ON THE MISUSE OF MANIFEST VARIABLES IN THE DETECTION OF MEASUREMENT BIAS

WILLIAM MEREDITH
UNIVERSITY OF CALIFORNIA, BERKELEY

ROGER E. MILLSAP
BARUCH COLLEGE, THE CITY UNIVERSITY OF NEW YORK

Measurement invariance (lack of bias) of a manifest variable $Y$ with respect to a latent variable $W$ is defined as invariance of the conditional distribution of $Y$ given $W$ over selected subpopulations. Invariance is commonly assessed by studying subpopulation differences in the conditional distribution of $Y$ given a manifest variable $Z$, chosen to substitute for $W$. A unified treatment of conditions that may allow the detection of measurement bias using statistical procedures involving only observed or manifest variables is presented. Theorems are provided that give conditions for measurement invariance, and for invariance of the conditional distribution of $Y$ given $Z$. Additional theorems and examples explore the Bayes sufficiency of $Z$, stochastic ordering in $W$, local independence of $Y$ and $Z$, exponential families, and the reliability of $Z$. It is shown that when Bayes sufficiency of $Z$ fails, the two forms of invariance will often not be equivalent in practice. Bayes sufficiency holds under Rasch model assumptions, and in long tests under certain conditions. It is concluded that bias detection procedures that rely strictly on observed variables are not in general diagnostic of measurement bias, or the lack of bias.

Key words: measurement bias, differential item functioning, Bayes sufficiency, local independence, exponential family.

Introduction

In this paper, we present a unified treatment of conditions that are supposed to allow the detection of measurement bias using statistical procedures involving only observed or manifest variables. In the typical situation, we have two or more populations of individuals defined by manifest variables such as age or gender. We seek to examine the measurement bias in some manifest variable $Y$ with respect to these populations. The variable $Y$ is believed to be systematically related to (or a measure of) a latent, unobserved variable $W$ that cannot be directly measured. For example, $Y$ might be a test item score and $W$ might be an ability believed relevant for performance on $Y$. As different examples, $Y$ might be salary and $W$ might be merit; or $Y$ might be job performance and $W$ might be "aptitude".

A common approach to this problem is to measure a second manifest variable $Z$ that can in some sense substitute for the unmeasured variable $W$. In the above examples, $Z$ might be a test score, an observed indicator of merit, or of aptitude, respectively. Bias is then studied by examining features of the conditional distribution of $Y$ given $Z$ (e.g., the regression function). Bias is said to exist when the populations differ on these features. The traditional chi-square procedures for the detection of item bias or differential item functioning (DIF) are examples of this approach (Berk, 1982; Ironson, 1982; Marascuilo & Slaughter, 1981; Scheuneman, 1979; Shepard, Camilli, &
Averill, 1981). A more recent example is the Mantel-Haenzsel procedure for DIF detection (Holland & Thayer, 1988; Mantel & Haenzsel, 1959). Examination of group differences in the regressions of job performance on observed aptitude measures is yet another example (Cleary, 1968; Reilly, 1986). A conclusion of this paper is that methods for studying bias that rely on manifest variables alone are not generally diagnostic of measurement bias, or the lack of bias. Although some of the results to be given are quite general, the emphasis in this paper will be on differential item functioning as it arises in mental test theory.

Preliminary Definitions

Throughout this paper \( X \) and \( X = \{Y, Z\} \) will be employed to denote the manifest multivariate measurement random variable of interest. For example, the elements of \( X \) might be item responses and/or test scores. The possibly multivariate latent random variable underlying \( X \) will be denoted by \( W \). Although it is often assumed that the variables in \( X \) are conditionally independent given \( W \), this assumption will not generally be made for reasons that will become clear in the sequel. Let \( V \) denote a vector-valued random variable whose elements are indicators of demographic characteristics such as age, gender, or ethnicity. To avoid trivialities, the following assumptions will be made throughout:

1. \( X \) and \( W \) are not independent,
2. For every realization of \( X \), the conditional distribution of \( W \) is not degenerate, and
3. \( V \) and \( W \) are not independent.

Suppose a parent population consisting of all individuals for whom measurement with \( X \) is deemed appropriate. Subpopulations of interest are selected from the parent on the basis of \( V \). Conceptualize this selection as occurring via a selection function \( \pi(v) = \Pr(\text{inclusion in subpopulation } | V = v) \). For example, the subpopulation of 18 year old white females would be obtained by letting \( \pi(v) = 1 \) for all realizations of \( V \) such that the coordinates of \( V \) corresponding to age, ethnicity, and gender assume the values corresponding to 18 years, white, and female, respectively; otherwise, \( \pi(v) = 0 \). There is no necessary requirement that \( \pi(v) = 0 \) or 1, merely \( 0 \leq \pi(v) \leq 1 \), allowing for probabilistically selected subpopulations.

Define measurement invariance, or lack of bias, as existing when the conditional distribution of \( X \) given \( W = w \) is the same in all subpopulations that can be derived from the parent by selection on \( V \), for all values of \( w \) for which \( \Pr(w) > 0 \) in the subpopulations. This definition is quite general, and includes examples that have been discussed in the literature. For example, Lord’s (1980) definition of lack of item bias is a special case, and Mellenbergh’s (1989) definition of lack of item bias is essentially identical to the above.

Basic Theorems

In what follows, we assume that the random variables \( X \), \( W \), and \( V \) are all discrete and finite, and argue that any continuous distributions of interest can be arbitrarily well-approximated by suitably chosen discrete distributions. In many situations, \( X \) is naturally discrete (e.g., responses to polychotomous items) and most coordinates of \( V \) are naturally discrete (e.g., gender, ethnicity). Continuity assumptions can be regarded as convenient idealizations in these cases.

Suppose that \( X \) can assume values \( x_h \) in Euclidean \( r \)-space, \( h = 1, \ldots, m \), \( W \) can
assume values $w_i, i = 1, \ldots, n$ in Euclidean $s$-space, and $V$ can assume values $v_j$ in Euclidean $t$-space, $j = 1, \ldots, p$. Let $\text{pr}(x_h)$ denote $\text{prob}(X = x_h)$, and so on, define $E(\cdot)$ as the mathematical expectation operator, and let $\pi_j = \pi(v_j)$. For a given selection function $\pi(v)$,

$$
\sum_{j=1}^{p} \text{pr}(w_i, v_j) \pi_j
$$

$$
\text{pr}(w_i|\pi) = \frac{\sum_{j=1}^{p} \text{pr}(v_j) \pi_j}{\sum_{j=1}^{p} \text{pr}(v_j) \pi_j},
$$

(1)

$$
\sum_{j=1}^{p} \text{pr}(x_h, w_i, v_j) \pi_j
$$

$$
\text{pr}(x_h, w_i|\pi) = \frac{\sum_{j=1}^{p} \text{pr}(v_j) \pi_j}{\sum_{j=1}^{p} \text{pr}(v_j) \pi_j}.
$$

(2)

Under the given assumptions, it is possible that $\text{pr}(w_i|\pi) = 0$ for some $w_i$, but for every fixed $i = I$, there exist selection functions such that $\text{pr}(w_i|\pi) > 0$. Since $\pi_i$ is constant is an admissible selection function, measurement invariance implies that for every $\pi(v)$, $\text{pr}(x_h|w_i, \pi) = \text{pr}(x_h|w_i)$ for every $h$ and those values of $i$ such that $\text{pr}(w_i|\pi) > 0$. The selection functions $\pi_j = 1, j = J, \pi_j = 0, j \neq J$ are admissible. The following theorem is central to the developments in the sequel.

**Theorem 1. (Fundamental Lemma of Measurement Invariance).** The conditional distribution of $X$ given $W = w_i$ is invariant with respect to selection on $V$ for $i = 1, \ldots, n$ if and only if $X$ and $V$ are locally independent when conditioned on $W$.

**Proof.** (If part) It is shown that local independence implies that for any given selection function, $\text{pr}(x_h|w_i, \pi) = \text{pr}(x_h|w_i)$ for all $h$ and all $i$ such that $\text{pr}(w_i|\pi) > 0$. From (1), (2), and local independence, we have for all $h$,

$$
\sum_{j=1}^{p} \text{pr}(x_h|w_i) \text{pr}(v_j|w_i) \text{pr}(w_i) \pi_j
$$

$$
\text{pr}(x_h|w_i, \pi) = \frac{\sum_{j=1}^{p} \text{pr}(v_j|w_i) \text{pr}(w_i) \pi_j}{\sum_{j=1}^{p} \text{pr}(v_j|w_i) \text{pr}(w_i) \pi_j} = \text{pr}(x_h|w_i),
$$

(3)

whenever $\text{pr}(w_i|\pi) > 0$. This must be true for every $i$ since there exists a selection function such that $\text{pr}(w_i|\pi) > 0$. (Only if part) We show that $\text{pr}(x_h|w_i, \pi) = \text{pr}(x_h|w_i)$ for all $h, i$ such that $\text{pr}(w_i|\pi) > 0$ and all $\pi(v)$ implies

$$
\text{pr}(x_h, w_i, v_j) = \text{pr}(x_h|w_i) \text{pr}(v_j|w_i) \text{pr}(w_i).
$$

for every $h, i$. We have
\[
pr(x_h|w_i, \pi) = \frac{\sum_{j=1}^{p} pr(x_h|w_i, v_j) pr(w_i, v_j) \pi_j}{\sum_{j=1}^{p} pr(w_i, v_j) \pi_j}, \tag{4}
\]

for all \(h, i\) such that \(pr(w_i|\pi) > 0\). From (4) it follows that

\[
\sum_{j=1}^{p} [pr(x_h|w_i, v_j) - pr(x_h|w_i)] pr(w_i, v_j) \pi_j = 0, \tag{5}
\]

for all \(h, i\), since \(pr(w_i, v_j) \pi_j = 0\) if \(pr(w_i|\pi) = 0\). Regarding \(\pi(v)\) as a \(p\)-dimensional vector, there must exist linearly independent sets of \(p\) such vectors among the set of all possible \(\pi(v)\). Because every possible \(\pi(v)\) can be included in some linearly independent set, if (5) is true for all \(\pi(v)\), it follows that

\[
pr(x_h|w_i, v_j) pr(w_i, v_j) = pr(x_h|w_i) pr(w_i|v_j) pr(v_j), \tag{6}
\]

or that

\[
pr(x_h, w_i, v_j) = pr(x_h|w_i) pr(v_j|w_i) pr(w_i), \tag{7}
\]

for every \(h, i, j\).

It is fairly easy to prove Theorem 1 for continuous and mixed continuous/discrete distributions of \(X\) and \(W\), as long as \(V\) is regarded as discrete. The key is that the only vector orthogonal to \(p\) linearly independent vectors of dimension \(p\) is the null vector, leading to (7) for all values of \(X\), \(W\), and \(V\).

Now partition \(X\) into two component random variables \(X = \{Y, Z\}\), both possibly multivariate. Suppose \(Y\) can take values \(y_g\) in Euclidean \(q\)-space, \(g = 1, \ldots, k\), and for parsimony in notation, suppose \(Z\) can take values \(z_h\) in Euclidean \(r\)-space, \(h = 1, \ldots, m\). We want to study the conditional distribution of \(Y\) given \(Z = z_h\) in selected subpopulations. We have

\[
pr(y_g, z_h|\pi) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{p} pr(y_g, z_h, w_i, v_j) \pi_j}{\sum_{j=1}^{p} pr(v_j) \pi_j} = \frac{\sum_{j=1}^{p} pr(y_g, z_h, v_j) \pi_j}{\sum_{j=1}^{p} pr(v_j) \pi_j}, \tag{8}
\]

and
WILLIAM MEREDITH AND ROGER E. MILLSAP 293

\[
\frac{\sum_{j=1}^{p} \Pr(z_h, v_j) \pi_j}{\sum_{j=1}^{p} \Pr(v_j) \pi_j} = \frac{\Pr(z_h \mid \pi)}{\Pr(v_j) \pi_j}.
\]  

(9)

We define invariance of the conditional distribution of \( Y \) given \( Z \) to mean that \( \Pr(y \mid z_h, \pi) \) is unaffected by selection on \( V \), or that \( \Pr(y \mid z_h) = \Pr(y \mid z_h) \) for all \( g, h \) such that \( \Pr(z_h \mid \pi) > 0 \) and for all \( \pi(v) \). Then Theorem 1 leads to the following corollary.

**Corollary 1.1.** The conditional distribution of \( Y \) given \( Z \) is invariant with respect to selection on \( V \) if and only if \( Y \) and \( V \) are locally independent when conditioned on \( Z \).

**Proof.** Same as that of Theorem 1 with \( X \) replaced by \( Y \) and \( W \) replaced by \( Z \), and employing (8) and (9). \( \square \)

Measurement invariance has not been invoked in the proof of Corollary 1.1. Invariance of the conditional distribution of \( Y \) given \( Z \) can hold even though the measurement properties of \( X = \{ Y, Z \} \) vary as a function of \( V \), or when bias is present. A simple example occurs when \( Z \) is an unweighted total test score including \( Y \), all items are dichotomous Rasch items, and a constant that depends only on \( V \) is added to each item difficulty parameter. This situation represents a "constant bias" that affects all items equally, and is known to create problems for most methods of DIF detection (Ironson, 1982; Thissen, Steinberg, & Wainer, 1988). In general, the conditions required for measurement invariance in Theorem 1 are distinct from those required for invariance of the conditional distribution of \( Y \) given \( Z \) in Corollary 1.1. In the next section, we present some conditions leading to equivalence between these two forms of invariance.

**The Measurement Invariant Case**

**Bayes Sufficiency**

Assume that the conditions for measurement invariance hold: \( \{ Y, Z \} \) and \( V \) are locally independent when conditioned on \( W \). When will the conditional distribution of \( Y \) given \( Z \) be invariant with respect to selection on \( V \)? We can rewrite (8) and (9) as

\[
\sum_{i=1}^{n} \sum_{j=1}^{p} \Pr(y_g, z_h \mid w_i) \Pr(w_i, v_j) \pi_j
\]

\[
\Pr(y_g, z_h \mid \pi) = \frac{\sum_{j=1}^{p} \Pr(v_j) \pi_j}{\sum_{j=1}^{p} \Pr(v_j) \pi_j}.
\]  

(10)

\[
\sum_{i=1}^{n} \sum_{j=1}^{p} \Pr(z_h \mid w_i) \Pr(w_i, v_j) \pi_j
\]

\[
\Pr(z_h \mid \pi) = \frac{\sum_{j=1}^{p} \Pr(v_j) \pi_j}{\sum_{j=1}^{p} \Pr(v_j) \pi_j}.
\]  

(11)
To simplify the presentation, we introduce some matrix notation. Let

i. \( p_{gh} \) denote the \( n \times 1 \) vector \( \{pr(y_g, z_h|w_i)\}_{i=1}^n \);

ii. \( p_h \) denote the \( n \times 1 \) vector \( \{pr(z_h|w_i)\}_{i=1}^n \);

iii. \( P_{wv} \) denote the \( n \times p \) matrix \( \{pr(w_i, v_j)\}_{i=1}^n, j=1, \ldots, p \);

iv. \( \pi \) denote the \( p \times 1 \) vector \( \{\pi_j\}_{j=1}^p \);

v. \( e_{gh} \) be an \( n \times 1 \) vector in the orthogonal complement of the column space of \( P_{wv} \) (i.e., \( e_{gh}' P_{wv} = 0' \));

vi. \( c_{gh} = pr(y_g|z_h) \);

vii. \( q_{gh} \) denote the \( p \times 1 \) vector \( \{pr(y_g, z_h, v_j)\}_{j=1}^p \).

We can rewrite (10) and (11) in this notation as
\[
pr(y_g, z_h|\pi) = \frac{p_{gh}' P_{wv} \pi}{1' P_{wv} \pi},
\]
(12)
\[
pr(z_h|\pi) = \frac{p_h' P_{wv} \pi}{1' P_{wv} \pi},
\]
(13)
where \( 1 \) is an \( n \times 1 \) vector of ones. The following theorem answers the question posed above.

**Theorem 2.** Given measurement invariance for \( \{Y, Z\} \), the conditional distribution of \( Y \) given \( Z \) is invariant with respect to selection on \( V \) if and only if \( p_{gh} = c_{gh} p_h + e_{gh} \), where \( e_{gh}' P_{wv} = 0' \), for all \( g, h \).

**Proof.** (If part) Given \( p_{gh} = c_{gh} p_h + e_{gh} \), one has, using (12) and (13) for every \( g \) and \( h \) such that \( pr(z_h|\pi) > 0 \),
\[
pr(y_g|z_h, \pi) = \frac{c_{gh} p_h' P_{wv} \pi}{p_h' P_{wv} \pi} = c_{gh} = pr(y_g|z_h).
\]
(14)
(Only if part) Observe that
\[
q_{gh}' = p_{gh}' P_{wv}.
\]
(15)
Invariance of the conditional distribution of \( Y \) given \( Z \) with respect to selection on \( V \) and Corollary 1.1 imply
\[
pr(y_g, z_h, v_j) = pr(y_g|z_h) pr(v_j|z_h) pr(z_h) = c_{gh} pr(z_h, v_j),
\]

hence,
\[
q_{gh}' = c_{gh} p_h' P_{wv},
\]
(16)
for all \( g, h \). Combining (15) and (16) yields
\[
(p_{gh} - c_{gh} p_h) P_{wv} = e_{gh}' P_{wv} = 0',
\]
(17)
for all \( g, h \). But (17) implies that the vector \( e_{gh} \) lies in the orthogonal complement of the column space of \( P_{wv} \); hence,
\[
p_{gh} = c_{gh} p_h + e_{gh}, \quad \text{with } e_{gh}' P_{wv} = 0' \quad \text{for all } g, h. \]
Theorem 2 describes the conditions required for invariance of the conditional distribution of \( Y \) given \( Z \) when measurement invariance holds. In the special case when \( e_{gh} = 0 \) for all \( g, h \), we have

\[
\Pr (y_g, z_h | w_i) = \Pr (y_g | z_h) \Pr (z_h | w_i),
\]

and consequently, \( Y \) and \( W \) are locally independent when conditioned on \( Z \). Lehmann (1986) characterizes this condition as Bayes sufficiency of \( Z \) for \( W \). One situation leading to \( e_{gh} = 0 \) for all \( g, h \) arises when the rank of \( P_{wv} \) is \( n \). Other conditions leading to Bayes sufficiency will be described later. Given measurement invariance and \( e_{gh} \neq 0 \) for some \( g, h \) these vectors must satisfy some conditions in addition to those in Theorem 2:

\[
\sum_{g=1}^{k} e_{gh} = 0 \text{ for every } h,
\]

\[
0 < \Pr (y_g, z_h | w_i) + e_{ghi} < 1 \text{ for every } g, h, i.
\]

Furthermore, since \( \Pr (z_h | w_i) = 0 \) implies that \( \Pr (y_g, z_h | w_i) = 0 \), every element of \( e_{gh} \) corresponding to a zero element in \( p_h \) must equal zero.

Theorem 2 can be reformulated by defining the following: Let \( t_{gh} \) denote the \( n \times 1 \) vector \( t_{gh} = \{\Pr (y_g | z_h, w_i)\}\) for \( i = 1, \ldots, D_h \) the \( n \times n \) diagonal matrix created by arraying the elements of \( p_h \) along the diagonal, and \( e_{gh} \) a \( n \times 1 \) vector orthogonal to \( D_h P_{wv} \). Theorem 2 implies the following corollary, stated without proof.

**Corollary 2.1.** Given measurement invariance, the conditional distribution of \( Y \) given \( Z \) is invariant with respect to selection on \( V \) if and only if \( t_{gh} = c_{gh} I + e_{gh} \), where \( e'_{gh} D_h P_{wv} = 0 \), for all \( g, h \).

If the \( ii \)-th diagonal of \( D_h \) is zero, the \( i \)-th element of \( e_{gh} \) is irrelevant. Also, if \( e_{gh} \) is not a null vector, it must satisfy additional constraints as does \( e_{gh} \) in Theorem 2.

The foregoing discussion of Bayes sufficiency leads to the following two additional corollaries to Theorem 2, stated without proof.

**Corollary 2.2.** Given measurement invariance for \( \{Y, Z\} \), if \( Z \) is Bayes sufficient for \( W \), then the conditional distribution of \( Y \) given \( Z \) is invariant with respect to selection on \( V \).

**Corollary 2.3.** Given measurement invariance for \( \{Y, Z\} \), if the rank of \( P_{wv} \) is \( n \), the conditional distribution of \( Y \) given \( Z \) is invariant with respect to selection on \( V \) if and only if \( Z \) is Bayes sufficient for \( W \).

Bayes sufficiency of \( Z \) leads to invariance of the conditional distribution of \( Y \) given \( Z \) in the measurement invariant case. But Bayes sufficiency is not a necessary condition for invariance of the conditional distribution of \( Y \) given \( Z \). Artificial examples can be constructed in which Bayes sufficiency fails, yet the conditional distribution of \( Y \) given \( Z \) is invariant over selection on \( V \). Hence, it is of interest to explore conditions other than Bayes sufficiency of \( Z \) that may determine relations between the two forms of invariance when measurement invariance holds. The following two sections present such conditions.

**Local Independence**

The variables \( Y \) and \( Z \) are said to be locally independent given \( W \) and \( V \) when the following is true
for all $g$, $h$, $i$, $j$. Under standard latent trait assumptions, local independence would hold, for example, if $Y$ is a test item score variable and $Z$ is a total score variable that excludes $Y$, with all items fitting latent trait models under a common trait, or set of traits. Local independence in (19) will contradict Bayes sufficiency in (18) under certain conditions. For example, if measurement invariance holds for $Y$ and (19) also holds, then Bayes sufficiency in (18) cannot hold given our earlier restrictions on marginal independence between $X = \{Y, Z\}$ and $W$. Yet, as shown by the following theorem, when measurement invariance holds for $Y$, it is still possible for the conditional distribution of $Y$ given $Z$ to be invariant under local independence.

Theorem 3. Suppose that (19) holds and that

1. $\Pr(Y_{gl}W_{i}, v_{j}) = \Pr(Y_{gl}W_{i})$,
2. $\Pr(W_{i}Z_{h}, v_{j}) = \Pr(W_{i}Z_{h})$,

for all $g$, $h$, $i$, $j$. Also, suppose that for some $h$, $\Pr(W_{i}Z_{h})$ is not a degenerate distribution. Then

$$\Pr(Y_{gl}Z_{h}, v_{j}) = \Pr(Y_{gl}Z_{h}),$$  \hspace{1cm} (20)

for all $g$, $h$, $j$.

Proof. Equations (19) and (i) imply

$$\Pr(Y_{gl}Z_{h}, W_{i}, v_{j}) = \Pr(Y_{gl}W_{i}).$$

Then, using (ii),

$$\Pr(Y_{gl}Z_{h}, v_{j}) = \sum_{i=1}^{n} \Pr(Y_{gl}W_{i}) \Pr(W_{i}Z_{h})$$

$$= \Pr(Y_{gl}Z_{h}),$$

establishing (20).

Note that (ii) in Theorem 3 implies that $Z$ is biased, or that

$$\Pr(Z_{h}W_{i}, v_{j}) \neq \Pr(Z_{h}W_{i}),$$ \hspace{1cm} (21)

for some $h$, $i$, $j$, given our restrictions on marginal independence between $V$ and $W$. The converse need not be true: (ii) need not hold when $Z$ is biased. If measurement invariance holds for both $Y$ and $Z$, the conditional distribution of $Y$ given $Z$ will not be invariant under local independence. This is established in the following theorem.

Theorem 4. Suppose that (19) holds and that

$$\Pr(Y_{gl}W_{i}, v_{j}) = \Pr(Y_{gl}W_{i}),$$ \hspace{1cm} (22)

$$\Pr(Z_{h}W_{i}, v_{j}) = \Pr(Z_{h}W_{i}).$$ \hspace{1cm} (23)

Also, suppose that for some $h$, $j$, $\Pr(W_{i}Z_{h}, v_{j})$ is not a degenerate distribution. Finally, suppose that for some $g$,
pr \((y_g | z_h) \neq pr \((y_g | z_h^*),\) \tag{24}

with \(h\) and \(h^*\) being distinct values of \(h\), for at least \(H > 2\) values of \(h\). Then

\[ pr \((y_g | z_h, v_j) \neq pr \((y_g | z_h), \]

for some values \(g, h, j\).

**Proof.** From (19), (22), and (23), it is easily shown that

\[ pr \((y_g, z_h | w_i, v_j) = pr \((y_g, z_h|w_i), \]

\[ pr \((y_g | z_h, w_i) = pr \((y_g|w_i), \]

for all \(g, h, i, j\). To begin, assume that the conclusion of the theorem is false, or that

\[ pr \((y_g | z_h, v_j) = pr \((y_g | z_h), \]

for all \(g, h, j\). Corollary 2.1 says that (25) and (27) imply

\[ t_{gh} = c_{gh} 1 + e_{ghi}. \tag{28} \]

Let \(e_{ghi}\) be the \(i\)-th element of \(e_{gh}\). Given (26), we can express (28) in scalar form as

\[ pr \((y_g | w_i) = pr \((y_g | z_h) + e_{ghi}. \tag{29} \]

Equation (29) must hold for all \(h\), given \(g\) and \(i\). Let \(t_g = \{pr \((y_g | w_i)\}\) be an \(n \times 1\) vector and let \(P_{wv}\) be the \(j\)th column of \(P_{wv}\). Then

\[ t'_g D_h P_{wv} = \sum_{i=1}^{n} pr \((y_g | w_i) pr \((z_h | w_i) pr \((w_i, v_j) \]

\[ = pr \((y_g, z_h, v_j), \tag{30} \]

using (22), (23), and (26). Pick any specific value of \(h\), say \(h^*\). Then (28) through (30) imply that if we postmultiply \(t_g\) by \(D_{h^*} P_{wv}\), the following must hold for all \(h\):

\[ pr \((y_g, z_{h^*}, v_j) = pr \((y_g | z_h) pr \((z_h^* | v_j) + \sum_{i=1}^{n} e_{ghi} pr \((z_h^*, w_i, v_j), \tag{31} \]

using (23). Note that when \(h = h^*\), the second term on the right in (31) is zero, given the definition of \(e_{gh}\). Dividing both sides of (31) by \(pr \((z_h^*, v_j)\) and using (27), we have

\[ pr \((y_g | z_{h^*}) = pr \((y_g | z_h) + \sum_{i=1}^{n} e_{ghi} pr \((w_i | z_{h^*}, v_j), \tag{32} \]

for all pairs \(h, h^*\), since the choice of \(h^*\) is entirely arbitrary. To simplify notation, let \(b_{hj} = \{pr \((w_i | z_h, v_j)\}\) be an \(n \times 1\) vector. If (32) holds for all pairs \(h, h^*\), it must be true that

\[ pr \((y_g | z_{h^*}) = pr \((y_g | z_h) + e'_{ghh} b_{h'j}, \tag{33} \]

\[ pr \((y_g | z_h) = pr \((y_g | z_{h^*}) + e'_{gh} b_{hj}, \tag{34} \]

for all pairs \(h, h^*\), with \(e'_{ghh} b_{hj} = 0\) (when \(h = h^*\)). If (33) and (34) hold for all pairs \(h, h^*\), then
\[ e_{gb} b_{h*} + e_{gb} b_{hj} = 0, \]  

(35)

for all such pairs. The terms on the left side of (35) cannot both be zero for all pairs \( h, h^* \) because this contradicts (24). Also, it is assumed that for some \( g \), there are at least \( H > 2 \) values of \( h \) giving distinct values of \( \text{pr} (y_g | z_h) \). But (35) cannot hold simultaneously for more than \( H = 2 \) values of \( h \), assuming nonzero values for the terms on the left side of (35). Thus one is led to a contradiction by assuming (27), and so (27) must be false.

The assumption in Theorem 4 concerning \( H \) is trivial in practice because in DIF applications we would expect \( H > 2 \) distinct values of \( \text{pr} (y_g | z_h) \) for some \( g \). Theorem 4 establishes that when measurement invariance holds for both \( Y \) and \( Z \), local independence precludes invariance of the conditional distribution of \( Y \) given \( Z \). This theorem has implications for the discussions of Holland and Thayer (1988) and Zwick (1990) concerning whether \( Z \) should include \( Y \) when \( Y \) is a dichotomous item score and \( Z \) is the sum of item scores on the test. Under local independence assumptions, if \( Y \) is omitted, the conditional distribution of \( Y \) given \( Z \) will not be invariant when \( Y \) and \( Z \) are unbiased.

**Stochastic Ordering**

In the measurement invariant case, failure of Bayes sufficiency can effectively preclude invariance of the conditional distribution of \( Y \) given \( Z \) when the distribution of \( W \) given \( V \) is stochastically ordered. Zwick (1990) demonstrates this fact in relation to the Mantel-Haenszel detection procedure when \( Y \) is a dichotomous test item score, \( Z \) is a total test score, and \( W \) is a univariate latent trait. This topic is of interest because in practice, it may be reasonable to suppose that the distributions of \( W \) are stochastically ordered among populations defined by \( V \). To define stochastic ordering, let \( F(U) \) be the cumulative probability for a discrete univariate random variable \( U \), so that

\[ F(u_q) = \sum_{s=1}^{q} \text{pr} (u_s), \]

where \( u_1 < u_2 < \cdots < u_r \) are realizations. In what follows, it will be said that given two distributions \( F_1(U) \) and \( F_2(U) \), \( F_2(U) \) is "stochastically greater" than \( F_1(U) \) if

\[ 1 - F_2(u_q) > 1 - F_1(u_q), \]

for all \( q \), with the inequality being strict for some values of \( U \). "Stochastically smaller" will imply the converse. In practice, \( F(W) \) may be stochastically ordered among populations defined by \( V \). For example, it is often assumed that the distributions of \( W \) are normal with means dependent on \( V \), and a common variance. Turn now to some general results regarding invariance of the conditional distribution of \( Y \) given \( Z \) when stochastic ordering assumptions are deemed realistic.

Begin with the following lemma, stated without proof.

**Lemma 1.** Suppose that \( U \) is a discrete univariate random variable taking \( r \) real values \( u_1 < u_2 < \cdots < u_r \). Let \( F_1(U) \) and \( F_2(U) \) be two cumulative distribution functions and suppose that \( F_2(U) \) is stochastically greater than \( F_1(U) \). Then, for any function \( \Psi(U) \) that is increasing in \( U \) and strictly increasing for at least one value \( u_q \) for which \( 1 - F_2(u_{q-1}) > 1 - F_1(u_{q-1}) \),
\[ E_2[\Psi(U)] > E_1[\Psi(U)]. \]

Similar lemmas appear in Lehmann (1955) and in Lord and Novick (1968). Before using this lemma, we define some new functions that will be useful. Equations (12) and (13) lead to

\[ \text{pr} \left( y_g \mid z_h, \pi \right) = \sum_{i=1}^{n} \text{pr} \left( y_g \mid z_h, w_i \right) \text{pr} \left( w_i \mid z_h, \pi \right), \] (36)

for all \( g, h \). Let \( \Phi(Y) \) be any (many-to-one) function of \( Y \). If \( Y \) is multivariate, \( \Phi(Y) \) could be a single coordinate of \( Y \). Then

\[ E[\Phi(Y) \mid z_h, \pi] = \sum_{i=1}^{n} \Psi_h(w_i) \text{pr} \left( w_i \mid z_h, \pi \right), \] (37)

where

\[ \Psi_h(w_i) = E[\Phi(Y) \mid z_h, w_i]. \] (38)

Although \( W \) may be multivariate, we will restrict attention to functions \( \Psi_h(W) \) in which \( W \) enters the function in a univariate way. In other words, we will assume that there exist many-to-one functions \( U_h(W) \) with values \( u_{hq}, q = 1, \ldots, r \), such that

\[ \Phi_h(w_i) = \Gamma_h(u_{hq}), \] (39)

for all \( w_i \) such that \( U_h(W) = u_{hq} \). As a special case, \( U_h(W) \) could be a single coordinate of a multivariate \( W \). Alternatively, \( U_h(W) \) could equal \( W \) itself if \( W \) is univariate.

The conditional probability of \( U_hq \) given \( Z_h \) and \( \pi \) is

\[ \text{pr} \left( u_{hq} \mid z_h, \pi \right) = \sum_{S_q} \text{pr} \left( w_i \mid z_h, \pi \right), \]

where \( S_q \) is the subset of \( w_i \) for which \( U_h(w_i) = u_{hq} \). Then (37) becomes

\[ E[\Phi(Y) \mid z_h, \pi] = \sum_{q=1}^{r} \Gamma_h(u_{hq}) \text{pr} \left( u_{hq} \mid z_h, \pi \right) = E[\Gamma_h(U(W)) \mid z_h, \pi]. \] (40)

Having defined \( \Gamma_h(u_{hq}) \), we now prove the following theorem.

**Theorem 5.** Suppose that

i. measurement invariance holds for \( \{Y, Z\} \),
ii. the conditional distribution of \( Y \) given \( Z \) is invariant with respect to selection on \( V \),
iii. \( Y \) and \( W \) are not locally independent when conditioned on \( Z \), and
iv. there exists a function \( \Phi(Y) \) such that

\[ E[\Phi(Y) \mid z_h, w_i] = \Gamma_h[U_h(w_i)] = \Gamma_h(u_{hq}), \]

with \( U_h(W) \) univariate and \( u_{h1} < u_{h2} < \cdots < u_{hr} \). Then
I. for any \( h \) and any two selection functions \( \pi(v) \) and \( \theta(v) \) such that \( \Pr(z_h | \pi) > 0 \), \( \Pr(z_h | \theta) > 0 \), and the conditional distribution of \( U_h(W) \) given \( z_h \) and \( \theta(v) \) is stochastically greater (smaller) than the conditional distribution of \( U_h(W) \) given \( z_h \) and \( \pi(v) \), then \( \Phi(Y) \) which defines \( \Gamma \) is such that \( \Gamma_h(u_{hq}) \) is not strictly increasing in \( u_{hq} \).

II. for any \( h \) and \( \Phi(Y) \) such that \( \Gamma_h(u_{hq}) \) is strictly increasing in \( u_{hq} \), there exist no pairs of selection functions \( \pi(v) \) and \( \theta(v) \) such that the conditional distribution of \( U_h(W) \) given \( z_h \) and \( \theta(v) \) is stochastically greater (smaller) than that of \( U_h(w) \) given \( z_h \) and \( \pi(v) \).

**Proof.** Lemma 1 implies that if both

1. \( \Gamma_h(u_{hq}) \) is strictly increasing in \( u_{hq} \),
2. the conditional distribution of \( U_h(W) \) given \( z_h \), \( \theta(v) \) is stochastically greater (smaller) than that of \( U_h(W) \) given \( z_h \), \( \pi(v) \), then from (40)

\[
E[\Phi(Y) | z_h, \pi] \neq E[\Phi(Y) | z_h, \theta].
\]

But by Assumption ii, this cannot be true, because it implies that the conditional distribution of \( Y \) given \( Z \) is not invariant with respect to selection on \( V \). Hence, Conditions 1 and 2 above cannot hold simultaneously under assumptions (i) through (iv).

The central implication of Theorem 5 is that under measurement invariance of \{\( Y, Z \)\}, invariance of the conditional distribution of \( Y \) given \( Z \) with respect to selection on \( V \) will seldom occur unless \( W \) and \( Y \) are conditionally independent given \( Z \), that is, unless \( Z \) is Bayes sufficient for \( W \). In practice, some functions \( \Phi(Y), U_h(W), \pi(v), \) and \( \theta(v) \) can often be found to contradict statements (a) or (b) in Theorem 5. For example, suppose that \( W \) is a row vector, let \( a \) be a column vector, \( \delta \) a constant, and let \( Y \) be dichotomous (0, 1) such that

\[
\Pr(Y = 1 | W_i) = \frac{\exp(w_i a - \delta)}{1 + \exp(w_i a - \delta)}.
\]

Let \( \Phi(Y) = Y \). If \( Z \) and \( Y \) are locally independent given \( W \) (implying that \( Y \) and \( W \) are not locally independent given \( Z \)), we have

\[
E(Y | z_h, W_i) = \Pr(Y = 1 | W_i),
\]

and \( U_h(W) = Wa \), a linear composite of the components of \( W \). More generally, we can consider situations in which \( a_k \) varies with the value of \( z_h \), giving \( U_h(W) = Wa_h \). A factor analytic example could have \( \Phi(Y) \) be a component of multivariate \( Y \) with \( E[\Phi(Y) | z_h, W_i] = w_i a \) if \( Y \) and \( Z \) are locally independent. As in the previous example, \( U_h(W) = Wa \). In these examples, if selection functions \( \pi(v) \) and \( \theta(v) \) can be found that lead to the required stochastic ordering, I and II are contradicted.

One condition leads naturally to the stochastic ordering required in Theorem 5, as shown in the next theorem. In the following, we will say that \( w_i > w_i^* \) if every component of \( w_i \) is greater than or equal to the corresponding component in \( w_i^* \), with the inequality being strict for at least one component.

**Theorem 6.** Given measurement invariance for \( Z \), suppose that there exist selection functions \( \pi(v) \) and \( \theta(v) \) such that
if \( w_i > w^*_i \) (monotone likelihood ratio). Then, for every value of \( z_h \) such that \( \text{pr}(z_h|\pi) > 0 \) and \( \text{pr}(z_h|\theta) > 0 \),

\[
\frac{\text{pr}(w_i|z_h, \theta)}{\text{pr}(w_i|z_h, \pi)} - \frac{\text{pr}(w^*_i|z_h, \theta)}{\text{pr}(w^*_i|z_h, \pi)} > 0,
\]

if \( w_i > w^*_i \).

**Proof.**

\[
\frac{\text{pr}(w_i|z_h, \theta)}{\text{pr}(w_i|z_h, \pi)} = \frac{\text{pr}(z_h|w_i) \text{pr}(w_i|\theta)}{\text{pr}(z_h|\theta)}
\]

hence

\[
\frac{\text{pr}(w_i|z_h, \theta)}{\text{pr}(w_i|z_h, \pi)} - \frac{\text{pr}(w^*_i|z_h, \theta)}{\text{pr}(w^*_i|z_h, \pi)} = \frac{\text{pr}(w_i|\theta) - \text{pr}(w^*_i|\theta)}{\text{pr}(w_i|\pi) - \text{pr}(w^*_i|\pi)} \frac{\text{pr}(z_h|\theta)}{\text{pr}(z_h|\pi)},
\]

which is non-negative if \( w_i > w^*_i \).

A similar theorem was given by Zwick (1990) when \( W \) represents the true score in classical test theory. Theorem 6 suggests that when \( W \) is univariate, a monotone likelihood ratio for the distribution of \( W \) given \( \theta(v) \) and \( \pi(v) \) implies that the conditional distribution of \( W \) given \( z_h \) and \( \theta \) is stochastically greater than the conditional distribution of \( W \) given \( z_h \) and \( \pi \) for all \( z_h \) such that \( \text{pr}(z_h|\theta) > 0 \) and \( \text{pr}(z_h|\pi) > 0 \). In this monotone likelihood situation, Theorem 5 implies that if (39) is strictly increasing in \( W \), the conditional distribution of \( Y \) given \( Z \) will be invariant only if \( Z \) is Bayes sufficient for \( W \).

To summarize, Theorems 3 through 6 identify some conditions that, given measurement invariance, determine the possible invariance in the conditional distribution of \( Y \) given \( Z \) when Bayes sufficiency does not hold for \( Z \). Theorem 3 illustrates that the conditional distribution of \( Y \) given \( Z \) may be invariant under local independence of \( Y \) and \( Z \). Invariance is possible in this case if \( Z \) is biased. If measurement invariance holds for both \( Y \) and \( Z \), Theorem 4 establishes that local independence of \( Y \) and \( Z \) precludes invariance of the conditional distribution of \( Y \) given \( Z \). Stochastic ordering among the distributions of \( W \) given \( V \) will also preclude this invariance under fairly general conditions, as illustrated in Theorem 5. Finally, Theorem 6 describes a monotone likelihood ratio condition that implies stochastic ordering as required in Theorem 5.

In the next section, we consider conditions that lead to a form of invariance for the conditional distribution of \( Y \) given \( Z \) regardless of measurement invariance in \( \{Y, Z\} \). In the subsequent two sections, we describe two broad conditions leading to Bayes sufficiency of \( Z \).

**Apparent Invariance**

Corollary 1.1. gives necessary and sufficient conditions for the invariance of the conditional distribution of \( Y \) given \( Z \) with respect to selection on \( V \) under all selection
functions \( \pi(v) \). Suppose that these conditions are not fulfilled. Given a particular selection function \( \pi(v) \), can there exist a second selection function \( \theta(v) \) such that

\[
\text{pr} \left( y_g | z_h, \pi \right) = \text{pr} \left( y_g | z_h, \theta \right),
\]

for all \( g, h \) such that \( \text{pr} \left( z_h | \pi \right) > 0 \) and with \( \theta(v) \) not proportional to \( \pi(v) \)? When will such a function exist? When the conditions of Corollary 1.1 are not met, but selection functions \( \pi(v) \) and \( \theta(v) \) exist that satisfy (41), the resulting invariance will be referred to as “apparent invariance”. Apparent invariance of the conditional distribution of \( Y \) given \( Z \) is illusory in that the “invariance” is confined to a specific set of selection functions. Under apparent invariance, comparisons of the conditional distribution of \( Y \) given \( Z \) among groups defined by \( \pi(v) \) and \( \theta(v) \) could lead to erroneous conclusions of measurement invariance for \( Y \). Hence, conditions leading to apparent invariance are of practical interest.

In practice, interest is restricted to “complementary” selection functions (e.g., in comparing a subpopulation of males to a subpopulation of females). Suppose the components of \( V \) and \( \pi \) are ordered such that \( \pi' = \{ \pi_1' : 0' : 0' \} \) and consider a homologically partitioned \( \theta' = \{ 0' : 0_2' : 0' \} \). Application of \( \pi \) and \( \theta \) would yield two nonoverlapping subpopulations from the total or parent population. Define the \( p \times 1 \) vectors \( s_{gh} = \{ \text{pr} \left( y_g, z_h | v_j \right) \} \) and \( s_h = \{ \text{pr} \left( z_h | v_j \right) \} \), and let \( D_v \) be the \( p \times p \) diagonal matrix \( D_v = \text{diag} \{ \text{pr} \left( v_j \right) \} \). Partition \( s_{gh} \) as

\[
s'_{gh} = \{ s'_{gh(1)} | s'_{gh(2)} | s'_{gh(3)} \},
\]

homologically with \( \pi \) and \( \theta \) above, and similarly partition \( s_h \). Finally, partition \( D_v \) as

\[
D_v = \begin{bmatrix}
D_1 & 0 & 0 \\
0 & D_2 & 0 \\
0 & 0 & D_3
\end{bmatrix}.
\]

One may write

\[
\text{pr} \left( y_g | z_h, \pi \right) = \frac{s'_{gh(1)} D_1 \pi_1}{s'_{h(1)} D_1 \pi_1}, \tag{42}
\]

and

\[
\text{pr} \left( y_g | z_h, \theta \right) = \frac{s'_{gh(2)} D_2 \theta_2}{s'_{h(2)} D_2 \theta_2}, \tag{43}
\]

for all \( g, h \). We have the following theorem.

**Theorem 7.** Suppose

i. there exists a matrix \( A \) such that \( s'_{gh(2)} = s'_{gh(1)} A \) for all \( g, h \), and

ii. \( A \) has a right inverse, \( AA^{-1} = I \), all of whose elements are positive.

Then, for every \( \pi' = \{ \pi_1' : 0' : 0' \} \), there exists a \( \theta' = \{ 0' : 0_2' : 0' \} \) such that

\[
\text{pr} \left( y_g | z_h, \pi \right) = \text{pr} \left( y_g | z_h, \theta \right),
\]

for all \( g, h \).

**Proof.** Let \( \theta_2 = aD_2^{-1} A^{-1} D_1 \pi \), where \( a \) is a positive constant chosen so that the largest element in \( \theta_2 \) is less than or equal to one. Note that every element of \( \theta_2 \) is
non-negative so \( \theta \) is an admissible selection function. Substitution in (29) using (i) and (ii) yields

\[
\Pr(y_g | z_h, \theta) = \frac{s_{gh(1)} A D_2 D_2^{-1} A^{-1} D_1 \pi_1}{s_{h(1)} A D_2 D_2^{-1} A^{-1} D_1 \pi_1} = \frac{s_{gh(1)} D_1 \pi_1}{s_{h(1)} D_1 \pi_1} = \Pr(y_g | z_h, \pi),
\]

for all \( g, h \).

Note that Theorem 7 requires that the vectors \( s_{gh(2)} \) all lie in the subspace spanned by the vectors \( s_{gh(1)} \).

An alternative theorem goes as follows:

**Theorem 8.** For a given \( \pi' = \{\pi'_1 : \theta' : \theta'\} \), suppose there exist vectors \( a \) and \( b_{gh} \) such that

i. \( s_{gh(2)} D_2 = (s'_{gh(1)} D_1 \pi_1) a' + b'_{gh} \) for all \( g, h \),

ii. where \( a \) is a vector of positive numbers,

iii. and \( a' b_{gh} = 0 \) for all \( g, h \).

Then there exists a \( \theta' = \{\theta'_1 : \theta'_2 : \theta'\} \) such that

\[
\Pr(y_g | z_h, \pi) = \Pr(y_g | z_h, \theta),
\]

for all \( g, h \).

**Proof.** Let \( \theta_2 = f a/a' a \), where \( f \) is chosen so that the largest element in \( \theta_2 \) is less than or equal to one, and substitute in (43) using i, ii, and iii.

Theorem 8 requires that the \( km \) vector composed of the elements of \( s'_{gh(1)} D_1 \pi_1 \) lies in the column space of the matrix whose rows are the \( s'_{gh(2)} D_2 \).

Theorems 7 and 8 establish conditions leading to apparent invariance for complementary selection functions. Similar conditions can be developed for more general selection functions, but these will not be pursued here. None of the conditions in Theorems 7 and 8 require assumptions about \( W \), or its relation to \( Y, Z \), or \( V \). Apparent invariance can hold when measurement invariance fails for \( Y \) and/or \( Z \).

**Asymptotic Bayes Sufficiency**

Theorem 2 and its corollaries demonstrate the importance of Bayes sufficiency of \( Z \) in determining the possible equivalence between measurement invariance of \( \{Y, Z\} \) and invariance of the conditional distribution of \( Y \) given \( Z \). In this section and the following, consideration is given to some conditions leading to Bayes sufficiency beginning with the relationship of \( Z \) to \( W \) in very long tests.

Suppose that \( Z \) is a perfectly reliable measure of \( W \) in the sense that the joint distribution of \( Z \) and \( W \) is degenerate. Let \( P_{zw} \) denote the \( m \times n \) matrix \( P_{zw} = \{\Pr(z_h, w_i)\} \). Degeneracy implies \( m > n \), every column of \( P_{zw} \) contains only one nonzero element, and every row of \( P_{zw} \) contains no more than one nonzero element. Degener-
acy also implies that measurement invariance holds for $Z$, and that $Z$ is Bayes sufficient for $W$.

Under very general conditions, when $Z$ is an unweighted total test score, we would expect the reliability of $Z$ to approach unity as test length increases. To illustrate, suppose $X_0, X_1, X_2, \ldots, X_q, \ldots$, denote a countably infinite set of dichotomous items that are locally independent when conditioned on a univariate latent trait $W$, and $V$. (See Junker, 1990; Shealy & Stout, 1990; and Stout, 1990; for similar developments in the multivariate, essentially independent case). Let

$$\Phi_{qj}(w_i) = \text{pr} \left( X_q = 1 \mid w_i, v_j \right),$$

for $q = 0, 1, \ldots, j = 1, \ldots, p$, be the item response functions. Further suppose that

$$\lim_{r \to \infty} \left( \frac{1}{r} \sum_{q=1}^{r} \Phi_{qj}(w_i) = w_i^*, \right)$$

with $w_i > w_i', w_i^* > w_i'^*$, for all $i, i'$. Note the assumption that the limit is the same for all values of $v_j$ which would occur if items $X_1, X_2, \ldots$ were measurement invariant, or if only a finite (small) number of items were biased, or if separate biases were somehow counteractive. Now, let $Y = X_0$ and

$$Z_r = \frac{\sum_{q=1}^{r} X_q}{r}.$$

**Theorem 9.** For each and every fixed value of $\{w_i, v_j\}$, the sequence $Z_1, Z_2, \ldots, Z_T, \ldots$ converges almost surely to $w_i^*$.

**Proof.** Follows directly from the Kolmogorov Theorem (Strong Law of Large Numbers) (Rao, 1973).

**Corollary 9.1.** The limiting joint distribution of $Z_r$ and $W$ is degenerate.

**Proof.** From Theorem 9 it follows that for each value $w_i$,

$$\text{pr} \left( \lim_{r \to \infty} Z_r = w_i^* \right) = 1.$$

These results suggest that to detect bias in $Y$, it is reasonable to perform subgroup comparisons of the conditional distributions of $Y$ given a highly reliable (long) $Z$. Two caveats should be noted. When $Z$ is univariate but $W$ is multivariate, it is unlikely that degeneracy of the joint distribution of $Z$ and $W$ will be achieved even in long tests. Bayes sufficiency may not be reached asymptotically in this case. Secondly, statistical significance of subgroup comparisons may occur in large samples even when no bias is present.

**Exponential Families**

We now consider a large class of parametric models relating $\{Y, Z\}$ and $W$ that lead to Bayes sufficiency. This class is of practical interest because it includes the logistic
item response models often used in dichotomous item data. All of the situations to be examined require that \( Y \) be "part of" \( Z \), and we consider only the case of a univariate latent trait \( W \). Generalizations to the multivariate case are possible. Let \( X = \{X_1, X_2, \ldots, X_q, \ldots, X_r\} \) be such that the components of \( X \) are locally independent when conditioned on \( W \). Further assume that \( X \) is measurement invariant. Let

\[
\text{pr} \left(X_q | w_i \right) = \exp \left\{ A_q(x_q) B(w_i) + C_q(x_q) + D_q(w_i) \right\},
\]

where \( A, B, C, \) and \( D \) denote functions of their argument. Further suppose that \( B(w_i) \) is a one-to-one function, and simplify by writing

\[
\text{pr} \left(X_q | w_i \right) = \exp \left\{ A_q(x_q) w_i + C_q(x_q) + D_q(w_i) \right\}.
\]  

(44)

The conditional distribution of \( X_q \) is a member of the exponential family with natural parameter \( w_i \). Let \( Y = X_1 \), and

\[
Z = \sum_{q=1}^{r} A_q(X_q),
\]

with realizations \( z_h, h = 1, \ldots, m \). We have the following theorem:

**Theorem 10.** Under the foregoing conditions,

\[
\text{pr} \left(z_h | w_i \right) = \exp \left\{ z_h w_i + C(z_h) + D(w_i) \right\},
\]  

(45)

where

\[
D(w_i) = \sum_{q=1}^{r} D_q(w_i),
\]

and \( C(z_h) \) is such that

\[
\sum_{h=1}^{m} \text{pr} \left(z_h | w_i \right) = 1.
\]

**Proof.** The characteristic function of \( Z \) is

\[
\prod_{q=1}^{r} \exp \{D_q(w_i) - D_q(w_i + it)\} = \exp \left\{ \sum_{q=1}^{r} D_q(w_i) - \sum_{q=1}^{r} D_q(w_i + it) \right\}
\]

\[
= \exp \{D(w_i) - D(w_i + it)\},
\]

which in turn is the characteristic function of a random variable \( Z \) with \( \text{pr} \left(z_h | w_i \right) \) as stated. \( \square \)

As an immediate corollary, there is:

**Corollary 10.1.** The joint conditional distribution function of \( X_q, q = 1, \ldots, r \), given \( Z = z_h \) is

\[
\text{pr} \left(x_1, x_2, \ldots, x_q | z_h \right) = \exp \left\{ \sum_{q=1}^{r} C_q(x_q) - C(z_h) \right\},
\]

\[\text{pr} \left(x_1, x_2, \ldots, x_q | z_h \right) = \exp \left\{ \sum_{q=1}^{r} C_q(x_q) - C(z_h) \right\},\]
Proof. Follows directly by substitution using (44), (45), and local independence.

It is well-known that within the exponential family, $Z$ has the following property.

**Theorem 11.** Under the foregoing conditions, $Z$ is sufficient for each value of $W = w_i$.

**Proof.**

\[
\begin{align*}
\text{pr}(x_1, x_2, \ldots, x_q | w_i) &= \prod_{q=1}^r \exp \left\{ A_q(x_q)w_i + C_q(x_q) + D_q(w_i) \right\} \\
&= \exp \left\{ \sum_{q=1}^r C_q(x_q) \right\} \exp \left\{ z_h w_i + D(w_i) \right\}.
\end{align*}
\]

By the factorization theorem, $Z$ is Bayes sufficient for $W$.

**Corollary 11.1.** Defining $Y$ and $Z$ as above,

\[
\text{pr}(y_{\bar{g}} | z_h, w_i) = \text{pr}(y_{\bar{g}} | z_h).
\]

**Proof.** Follows directly from Corollary 2.2.

As an example, suppose that $X_q$, $q = 1, \ldots, r$, are dichotomous item responses with logistic item response functions

\[
\text{pr}(X_q = 1 | w_i) = \frac{\exp \{ a_q w_i + c_q \}}{1 + \exp \{ a_q w_i + c_q \}}.
\]

One can rewrite

\[
\text{pr}(X_q = 1 | w_i) = \exp \{ a_q x_q w_i + c_q w_i - \ln [1 + \exp \{ a_q w_i + c_q \}] \},
\]

which is a special case of the general exponential family. Let

\[
Z = \sum_{q=1}^r a_q X_q,
\]

be the Bayes sufficient statistic for $W$ in the logistic test model (Lord & Novick, 1968) and let $Y = X_1$. Further, let $T(z_h)$ denote the subset of realizations of $x_q$, $q = 1, \ldots, r$ such that $Z = z_h$ for $\{x_1, x_2, \ldots, x_r\} \in T(z_h)$. Then for this test model, there is the following theorem and corollary.

**Theorem 12.** For the logistic test model,

\[
\text{pr}(z_h | w_i) = G(w_i) \left[ \exp \left( z_h w_i \right) \right] \left[ \sum_{T(z)} \exp \left\{ \sum_{q=1}^r c_q x_q \right\} \right].
\]

**Proof.** The proof follows directly from Theorem 12 with
\[ G(w_i) = \exp \left[ \sum_{q=1}^{r} \ln \left( 1 + \exp \{ a_q w_i + c_q \} \right)^{-1} \right] \]
\[ = \prod_{q=1}^{r} \left( 1 + \exp \{ a_q w_i + c_q \} \right)^{-1}, \]
\[ C(z_h) = \ln \left[ \sum_{T(z_h)} \exp \left\{ \sum_{q=1}^{r} c_q x_q \right\} \right]. \]

**Corollary 12.1.** The conditional distribution of \( X_q, q = 1, \ldots, r \), given \( Z \) is
\[ \Pr (x_1, x_2, \ldots, x_r \mid z_h) = \frac{\exp \left\{ \sum_{q=1}^{r} c_q x_q \right\}}{\sum_{T(z)} \exp \left\{ \sum_{q=1}^{r} c_q x_q \right\}}, \]
for \( \{x_1, x_2, \ldots, x_r\} \in T(z_h) \) and is independent of \( W \).

**Proof.** Follows directly from Corollary 10.1 and Theorem 12.

Now let \( T(1)(z_h - a_1) \) denote the subset of realizations \( x_q, q = 2, \ldots, r \), such that
\[ z_h = a_1 + \sum_{q=2}^{r} a_q x_q, \]
for \( \{x_2, x_3, \ldots, x_r\} \in T(1)(z_h - a_1) \) and let \( T(1)(z_h) \) denote the subset of realizations \( x_q, q = 2, \ldots, r \), such that
\[ z_h = \sum_{q=2}^{r} a_q x_q. \]

Note that
\[ T(z_h) = T(1)(z_h) + T(1)(z_h - a_1). \]

Then
\[ \sum_{T(z)} \exp \left\{ \sum_{q=1}^{r} c_q x_q \right\} = \exp (c_1) \sum_{T(1)(z - a_1)} \exp \left\{ \sum_{q=1}^{r} c_q x_q \right\} + \sum_{T(1)(z)} \exp \left\{ \sum_{q=1}^{r} c_q x_q \right\} = \exp (c_1) G[c_2, c_3, \ldots, c_r \mid z_h - a_1] + G[c_2, c_3, \ldots, c_r \mid z_h]. \]
With this notation, we have the following theorem.

**Theorem 13.** For \( Y = X_1 \),

\[
\Pr (y_g = 1|z_h) = \frac{\exp (c_1)G[c_2, c_3, \ldots, c_r|z_h - a_1]}{\exp (c_1)G[c_2, c_3, \ldots, c_r|z_h - a_1] + G[c_2, c_3, \ldots, c_r|z_h]},
\]

and

\[
\frac{\Pr (y_g = 1|z_h)}{\Pr (y_g = 0|z_h)} = \frac{\exp (c_1)G[c_2, c_3, \ldots, c_r|z_h - a_1]}{G[c_2, c_3, \ldots, c_r|z_h]}.
\]

**Proof.** Substitute the foregoing in Corollary 12.1 and sum over \( T_{(1)}(z_h - a_1) \). \( \square \)

As a special case of the logistic model, consider the Rasch model for which \( a_q = a \), \( q = 1, \ldots, r \), with \( a = 1 \) without loss of generality. Then \( Z \) assumes the integer values \( \{0, 1, 2, \ldots, r\} \). Let

\[ S_z[\exp (c_1), \exp (c_2), \ldots, \exp (c_r)], \]

and

\[ S_{z-1}[\exp (c_2), \exp (c_3), \ldots, \exp (c_r)], \]

denote the elementary symmetric functions. The result of Theorem 13 becomes

\[
\Pr (y_g = 1|z_h) = \frac{\exp (c_1)S_{z-1}[\exp (c_2), \exp (c_3), \ldots, \exp (c_r)]}{S_z[\exp (c_1), \exp (c_2), \ldots, \exp (c_r)]},
\]

as shown by Holland and Thayer (1988). In this situation,

\[
0 < \Pr (y_g = 1|z_h) < 1 \text{ if } 0 < z_h < r,
\]

\[
\Pr (y_g = 1|z_h = 0) = 0,
\]

\[
\Pr (y_g = 1|z_h = r) = 1.
\]

As Holland and Thayer have shown, if items 2 through \( r \) fit a Rasch model with the \( c_q \) identical across selected groups, then the item-total test regression functions \( \Pr (y_g = 1|z_h) \) for \( Y = X_1 \) can be compared to detect DIF.

Deviations from this Rasch parameterization that disrupt Bayes sufficiency may also disrupt the invariance in the conditional distribution of \( Y \) given \( Z \). To illustrate, consider the following example. Suppose that we have a logistic model with \( a_1 = 1/2 \), \( a_q = 1 \) for \( q = 2, \ldots, r \), \( c_q = 0 \) for \( q = 1, \ldots, r \), and let \( Y = X_1, Z = \sum_{q=1}^{r} X_q \), and \( W^* = \exp (W/2) \). Then,

\[
\Pr (z_h = 1|w^*_i) = \frac{[w^*_i + (r - 1)w^*_i^2]}{(1 + w^*_i^2)(1 + w^*_i^2)^{r-1}},
\]

and

\[
\Pr (y_g = 1|z_h = 1, w^*_i) = [1 + (r - 1)w^*_i]^2^{-1},
\]
which is clearly a function of $W^*$ (or $W$). In this case, the conditional distribution of $Y$ given $Z$ will generally not be invariant with respect to selection on $V$. Since $\Pr(Y_g = 1|z_h = 1, w*_q)$ is a strictly decreasing function of $w*$, lack of invariance is generally assured if the prior distributions are stochastically ordered. Similar results hold for $Z = 2, 3, \ldots, r - 1$. In this example, the hypothesis of measurement invariance would generally be rejected although no bias is present.

As a second example, assume we have a logistic guessing model with

$$\Pr(u_r = 1|w_i) = \frac{\exp \{aq(w_i - b_q)\} + k_q}{1 + \exp \{aq(w_i - b_q)\}},$$

Assume that $a_q = 1, b_q = 0$ for $q = 1, \ldots, r$, and let $Y = X_1, Z = \sum_{q=1}^r X_q$, and $W^* = \exp(W)$. Then,

$$\Pr(z_h = 1|w*_q) = (1 + w*_q)^{-r} \sum_{q=1}^r \left( (w*_q + k_q)(1 - k_q)^{-1} \prod_{m=1}^r (1 - k_m) \right),$$

and

$$\Pr(y_g = 1|z_h = 1, w*_q) = \frac{w*_q + k_1}{1 - k_1},$$

$$\sum_{q=1}^r \left[ \frac{w*_q + k_q}{1 - k_q} \right]$$

which is a strictly decreasing function of $W$. Analogous, but more complex, results hold for other values of $Z$. Again, stochastic ordering of the prior distributions $\Pr(w_i)$ will generally lead to rejection of the hypothesis of measurement invariance, although no bias exists. The above arguments hold even if $k_q = 0$ for $q = 2, \ldots, r$ but $k_1 \neq 0$, or if $k_1 = 0$ but $k_q \neq 0$ for $q = 2, \ldots, r$. If $k_q = k$ for $q = 1, \ldots, r$, we can make the change in variable $W^* = \exp(W) - k$, and obtain

$$\Pr(y_g = 1|z_h = 1, w*_q) = \frac{1}{r},$$

and so forth.

Discussion

The foregoing results have demonstrated that methods for studying bias that rely exclusively on manifest variables are not generally diagnostic of measurement bias, or the lack of bias. Corollary 1.1 establishes that measurement invariance of $\{Y, Z\}$ is not required for invariance of the conditional distribution of $Y$ given $Z$. Given measurement invariance for $\{Y, Z\}$, Theorem 2 establishes that the necessary conditions for invariance of the conditional distribution of $Y$ given $Z$ are restrictive. In particular, Bayes sufficiency of $Z$ with respect to $W$ is often required in practice when the distributions of $W$ are stochastically ordered. We have also shown that certain conditions preclude invariance of the conditional distribution of $Y$ given $Z$ (e.g., Theorem 4).

Bayes sufficiency (or "near" Bayes sufficiency) of $Z$ for $W$ will occur when $Z$ is a highly reliable measure of $W$. If $Z$ is a sum of item scores, we might expect the reliability of $Z$ to increase with the number of items. A question to be investigated
concerns the lower limit for reliability of $Z$ that will yield near invariance of the conditional distribution of $Y$ in the measurement invariant case. Also, what is the lower limit of reliability that will allow accurate detection of bias when $Y$ is not measurement invariant? Simulation evidence will be necessary to answer these questions.

A second condition leading to Bayes sufficiency of $Z$ for $W$ comes about when $\text{pr}(Y, Z | W)$ is a member of the exponential family, and $Z$ is the sufficient statistic for $W$ within this parameterization. This case is of particular interest given the widespread use of logistic latent trait models in mental test theory. In the general logistic case, the sufficient statistic $Z$ is a weighted sum of item scores. Even with very large samples, the proportion of examinees for which $Z = z$ will be negligibly small if more than a dozen items are involved. In addition, inaccuracy in the estimation of the coefficients in the composite will degrade the sufficiency of $Z$. Thus, there are serious difficulties in the way of the use of $Z$ in the general logistic case. This paper has presented results on the exponential family only for the case of a univariate $W$. It should be noted that when $W$ is multivariate, sufficiency generally will require a multivariate $Z$.

One might note that for any test theory model, the item response vector or pattern $(X_q, q = 1, \ldots, r)$ is sufficient for $W$, and one can let $Z = (X_q, q = 1, \ldots, r)$ and $Y = X_1$. But for any $Z = z$, $\text{pr}(y = 1 | z)$ is one or zero depending on whether $X_1$ is one or zero in $Z$ for any and all groups. Grouping on the basis of $Z$ destroys the sufficiency property.

One is then left with the special case of the Rasch model, with $Z$ an unweighted sum of item scores. If the sum leading to $Z$ omits $Y$, Theorem 4 establishes that in the measurement invariant case, the conditional distribution of $Y$ given $Z$ will not be invariant. This result suggests that $Z$ should include $Y$, as concluded by Holland and Thayer (1988). It has been shown that violations of the Rasch model assumptions can lead to variation in the conditional distribution of $Y$ given $Z$ over selection on $V$ even if $(Y, Z)$ are measurement invariant.

In sum, bias detection procedures which rely exclusively on manifest variables are not generally diagnostic of bias, or lack of bias. Logical alternatives to manifest variable procedures are procedures that model the manifest variables in terms of latent variables, and attempt to assess measurement invariance directly. Clearly, the success of these latent variable procedures depends heavily on having the correct model. When the chosen model is inadequate, use of the model may lead to incorrect decisions regarding bias, or the lack of bias. The numerous practical difficulties of estimation and fit assessment in latent variable procedures have not been discussed. In spite of these difficulties, the results presented here suggest that progress in the development of accurate bias detection procedures will most likely follow from refinement of latent variable procedures.

References

Manuscript received 7/31/90
Final version received 8/14/91