2895	0.6223
$h_Y$	0.4462	0.8094	1.0027	0.2895	0.5706
$a_X$	0.9715	0.9715	0.9971	0.9971	0.9869
$a_Y$	0.9795	0.9795	0.9973	0.9973	0.9896

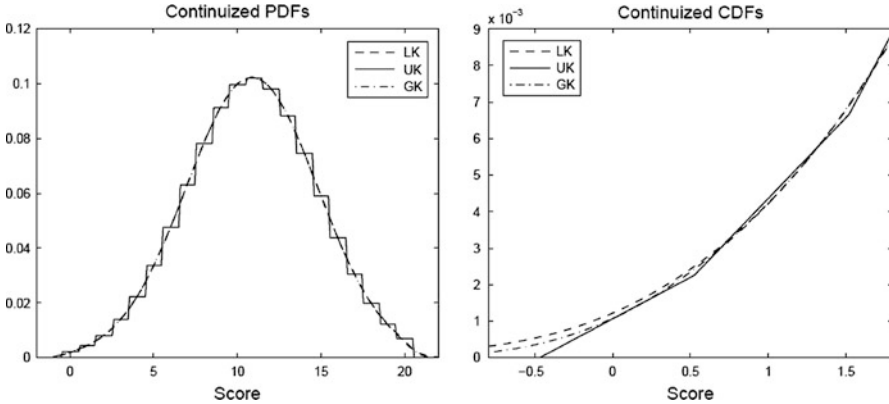
The optimal bandwidths using SLK, RLK, SUK, RUK, and GK are listed in Table 10.1. The first finding comes from the comparison between the standard and rescaled versions of LK or UK. Suppose the former has mean 0 and standard deviation  $\sigma_1$ , while the latter has mean 0 and standard deviation  $\sigma_2$ . Then the corresponding optimal bandwidths,  $h_1$  and  $h_2$ , for  $X$  or  $Y$ , satisfy the following equality:

$$\sigma_1 h_1 = \sigma_2 h_2. \quad (10.9)$$

In other words, the scale difference in different versions is adjusted by the selected bandwidths. Different versions of kernel function produce identical continuized score distributions as long as they come from one family of distributions.

The second finding is obtained through the comparison among GK, RLK, and RUK. Their first three moments (i.e., mean, variance, and skewness) are the same, but their major differences in shape can be characterized by the fourth moment or, equivalently, the kurtosis. Among the three, RLK has the largest kurtosis and RUK has the smallest kurtosis (the larger the kurtosis, the heavier the tails). Table 10.1 indicates that the heavier the tails of a kernel function, the larger the optimal bandwidth. Note that kurtosis is a standardized measure, so different versions from the same family of distributions have the same kurtosis. This observation can be generalized to SLK and SUK through Equation 10.9: the heavier the tails of a kernel function, the larger the product of its standard deviation and the optimal bandwidth.

Figure 10.2 displays the  $f_{h_X}(x; \hat{\mathbf{r}})$  and the left tail of  $F_{h_X}(x; \hat{\mathbf{r}})$  for LK, UK, and GK with optimal bandwidths. The graph in the left panel reveals that the  $f_{h_X}(x; \hat{\mathbf{r}})$  for LK and GK are smooth functions and hard to be distinguished. The  $f_{h_X}(x; \hat{\mathbf{r}})$  for UK is piecewise constant and appears to outline the histogram of the fitted frequencies for test  $X$  in Figure 10.1. The right panel only presents the portion of  $F_{h_X}(x; \hat{\mathbf{r}})$  within the range of -1 to 2, for the difference between curves may not be seen easily when graphed against the whole score range. Apparently, the  $F_{h_X}(x; \hat{\mathbf{r}})$  for LK has heavier left tail than that for GK, which corresponds to the fact that LK has heavier tails than GK. The use of UK results in a piecewise linear  $F_{h_X}(x; \hat{\mathbf{r}})$ , which is how linear interpolation functions in the percentile-rank method. Yet, we improved upon the linear interpolation by rescaling it in Equation 10.3 so that its continuous approximation preserves not only the mean but also the variance of the discrete score



**Fig. 10.2** Continuized score distributions of  $X$ : probability density functions (PDFs) and left tail of cumulative distribution functions (CDFs)

distribution. It is clear that the distributional characteristics of kernel functions are inherited by the corresponding continuous approximations. Figures for  $g_{h_Y}(y; \hat{s})$  and  $G_{h_Y}(y; \hat{s})$  exhibit the same properties, so they are omitted.

According to Equation 10.8, two phenomena are anticipated in assessing the continuized score distributions. First, the smaller the  $h_X$ , the closer the  $\kappa_n(h_X)$  to  $\kappa_{n,X}$ . Second, for a fixed  $h_X$ , if  $\kappa_{n,X}$  and  $\kappa_{n,V}$  have the same sign, the corresponding  $V$  will yield a closer continuous approximation in terms of cumulants. Attentions should be paid especially to even cumulants, as the three kernels have 0 odd cumulants. Recall that all even cumulants for LK are positive; the fourth and sixth cumulants for UK are negative and positive, respectively; and the even cumulants for GK of orders higher than three are 0. In this data example, cumulants were calculated numerically and found to coincide with theoretical findings from Equation 10.8. In Table 10.2, the second column shows the cumulants of the fitted, discrete score distributions; the fourth cumulant is negative and the sixth cumulant is positive. Because the comparison of cumulants involves two varying factors, type of kernel and bandwidth, it can be simplified by first examining cumulants for LK, UK, and GK with fixed bandwidths. The three chosen levels of bandwidth were small  $h_X$  ( $h_X = 0.2895$ ), moderate  $h_X$  ( $h_X = 0.6223$ ), and large  $h_X$  ( $h_X = 0.9280$ ). Each  $h_X$  is optimal for a certain kernel function, and the cumulants of optimal cases are highlighted in boldface. If we focus on one type of kernel and vary the level of bandwidth, it is evident that the cumulants under small  $h_X$  are closest to the corresponding cumulants of the fitted discrete score distributions.

For a fixed  $h_X$ , the performance of a kernel function from the viewpoint of how close its  $f_{h_X}(x, \hat{r})$  can approximate the histogram of  $\hat{r}$  has the following orders (from the best to the worst): UK, GK, LK for the fourth cumulant and LK, GK, UK for the sixth cumulant. The discrepancy in the orderings is due to the sign change in  $\kappa_{n,X}$  for  $n = 4$  and 6. In sum, UK best preserves  $\kappa_{n,X}$  for  $n \geq 3$  with its optimal

**Table 10.2** *Cumulants of  $X(h_X)$  With Fixed Bandwidths*

Order	Discrete	LK	UK	GK
$h_X = 0.2895$				
1	10.82	10.82	<b>10.82</b>	10.82
2	14.48	14.48	<b>14.48</b>	14.48
3	-3.57	-3.54	<b>-3.56</b>	-3.54
4	-63.16	-62.42	<b>-63.10</b>	-62.43
5	23.17	22.83	<b>22.81</b>	22.83
6	510.69	501.88	<b>501.86</b>	501.85
$h_X = 0.6223$				
1	10.82	10.82	10.82	<b>10.82</b>
2	14.48	14.48	14.48	<b>14.48</b>
3	-3.57	-3.44	-3.44	<b>-3.44</b>
4	-63.16	-59.74	-60.09	<b>-59.91</b>
5	23.17	21.70	21.75	<b>21.71</b>
6	510.69	472.19	471.72	<b>471.77</b>
$h_X = 0.9280$				
1	10.82	<b>10.82</b>	10.82	10.82
2	14.48	<b>14.48</b>	14.48	14.48
3	-3.57	<b>-3.28</b>	-3.27	-3.28
4	-63.16	<b>-55.49</b>	-57.13	-56.27
5	23.17	<b>20.06</b>	20.06	20.05
6	510.69	<b>432.12</b>	431.89	429.45

*Note:* Boldface indicates cumulants of a certain kernel function with its optimal bandwidth. LK = logistic kernel; GK = Guassian kernel; UK = uniform kernel

bandwidth. Cumulants of LK with its optimal bandwidth shrink by the most amount since they correspond to the largest bandwidth among the three kernels.

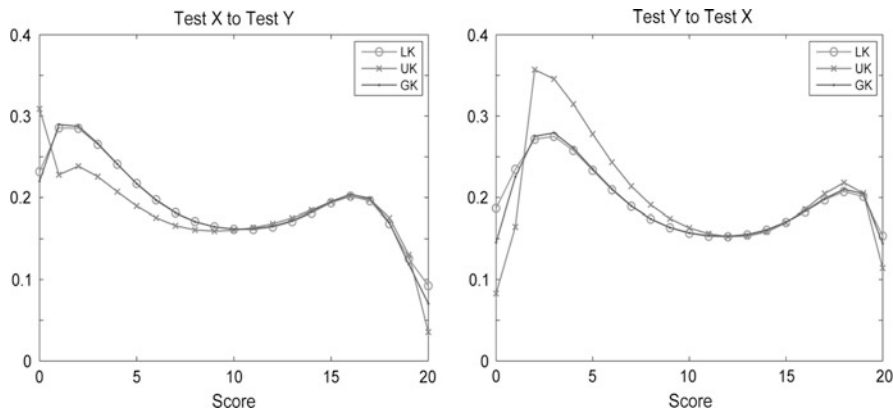
The conversion of scores from test X to test Y is based on the equation  $\hat{e}_Y(x) = G_{h_Y}^{-1}(F_{h_X}(x; \hat{r}); \hat{s})$  with optimal bandwidths. Similarly, the conversion of scores from test Y to test X is  $\hat{e}_X(y) = F_{h_X}^{-1}(G_{h_Y}(y; \hat{s}); \hat{r})$ . They are sample estimates of  $e_Y(x)$  and  $e_X(x)$  based on  $\hat{r}$  and  $\hat{s}$ . In Table 10.3, the equated scores for LK and GK are comparable, except for extreme scores, because their  $F_{h_X}(x; \hat{r})$  and  $G_{h_Y}(y; \hat{s})$  mainly differ at tails. UK tends to provide the most extreme equated scores among the three kernels. In addition, the average difference of equated scores between UK and GK is about twice as large as the average difference between LK and GK for the conversion from test X to test Y. For the inverse conversion, the average difference between UK and GK exceeds three times of the average difference between LK and GK. Overall, the maximal difference between any two kernels is about 0.18 raw-score point.

The sampling variability in  $\hat{e}_Y(x)$  or  $\hat{e}_X(y)$  is measured by the standard deviation of the asymptotic distribution, or the SEEs. It is known that distributions with heavier tails yield more robust modeling of data with more extreme values, and the same phenomenon is revealed when LK is employed. Figure 10.3 demonstrates that the SEEs for LK and GK do not differ remarkably. There is slightly less variation in the SEEs for LK. However, the SEEs for GK tend to have sharper drops at extreme scores, which are  $X = 0$  and  $20$  and  $Y = 0, 1,$  and  $20$  in this example. If the two forms to be equated have more discrepancy in the shape of their score distributions,

**Table 10.3** Equated Scores With Optimal Bandwidths

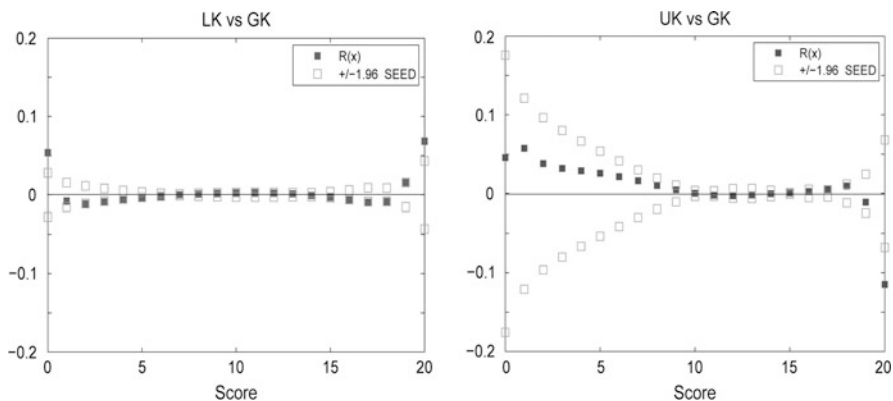
Score	X to Y			Y to X		
	LK	UK	GK	LK	UK	GK
0	0.447	0.439	0.394	-0.413	-0.227	-0.322
1	1.573	1.639	1.581	0.486	0.557	0.497
2	2.629	2.678	2.640	1.396	1.429	1.386
3	3.635	3.676	3.644	2.365	2.389	2.356
4	4.625	4.660	4.632	3.367	3.392	3.360
5	5.614	5.643	5.618	4.379	4.405	4.375
6	6.608	6.631	6.610	5.389	5.415	5.387
7	7.612	7.628	7.612	6.392	6.415	6.391
8	8.627	8.636	8.626	7.384	7.403	7.385
9	9.655	9.658	9.653	8.365	8.379	8.366
10	10.696	10.694	10.694	9.333	9.342	9.335
11	11.750	11.745	11.747	10.290	10.295	10.293
12	12.815	12.810	12.813	11.236	11.239	11.239
13	13.888	13.885	13.887	12.174	12.175	12.175
14	14.963	14.964	14.964	13.105	13.105	13.105
15	16.031	16.035	16.034	14.034	14.033	14.033
16	17.072	17.081	17.078	14.971	14.967	14.968
17	18.058	18.073	18.068	15.930	15.920	15.924
18	18.952	18.970	18.961	16.939	16.922	16.929
19	19.734	19.707	19.718	18.058	18.036	18.048
20	20.461	20.278	20.393	19.369	19.399	19.415

Note: LK = logistic kernel; GK = Gaussian kernel; UK = uniform kernel



**Fig. 10.3** Standard errors of equating (SEEs) for logistic (LK), uniform (UK), and Gaussian (GK) kernels

the equating functions for GK are likely to show sharp humps in the SEEs, but the SEEs for LK will remain less variable. On the other hand, the SEEs for UK do not display the same pattern and have greater variations from one conversion to another.



**Fig. 10.4** Difference and standard errors of equating difference (SEEDs) between  $\hat{e}_Y(x)$  of two kernel functions: logistic (LK) versus Gaussian (GK) and uniform (UK) versus Gaussian (GK)

It is straightforward to compare two estimated equating functions, for example,  $\hat{e}_{Y,LK}(x)$  and  $\hat{e}_{Y,GK}(x)$  for LK and GK by  $R(x) = \hat{e}_{Y,LK}(x) - \hat{e}_{Y,GK}(x)$  with an uncertainty measure, the SEED, to identify the 95% confidence interval for  $R(x)$ . Analogously,  $R(x) = \hat{e}_{Y,UK}(x) - \hat{e}_{Y,GK}(x)$  compares the estimated equating functions for UK and GK. The  $R(x)$  for comparisons between LK and GK and between UK and GK are plotted in Figure 10.4, all converting  $X$  to  $Y$ . Two curves representing  $\pm 1.96$  times of the SEEDs are also provided as the upper and lower bounds of the 95% confidence interval. For the comparison between LK and GK,  $R(0)$  and  $R(20)$  are significantly different at the 0.05 level, but the scale of SEED is less than 0.1 raw-score point, so the difference may still be negligible in practice. Again, the absolute values of  $R(x)$  and SEEDs increase as  $x$  approaches its boundaries, 0 and 20. The right panel shows that the difference between  $\hat{e}_{Y,UK}(x)$  and  $\hat{e}_{Y,GK}(x)$  is much larger than that of  $\hat{e}_{Y,LK}(x)$  and  $\hat{e}_{Y,GK}(x)$  for all score points outside the range of 9–17. The difference is nonsignificant at the 0.05 level, however, since the corresponding SEEDs are greater in scale.

## 10.5 Conclusions

Kernel equating is a unified approach for test equating and uses only GK smoothing to continuize the discrete score distributions. This chapter demonstrates the feasibility to incorporate alternative kernels in the five-step process and elaborates on tools for comparisons between various kernels. Equating through LK, UK, or GK has discrepancies in the continuized score distributions (e.g., heavier tails, piecewise continuity, or thinner tails that are inherited from the kernel functions) and hence in any product since the step of continuization. Although these discrepancies do not yield pronounced changes in the equated scores, certain desirable properties in the equating process could be achieved by manipulating the kernels.

## Chapter 10 Appendix

### 10.A.1 Computation of SEE With Alternative Kernels

The equating functions,  $e_Y(x)$  and  $e_X(y)$ , are functions of  $\mathbf{r}$  and  $\mathbf{s}$  through a design function and the composition of  $F_{h_X}(x; r)$  and  $G_{h_Y}(y; s)$ . The asymptotic variances of  $\hat{e}_Y(x)$  and  $\hat{e}_X(y)$ , which are equivalent to the squares of the SEEs, can be derived with the application of delta method given the asymptotic variances of  $\hat{\mathbf{r}}$  and  $\hat{\mathbf{s}}$ . Under the assumption that the bandwidths are fixed constants, for each conversion, the resulting asymptotic variance involves the matrix multiplication of three ingredients: the C-matrix from presmoothing, the Jacobian of the design function, and the Jacobian of the equating function. Changing kernel functions only affects the expression of the third ingredient, whose generalization to LK and UK is depicted in the remainder of this section. Details of the other two ingredients should refer to Chapter 5 of von Davier et al. (2004b).

Let  $J_{e_Y}$  denote the Jacobian of  $e_Y(x)$  with respect to  $\mathbf{r}$  and  $\mathbf{s}$ .  $J_{e_Y}$  is a  $1 \times (I + J)$  vector in which

$$J_{e_Y} = \left( \frac{\partial e_Y}{\partial \mathbf{r}}, \frac{\partial e_Y}{\partial \mathbf{s}} \right) = \left( \frac{\partial e_Y}{\partial r_1}, \dots, \frac{\partial e_Y}{\partial r_I}, \frac{\partial e_Y}{\partial s_1}, \dots, \frac{\partial e_Y}{\partial s_J} \right).$$

When the score distributions  $F(x)$  and  $G(y)$  have been approximated by sufficiently smoothed CDFs, the derivatives of the equating function can be computed,

$$\frac{\partial e_Y}{\partial r_i} = \frac{1}{G^{(1)}} \cdot \frac{\partial F_{h_X}(x; \mathbf{r})}{\partial r_i}, \quad (10.A.1)$$

$$\frac{\partial e_Y}{\partial s_j} = -\frac{1}{G^{(1)}} \cdot \frac{\partial G_{h_Y}(e_Y(x); \mathbf{s})}{\partial s_j}, \quad (10.A.2)$$

where  $G^{(1)} = g_{h_Y}(e_Y(x); \mathbf{s})$ . With some calculus, the partial derivative in Equation 10.A.1 is found to be

$$\frac{\partial F_{h_X}(x; \mathbf{r})}{\partial r_i} = K(R_{iX}(x)) - M_{iX}(x; \mathbf{r}) f_{h_X}(x; \mathbf{r}), \quad (10.A.3)$$

where

$$M_{iX}(x, r) = \frac{1}{2} (x - \mu_X) (1 - a_X^2) \left( \frac{x_i - \mu_X}{\sigma_X} \right)^2 + (1 - a_X) x_i,$$

and  $K(R_{iX}(x))$  is the CDF of LK or UK evaluated at  $R_{iX}(x)$ . In general,  $K(\cdot)$  can be the CDF of any kernel function. Replacement of  $X, x, F$ , and  $\mathbf{r}$  in Equation 10.A.3 by  $\mathbf{Y}, y, G$ , and  $\mathbf{s}$ , respectively, leads to the partial derivative  $\partial G_{h_Y}(y; \mathbf{s}) / \partial s_j$  in Equation 10.A.2. An estimate of  $J_{e_Y}$  can be achieved given  $\mathbf{r} = \hat{\mathbf{r}}$  and  $\mathbf{s} = \hat{\mathbf{s}}$ . Formulas for the Jacobian of  $e_X(y)$  and its estimate are similar in form to those for  $J_{e_Y}$ .

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