Advances in Educational and Psychological Testing: Theory and Applications

edited by
Ronald K. Hambleton
Jac N. Zaal

Kluwer Academic Publishers
Boston/London/Dordrecht
Historically, the use of tests has its roots in the necessity for selection and placement decisions in education, the army, and public administration. This is demonstrated in DuBois’s (1970) historiography of such cases as Binet’s pioneering work on developing a test for the assignment of retarded children to special education, the testing of conscripts for placement in the army during World War I, and the examination of applicants for the civil service in ancient China. It is no coincidence that, in each of these fields, decision making is characterized both by high visibility and a massive number of examinees. In such cases it seems logical to use tests on which to base decisions.

Inspired by these early successes, decision makers have been using tests ever since. Nowadays, the use of tests has pervaded such fields as the admission of students to schools, the selection of personnel in public as well as private settings, the assignment of clients to therapeutic treatments, the choice of careers in vocational guidance situations, pass-fail decisions in instructional units, certification, personnel review, tracking decisions in individualized study systems, and the evaluation of training programs. Although their contents and format may vary, it is inconceivable that the use of tests will ever leave these fields.
It is conspicuous that, although the practice of testing has its roots firmly in decision making, test theory has been developed mainly as a theory of measurements. This was already manifest in Spearman’s pioneering work on what is now known as classical test theory. In this theory, test scores are modeled as a linear combination of a true score and an error of measurement, and the concern is primarily in quantities such as the reliability coefficient, the standard error of measurement—as well as in their properties as a function of test length, item selection, and the like. Modern item response theory shows the same concern with measurement (ability estimation), and was not conceived as a theory of decision making either. The history of test theory shows a few exceptions, though, of which the publication of the Taylor-Russell (1939) tables, with their subsequent influence on the testing literature, and Cronbach and Gleser’s (1965) Psychological Tests and Personnel Decisions deserve special mention. To date, the latter has been the first and only monograph attempting to provide test-based decision making with a sound theoretical basis.

Recently, however, the situation has changed somewhat, and some test theorists are now seriously involved in attempts at modeling and optimizing the use of test scores for decision making. A major impetus for this change has come from the introduction of modern instructional systems such as individualized instruction, learning for mastery, and computer-aided instruction. In such systems, testing primarily serves instructional decision making, and an important task of their developers is to design optimal decision procedures. A seminal paper by Hambleton and Novick (1973) was the first to point at the challenge of these developments to classical test theory.

Although test theory has long ignored decision problems, at a more abstract level the study of optimal rules for decision making has had an extensive tradition in statistics dating back to early publications such as von Neumann and Morgenstern (1944) and Wald (1950). More up-to-date treatments of statistical decision theory can be found in DeGroot (1970) and Ferguson (1967); a short but excellent introduction is given in Lindgren (1976, ch. 8). The primary intention of this chapter is to demonstrate how the various types of decision-making problems in testing can be solved using the framework of statistical decision theory. In particular, in doing so, an (empirical) Bayesian point of view will be assumed.

In the next sections, first some decision-theoretic notions will be introduced. Then, a classification of all possible types of test-based decisions will be given. Subsequently, the main part of this chapter, showing how problems with respect to the various types of decisions can be solved using Bayesian decision theory, is addressed. In the final section, some results
will be presented on the application of decision theory to the problem of simultaneous optimization of combinations of different types of decisions in individualized instruction systems.

Elements of a Statistical Decision Problem

Statistical decision problems arise when a decision maker is faced with the need to choose a preferred action; the outcome of the action depends on the state of nature about which only partial information is available. A simple example is the vacationer who has to decide whether or not to go to the beach, but must rely on a forecast for information about the weather; a more sophisticated one is the researcher who has to decide on the basis of sample data which of his/her hypotheses holds for a given population.

The set of all possible states of nature relevant for the decision problem is known as the state space in decision theory. This set will be denoted by \( \Omega \); whereas a numerical parameter \( \omega \) will be used to index the individual states in \( \Omega \). (When \( \Omega \) is discrete, \( \omega_s, s = 1, \ldots, S \), will be used.) Let \( A \) be the set of all possible actions from which the decision maker can choose. Technically, \( A \) is known as the action space. Individual actions will be denoted by \( a \) (continuous action space) or \( a_t, t = 1, \ldots, T \), (discrete action space). For each action \( a \in A \), the decision maker is confronted with certain consequences. These consequences depend not only on the action chosen but also on the (unknown) state \( \omega \in \Omega \) nature is in. Some of these consequences may be positive; others negative. It is supposed that the decision maker is able to summarize the consequences of his/her actions given the true state of nature into an evaluation on a numerical scale. Because this scale is assumed to run from negative to positive evaluations, it is what is technically known as a utility scale. So in the following, the existence of a utility function \( u(\omega, a) \) on \( \Omega \times A \) is supposed. For each possible combination of an action and a state of nature, this function indicates how positively the decision maker evaluates the outcomes. When \( A \) is discrete, the utility function will be notated by \( u_\omega(a) \).

The decision problem would be easy if the true state of nature were known. If nature is in state \( \omega_n \), the best action is the one for which \( u(\omega_n, a) \) is maximal. In most decision problems, however, nature does not fully disclose its true state; all we have at hand are fallible data—for example, information from a sample, or subjective beliefs. A desirable way of formalizing this is to assume the existence of a random variable \( Z \) representing the outcome of some experiment or measurement conducted to get known the true state of nature, and whose distribution depends on
\( \omega \). In the following it is assumed that the family of probability distributions of \( Z \) with distribution functions \( F(z) \equiv F_Z(z; \omega) \) is known.

If the true state of nature is not known with certainty, the decisions based on \( \mathcal{A} \) are likely to be less than optimal. A rational approach, then, is to look for a decision rule that optimizes the outcomes across repetitions of the same decision problem, but with nature in possibly different states. The main purpose of decision theory is to find such rules and to study their properties. Formally, a (nonrandomized) decision rule is a prescription specifying for each possible value \( z \) of \( Z \) which action \( a \in \mathcal{A} \) has to be taken. Hence, a proper notation is to write the decision rule \( \delta \) as a mapping from the data \( Z \) to the action space \( \mathcal{A} : A = \delta(Z) \). Due to the fact that \( Z \) is a random variable, using \( \delta \) implies that the actions are also random. At first sight this may seem embarrassing. However, there is no way to escape the random nature of our data about \( \omega \). Moreover, the decision maker is free to choose whatever rule we would like to have from the set of possible mappings from \( Z \) to \( \mathcal{A} \), which gives him/her the opportunity to select an optimal one.

It is obvious that our criterion for selecting an optimal decision rule should have to do with the utilities of the decision outcomes, \( u(\omega, a) \). The function \( A = \delta(Z) \) implies that these utilities must be considered as realizations of a random variable \( U = u(\omega, \delta(Z)) \). In such cases it is natural to replace this variable by its expectation. Therefore, we define the expected utility as:

\[
R(\omega, \delta) \equiv E[u(\omega, \delta(Z))] \quad (1)
\]

\[
= \int u(\omega, \delta(z))dF(z)
\]

If nature is in state \( \omega \) and decision rule \( \delta \) is adopted, then equation (1) shows the utility to be expected. However, the actual actions still depend on the values taken by \( Z \), so their utility usually will vary about the value of equation (1).

**Bayes Rules**

It is still not possible to define a criterion for a uniformly best decision rule \( \delta \) using the expected utility function in equation (1). The reason is its dependency on the unknown value \( \omega \). One way out of this problem would be to make a sensible choice for this value, say \( \omega_0 \), and to select a \( \delta_0 \), such that the expected utility \( R(\omega_0, \delta) \) is maximal. This approach is taken, for instance, in minimax theory where \( \omega_0 \) is selected as a value of \( \Omega \) that represents the least favorable state of nature to the decision maker. As a
consequence, the decision maker is guaranteed that the minimum expected utility for \( \delta_0 \) is never smaller than \( R(\omega_0, \delta_0) \).

Another approach is not to select one special state \( \omega \) of nature, but to assume a distribution function \( G(\omega) \) over \( \Omega \). This may represent the decision maker's subjective probabilities of the possible states of nature or its empirical distribution. For both interpretations \( G(\omega) \) is known as the a priori distribution (or prior) because it represents knowledge about \( \omega \) available before the data \( Z = z \) are observed. Having \( G(\omega) \), we are now able to define the Bayes utility of decision rule \( \delta \) as

\[
B(\delta) = E[R(\omega, \delta)] = \int R(\omega, \delta) dG(\omega).
\]

In the literature, this quantity is also known as the Bayes risk of the decision procedure, although, strictly speaking, this name is only proper if a loss function instead of a utility function is used.

The Bayes utility in equation (2) only depends on the decision rule. It now seems obvious to select from the class of possible rules the one, say \( \delta^* \), that maximizes the Bayes utility:

\[
B(\delta^*) = \max_{\delta} B(\delta).
\]

Rules satisfying equation (3) are known as Bayes rules. Throughout this chapter it will be assumed that the quantities used in the above definition of a Bayes rule exist for the problem at hand (though it will not necessarily be true that the problem has a unique Bayes rule).

**Monotone Bayes Rules**

For the actual maximization in equation (3) it would be helpful if the (possibly infinitively) large set of all possible rules \( \delta \) could be restricted to some subset of a tractable form. In addition to this technical consideration, there is a less lofty reason for which the attention sometimes has to be restricted to a subset of possible rules. This has to do with the acceptability of some types of rules among those involved in the decision procedure. In education, for example, students, teachers, and administrators are familiar with selection decisions in which the decision rules have a monotone form: students are admitted to a program if their grade-point average or their test score are above a certain cutting point and rejected otherwise. It would be a shock to all parties if some institutions changed the form of their selection rule and started, for instance, admitting students with low or high scores while rejecting those with intermediate ones. However, the restriction to
monotone rules is only correct if they constitute an essentially complete class (Ferguson, 1967, p. 55); otherwise rules with a higher expected utility are wrongly ignored.

The conditions under which a class of monotone rules is essentially complete are known (Chuang, Chen, & Novick, 1981; Ferguson, 1967, sect. 6.1). Two conditions have to be met: first, the so-called posterior distribution of ω given Z-z should be stochastically increasing, that is, if F(ω|z) is the distribution function of ω given Z = z, z1 ≥ z2 must imply F(ω|z1) ≤ F(ω|z2) for all ω. Second, there should be an ordering of the actions for which the difference between utility functions for adjacent pairs of actions changes sign at most once. For the decision problems dealt with in this chapter, the conditions will be made more specific below.

A Classification of Test-Based Decision Making

The use of test scores for decision making in education and psychology can be classified in a simple way (van der Linden, 1985a, 1985b). In each of these settings three basic elements can be identified, and each type of decision making can be viewed as a specific configuration of these elements. In general, four different types of decision making can be distinguished. Further, for each of these types four possible restrictions or refinements can apply. This classification of test-based decision making will now be elaborated. In the next sections it is then shown how Bayesian decision theory can be applied to the problems in this classification.

Basic Elements of Test-Based Decisions

Each type of decision making can be identified as a specific configuration of one or more of the following elements:

1. A test that provides the scores on which the decisions are based;
2. One or more treatments with respect to which the decisions are made;
3. One or more criteria by which the successes of the treatments are measured.

The term test is used here because of the focus of this chapter. It could easily be replaced by any other measuring instrument or source of data without invalidating the content of this chapter. Likewise, treatment is a
generic term here, referring to whatever manipulation, experiment, or program is used to change the condition of individuals. Examples of treatments in educational and psychological settings are: instructional programs, applications of audiovisual-materials, or psychological therapies. It should be noted that information about the success of the treatment on the criterion may be provided by any source. In this chapter it is assumed that the information is quantitative by nature. In practice, the criterion is often measured by another test.

**Types of Decisions**

Four basic types of decisions are distinguished. Each type of decision can be represented by a unique flowchart containing one or more of the above elements.

1. **Selection Decisions.** In selection problems the decision is the acceptance or rejection of individuals for a treatment. A typical feature of the selection decision is that the test is administered before the treatment but that the criterion is measured afterwards. Well-known examples of selection decisions are selection of personnel in industry and admission of students to educational programs. Figure 5–1 gives a flowchart displaying the structure of the selection decision. Selection research has had a long tradition in educational and psychological testing in which the selection decision was viewed as a prediction problem. Until Cronbach and Gleser (1965), the usual approach was to establish regression lines or expectancy tables to predict criterion scores and to accept individuals with predicted criterion scores above a given threshold value.

2. **Mastery Decisions.** Unlike selection decisions, mastery decisions are made after a treatment. The content of the decision is whether individuals

![Flowchart of a Selection Decision](image)
who followed the treatment are successful on the criterion. A further feature is that the criterion is internal and not external to the test. It is the unreliability of the test as a representation of the criterion that creates the mastery decision problem. Due to measurement error, the possibility of false-negative and false-positive decisions exists, and it is the task of the decision maker to minimize their consequences. Figure 5–2 shows the formal structure of the mastery decision problem. Examples of mastery decisions are pass-fail and certification decisions in education.

3. Placement Decisions. In placement problems, several alternative treatments are available and it is the decision maker's task to assign individuals to the most promising treatment. The same test is administered to each individual and the success of each treatment is measured by the same criterion. Placement decisions differ from selection decisions by the fact that more than one treatment is available and that each individual is assigned to a treatment. The case of a placement decision with two treatments is shown in figure 5–3. Examples of placement decisions can be found in individualized instruction where students are allowed to follow different routes through instructional units but regardless of routes, the same criterion is appropriate. Aptitude-treatment interaction (ATI) re-
search has given the main impetus to interest in the placement problem. In ATI research, the traditional approach has been regression analysis with a separate regression line for each treatment assigning individuals to the treatment with the largest predicted criterion score.

4. **Classification Decisions.** As is clear from figure 5-4, the difference between placement and classification decisions is that in the latter, each treatment has its own criterion. Further properties of the two types of decisions are equal. Examples of classification decisions arise in vocational guidance situations where most promising careers or training programs must be identified. The most popular approach to classification decisions has been the use of linear-regression techniques. Each criterion is then mapped on a common utility scale, and the decision rule is to assign individuals to the treatment with the largest predicted utility.

**Further Restrictions and Generalizations**

It should be noted that the above types of decisions are not always met in their pure forms. These decisions often occur in combinations; also, further restrictions or generalizations may apply. An example of a combination of two types of decisions arises in a selection problem where the criterion is unreliable measured. If success on the criterion is defined by a threshold value, then, in fact, after the treatment a mastery decision has to be taken, and the problem is a selection-mastery decision problem. A combination of a selection and a classification problem is met if more than one treatment is available but not all individuals are accepted for a treatment. In individualized instruction—for instance, as implemented in CAI-systems—decision making can be viewed as guiding students through
a network of several of the above decisions. The simultaneous optimization of such networks will be discussed in the final section of the chapter.

For each basic type of decision, one or more of the following restrictions or refinements may apply:

1. **Multivariate Test Scores.** Instead of a single test, a battery of tests may be used as a basis for the decision. The use of test batteries has had a long history in the practice of personnel selection and vocational guidance. Formally, the use of test batteries implies decision rules defined on a vector of test scores instead of a single score.

2. **Sequential Testing.** In a sequential testing strategy, test items are administered until a decision can be made with a desired level of certainty (see, for example, van der Linden & Zwarts, 1989). The recent introduction of the computer in educational and psychological testing has stimulated the interest in sequential testing strategies for decision making (see the Hambleton, Zaal, & Pieters chapter). If tests are used in this mode, sequential Bayesian procedures have to be used (Lindgren, 1976, sect. 8.5).

3. **Multiple Criteria.** In some applications, the success of a treatment has to be measured on more than one criterion. Each individual criterion is then supposed to reflect a different aspect of the treatment. Formally, the presence of multiple criteria implies the necessity to define utility functions on a vector of criterion scores instead of a single score.

4. **Multiple Populations.** The presence of different populations of examinees reacting differently to the test items may create the problem of "fair" decision making. In education, the problems of fair selection and mastery decisions have been struggled with for a long time, in particular for populations defined by race or sex. Formally, the presence of populations reacting differently to test items implies different probability distributions of test and criterion scores for each population. In addition, the decision maker may have different utilities associated with different populations. As a result, a separate decision rule has to be established for each population.

5. **Quota restrictions.** So far it has been assumed that the number of vacancies in each treatment is unlimited. Due to the shortage of resources, however, these numbers are often constrained. Consequently, Bayes rules for quota-restricted decisions have to be found by methods of constrained optimization.
Conclusion

The above classification of test-based decisions shows four basic types of decisions that may occur separately or in combination. Furthermore, for each decision one or more refinements or generalizations may hold. In addition, the utility structure and probability distributions may vary from problem to problem. However, the key to finding Bayes rules for each possible decision is still the optimization of its Bayes utility, which will now be illustrated for a sample of decision problems.

Selection Decisions with Linear Utility

In selection decisions, test scores are used to decide on the acceptance or rejection of individuals for a treatment, with success measured on a future criterion. In order to apply the framework of Bayesian decision theory, the "unknown state of nature" should now be interpreted as the individual's unknown criterion score, and the "data" about this state are provided by the test score. For a randomly sampled individual, let the criterion be a continuous random variable \( Y \), with as possible states, success \( (Y \geq y_c) \) and failure \( (Y < y_c) \). The test score is assumed to be a discrete random variable \( X \) with possible values \( x = 0, 1, \ldots, n \) (number-right score). The information in \( X \) on \( Y \) is given by a joint probability function \( k(x, y) \). Since the conditions for a monotone decision are assumed to be met, the Bayes rule for the selection problem is a cutoff score \( x_c \), with an acceptance and a rejection decision for \( X \geq x_c \) and \( X < x_c \), respectively.

Formally, utility is a function defined on the true state of nature with a possibly different form for each action. A moment's reflection shows that in the present problem utility should be an increasing function of the criterion for the acceptance decision, but a decreasing function for the rejection decision: the higher the criterion score of an accepted individual, the higher the utility of the decision; whereas the opposite holds for a rejected individual. A linear utility function that meets this property is given in van der Linden and Mellenbergh (1977):

\[
u(y) = \begin{cases} 
  b_0(y_c - y) + a_0 & \text{for } x < x_c \\
  b_1(y - y_c) + a_1 & \text{for } x \geq x_c 
\end{cases}
\]

for \( b_0, b_1 > 0 \).

This function, which is shown in figure 5-5, consists of two additive components:

1. \( b_0(y_c - y) \) and \( b_1(y - y_c) \) represent amounts of utility dependent
on the difference between the criterion score and success threshold $y_c$ with constants of proportionality $b_0$ and $b_1$.

2. $a_0$ and $a_1$ are amounts of utility independent of the criterion score but dependent on the decision. They can be used, for instance, to allow for treatment costs.

When sampling individuals from the population, the Bayes utility for decision rule $x_c$ is equal to

$$R(x_c) = \sum_{x=0}^{x_c-1} \int [b_0(y_c - y) + a_0] k(x, y)dy + \sum_{x=x_c}^{n} [b_1(y - y_c) + a_1] k(x, y)dy. \quad (5)$$

Using $k(x, y) = g(y|x)h(x)$, $\int g(y|x)dy = 1$ and $\int yg(y|x)dy = E(Y|x)$, it follows

$$R(x_c) = \sum_{x=0}^{x_c-1} \{b_0[E(Y|x) - y_c] + a_0\} h(x)$$

$$- \sum_{x=x_c}^{n} \{b_1[E(Y|x) - y_c] - a_0\} h(x). \quad (6)$$
where $E(Y|x)$ is the regression function of $Y$ on $X$.

Completing the first sum

$$R(x_r) = \sum_{x=0}^{n} [b_0 E(Y|x) - y_c] + a_0] h(x)$$

$$- \sum_{x=x_r}^{n} (b_0 + b_1) [E(Y|x) - y_c] - (a_0 - a_1)] h(x).$$

(7)

Since the first sum is now a constant, $b_0 + b_1 > 0$, $h(x) \geq 0$ for all $x$, and the monotonicity conditions guarantee that $E(Y|x)$ is increasing in $x$, equation (7) is maximal for the value of the smallest value of $x_r$ for which

$$(b_0 + b_1) [E(Y|x) - y_0] + (a_0 + a_1)$$

(8)

is not negative. If the monotonicity conditions are not strict or it does not hold that $h(x) > 0$ in the neighborhood of the solution, this value of $x$ may not be unique. Throughout this chapter it will be assumed that conditions like these are fulfilled.

The regression function $E(Y|x)$ can easily be estimated by drawing a sample from the population and administering the treatment.

It should be noted that equation (8) is the difference between the two conditional expected utilities given $X = x$ associated with the acceptance and rejection decision. This is clear from inspection of the bracketed terms in equations (6) and (7). The expectations are known as posterior expected utilities; they can be considered the expected utilities after the observation $X = x$ has been made. For a monotone decision it holds in general that the optimal cutoff score is located at the point at which the posterior expected utilities cross (e.g., DeGroot, 1970, sect. 8.9). In the following, this property will be used without further validation.

A closed-form solution exists if the regression function is linear, that is, if

$$E(Y|x) = \beta x + a$$

(9)

Then, it follows from equation (8) that $x_r$ is the smallest value larger than

$$\frac{y_c - \alpha}{\beta} + \frac{a_0 - a_1}{\beta(b_0 + b_1)}$$

(10)

An interesting case arises if $a_0 = a_1$. Under this condition the second term in equation (10) vanishes, and the solution contains the regression parameters only. For this and other properties of the linear utility function, see van der Linden and Mellenbergh (1977).
Multiple Populations

As noted earlier, the presence of multiple populations in selection decisions creates the problem of fair selection if each population reacts differently to the test items. In such a case, the test items are often said to be "biased" against one or more of the populations.

The only thing needed to deal with multiple populations in selection decisions seems to include separate probability distributions of test and criterion scores for each population in the model to allow for differential item properties. However, the problem of fair selection often involves the notion of "disadvantagedness" as well, in particular when some of the populations are defined by race or sex. As Gross and Su (1975) and Novick and Petersen (1977) argue, this aspect of fair selection is only a question of utilities. A selection rule is "fair" if those involved in the decision process accept the utility structure underlying the decision rule. Hence, in addition to separate probability distributions, separate utility functions are needed to allow for different utility structures for the populations.

Now the above selection model can easily be adapted to the case of multiple populations (Mellenbergh & van der Linden, 1981). Let \( i = 1, \ldots, p \) denote the populations in the selection problem. Then, equation (4) has to be replaced by

\[
\begin{align*}
  u(t) & \quad \text{Mastery Decision} \\
  \quad & \quad \text{Nonmastery Decision}
\end{align*}
\]

\[ t_c \quad t \]

Figure 5–6. Threshold Utility Function for a Mastery Decision
APPLICATIONS OF DECISION THEORY

\[ u_i(y) = \begin{cases} 
    b_{0i}(y_c - y) + a_{0i} & \text{for } x < x_{ci} \\
    b_{1i}(y - y_c) + a_{1i} & \text{for } x \geq x_{ci}, \quad b_{0i}, \ b_{1i} > 0
\end{cases} \tag{11} \]

and the probability functions \( k_i(x, y), i = 1, \ldots, p, \) are now allowed to vary across populations. The Bayes utility for a random individual is defined analogous to equation (5) as \( R_i(x_{ci}) \). Let \( s_i \) be the relative size of population \( i \). Then, when sampling randomly from the total population under consideration, the Bayes utility is equal to

\[ R(x_{c1}, \ldots, x_{cp}) = \sum_{i=1}^{p} s_i R_i(x_{ci}). \]

But this is just a weighted sum of the Bayes utilities of the separate populations. Hence, the decision procedure is optimal if, analogous to equations (5) to (8), for each population a separate Bayes rule is derived. An example of the model with real test data is given in Mellenbergh and van der Linden (1981).

**Mastery Decisions with Threshold Utility**

In the mastery decision problem, the unknown state of nature is the individual’s true score on the criterion variable measured by the test. The data are the observed test scores. The true score is defined as the expected proportion of test items a given individual solves correctly. The true score of a random individual is denoted as \( T \), with possible values \( t \in [0, 1] \). A mastery level \( t_c \) is assumed, and an individual is considered to master the criterion if \( t \geq t_c \), and not to master it otherwise. Unlike the selection problem, there is no way whatsoever to measure the criterion variable directly. Hence, a test model is needed to derive the statistical relation of \( X \) to the true score \( T \).

For the purpose of illustration, a threshold utility function is chosen. Although figure 5–6 shows that this function has a jump at \( t_c \) that may be less realistic in some applications, it has been studied extensively in the mastery testing literature (Hambleton & Novick, 1973; Huynh, 1976; Mellenbergh, Koppelaar, & van der Linden, 1977; van der Linden, 1980, 1982; Wilcox, 1977). The threshold utility function is defined by the following four constants:

\[ u(t) = \begin{cases} 
    u_{10} & t < t_c, \ x < x_c \\
    u_{10} & t \geq t_c, \ x < x_c \\
    u_{01} & t < t_c, \ x \geq x_c \\
    u_{11} & t \geq t_c, \ x \geq x_c
\end{cases} \tag{12} \]
However, since the derivation following below holds for any positive linear rescaling of equation (12), it will be assumed for convenience that \( u_{10} = u_{11} = 0 \).

Let the joint distribution of \( X \) and \( T \) be given by the probability function \( k(x, t) \), and let \( g(t|x) = k(x, t)/h(x) \). For the sake of illustration, the optimal cutoff score \( x_c \) will now be derived using a comparison of the posterior expected utilities for the mastery and nonmastery decision.

If the mastery decision is taken for a random individual with test score \( X = x \), the posterior expected utility is equal to

\[
E_i(u(t)|x) = \int_0^1 u(t)g(t|x)dt = u_{01} \int_0^{x_c} g(t|x)dt. \tag{13}
\]

For the nonmastery decision

\[
E_0(u(T)|x) = \int_0^1 u(t)g(t|x)dt = u_{10} \int_{x_c}^{0} g(t|x)dt. \tag{14}
\]

Suppose \( x \) were continuous. Then equations (13) and (14) would cross at the value of \( x \) for which

\[
u_{01} \int_0^{x_c} g(t|x)dt = u_{01} \int_{x_c}^{0} g(t|x)dt \]

or

\[
\int_0^{x_c} g(t|x)dt = \frac{u_{10}}{u_{01} + u_{10}}. \tag{15}
\]

However, \( X \) is a discrete test score. Therefore, the Bayes utility is maximal if \( x_c \) is chosen to be the smallest integer value larger than the solution to equation (15).

It should be noted that this solution holds for any test model providing the probability function in the left-hand side of equation (15), and that it can be calculated only after such a model is specified.

A usual choice in mastery testing is the beta-binomial model (Huynh, 1976; Mellenbergh, Koppelaar, & van der Linden, 1977). In the model it is assumed that (1) the conditional distribution of \( X \) given \( T = t \) is the binomial,

\[
f(x|t) = \binom{n}{x} t^x(1 - t)^{n-x}, \tag{16}
\]

and (2), the marginal distribution of \( T \) is the beta distribution with probability density function

\[
b(t) = B^{-1}(v, w - n + 1)t^{v-1}(1 - t)^{w-n}, \quad v > 0, w > n - 1, \tag{17}
\]
where \( B(v, w - n + 1) \equiv \int_0^1 t^v(1 - t)^{w-n}dt \) is the complete beta function (e.g., Johnson & Kotz, 1970, ch. 24). The choice of equation (16) is motivated by the fact that the responses of a fixed individual to a series of test items can often be described as a sequence of Bernoulli trials, whereas equation (17) defines a flexible family of distributions on \([0,1]\) that contains most true-score distributions occurring in practice. Keats and Lord (1962) have shown that moment estimators of \( v \) and \( w \) exist that are a simple function of \( \mu_X \) and the KR-21 reliability coefficient. They also found a satisfactory fit of the beta-binomial model to test score distributions ranging widely in form.

For integer values of \( v \) and \( w \), it holds that

\[
\int_0^v g(t|x)dt = \int_0^{v+w} f(x|t)b(t)dt = \sum_{\gamma=v+x} f(\gamma|t_c)
\]

(18)

where \( f(\gamma|t_c) \) is the binomial probability function with success parameter \( t_c \) (Johnson & Kotz, 1970, sect. 24.6). Thus, a suitable estimate of the integral in equation (15) can be obtained via a table of the cumulative binomial.

A Numerical Example

Using the beta-binomial model, the result in equation (15) was used to calculate the optimal cutoff score \( x_r \) for tests of length \( n = 20 \), mastery threshold \( t_c = 14 \), and utility ratio \( u_{01}/u_{01} = 1 \). The data were simulated such that the average true score \( \mu_T \) and the KR-21 reliability coefficient varied systematically (van der Linden, 1984). Table 5–1 shows how \( x_r \) depends on these two parameters. For tests of high reliability, \( x_r \) varies hardly with \( \mu_T \), but this robustness is lost quickly for tests of lower reliability. For KR-21 = .05, \( \mu_T \leq 13 \) yields \( x_r \geq 20 \), but the optimal cutoff score drops immediately to below zero for \( \mu_T > 13 \). Only tests of high reliability yield stable cutoff scores.

An unexpected phenomenon in table 5–1 is the opposite direction in which \( x_r \) varies with \( \mu_T \). If the average true score \( \mu_T \) goes down, \( x_r \) goes up. Hence, for low performing populations the cutoff score should be set high, whereas it should be low for high performers. At first sight, this goes against our intuition. However, it is a logical consequence of our criterion of maximal expected utility. An analysis of equations (12) to (15) reveals the following: for high performing populations, almost all individuals are above the mastery threshold \( t_c \). Therefore, the Bayes utility of the decision
Table 5-1. Optimal Cutoff Scores for Populations Varying in Average True Score and Tests Varying in Reliability \((n = 20, \ell_c = 14, u_{10}/u_{01} = 1)\)

<table>
<thead>
<tr>
<th>KR-21</th>
<th>(t_{c})</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>11</th>
<th>13</th>
<th>15</th>
<th>17</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>.95</td>
<td></td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>.80</td>
<td>F</td>
<td>19</td>
<td>17</td>
<td>16</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>.65</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>.50</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>19</td>
<td>19</td>
<td>17</td>
<td>16</td>
<td>13</td>
<td>12</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>.35</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>20</td>
<td>16</td>
<td>13</td>
<td>P</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>.20</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>19</td>
<td>11</td>
<td>P</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>.05</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>20</td>
<td>P</td>
<td>P</td>
<td></td>
</tr>
</tbody>
</table>

Note: \(F\) indicates a cutoff score larger than 20 (all students fail); \(P\) indicates a cutoff score lower than 0 (all students pass).

The procedure tends to consist only of contributions from true masters and false nonmasters. Since the proportion of true masters among these two categories depends on the cutoff score \(x_c\), and the utility of a true master typically is larger than the utility of a false nonmaster, \(x_c\) will take a low value. This phenomenon was dubbed "the regression-from-the-mean effect" in van der Linden (1980). It is a reversion of the well-known regression-to-the-mean effect due to the fact that in the derivation of equation (15) the conclusion goes from the true score \(t_c\) to \(x_c\) instead of the other way around.

Placement Decisions with Normal-Ogive Utility

The typical feature of the placement decision is the presence of more than one treatment and the fact that each individual is assigned to a treatment. As before, the unknown state of nature is the individual’s criterion score after a treatment. Hence, although there is one criterion common to all treatments, each individual has a different unknown state of nature for each criterion.

Let \(Y\) be the criterion common to the treatments \(j = 1, 2, \ldots, m\), and let \(X\) denote the test score again. It is assumed that the treatments are ordered by the strict monotonicity conditions for the placement decision given in van der Linden (1981), so that an optimal cutoff score \(x_j\) exists for the decision between treatments \(j\) and \(j + 1\). For the sake of illustration,
the normal-ogive utility function introduced by Novick and Lindley (1978; see also Berhold, 1973) will be used (see also figure 5–7):

$$u_j(y) = \Phi \left( \frac{y - \mu_j}{\sigma_j} \right),$$

(19)

where $\Phi$ is the normal distribution function with parameters $\mu_j$ and $\sigma_j$.

The relevant quantity is the posterior expected utility for treatment $j$ after test score $X = x$. This will be denoted as $E_j(u_j(Y)|x)$. It is assumed that the conditional distribution of $Y$ given $X = x$ is normal with conditional mean and variance $E_j(Y|x)$ and $\text{Var}_j(Y|x)$, respectively. It follows that

$$E_j(u_j(Y)|x) = \Phi \left( \frac{E_j(Y|x) - \mu_j}{\sqrt{\text{Var}_j(Y|x) + \sigma_j^2}} \right).$$

(20)

The optimal cutoff score $x_j$ is now the smallest value of $x$ for which the posterior expected utility for treatment $j + 1$ is larger than for treatment $j$. That is, the smallest value of $x$ for which

$$\Phi \left( \frac{E_{j+1}(Y|x) - \mu_{j+1}}{\sqrt{\text{Var}_{j+1}(Y|x) + \sigma_{j+1}^2}} \right) - \Phi \left( \frac{E_j(Y|x) - \mu_j}{\sqrt{\text{Var}_j(Y|x) + \sigma_j^2}} \right)$$

(21)

is positive. But this is also the smallest value of $x$ for which
\[
\frac{E_{j+1}(Y|x) - \mu_{j+1}}{[\text{Var}_{j+1}(Y|x) + \sigma^2_{j+1}]^{1/2}} - \frac{E_{j}(Y|x) - \mu_{j}}{[\text{Var}_{j}(Y|x) + \sigma^2_{j}]^{1/2}}
\]

is positive. If it is known how the posterior expectations and variances depend on \( x \), the optimal cutoff score can be found graphically or numerically from equation (22).

Suppose now that the posterior expectations are linear and that homoscedasticity may be assumed. That is,

\[
E_{j}(Y|x) = \beta_{j}x + \alpha_{j}, \quad j = 1, \ldots, m
\]

and

\[
\text{Var}_{j}(Y|x) = \text{Var}_{j}(Y,X) \quad x = 0, \ldots, n
\]

\[
j = 1, \ldots, m,
\]

where \( \text{Var}_{j}(Y,X) = (1 - [\text{Cor}_{j}(Y,X)]^2) \text{Var}_{j}(Y) \) and \( \text{Cor}_{j}(Y,X) \) is the linear correlation coefficient between \( Y \) and \( X \) for \( j \). Substituting equations (23) and (24) into equation (22) yields as a result that \( x_{j} \) is the smallest value larger than

\[
[\epsilon_{j}(\mu_{j+1} - \alpha_{j+1}) - \epsilon_{j+1}(\mu_{j} + \alpha_{j})]/(\epsilon_{j}\beta_{j+1} - \epsilon_{j+1}\beta_{j})
\]

with

\[
\epsilon_{j} = [\text{Var}_{j}(Y,X) + \sigma^2_{j}]^{1/2}.
\]

More theory on placement decisions is given in van der Linden (1981).

**Quota-Restricted Placement**

In the above placement model, it was assumed that the number of vacant places in the treatments was unrestricted. In practice, however, quota restrictions regularly apply, and then a modification of the decision rule is necessary.

Following Chuang, Chen, and Novick (1981), three kinds of quota restrictions are distinguished:

1. Exactly \( N_{j} \) individuals should be assigned to treatment \( j = 1, \ldots, m \), where \( \sum_{j=1}^{n} N_{j} = N \) (number of examinees);

2. At least \( N_{j} \) individuals should be assigned to treatment \( j = 1, \ldots, m \).
where \( \sum_{j=1}^{m} N_j \leq N \);

3. At most \( N_j \) individuals should be assigned to treatment \( j = 1, \ldots, m \).

where \( \sum_{j=1}^{m} N_j \leq N \);

In a (strictly) monotone placement problem, the posterior expected utilities of adjacent pairs of treatments cross at most once in the range of test scores. Therefore, if the \( N \) examinees are arranged in decreasing order of test score, the following placement rules maximize the Bayes utility in the above three cases:

1. Beginning with the highest scoring examinee, \( N_m \) places in treatment \( m \) are filled, then \( N_{m-1} \) places in treatment \( m-1 \), and so on.

2. Suppose the \((k-1)\)th examinee in the order of examinees has been assigned to treatment \( j \). Then the following rule for \( k \) is optimal:
   i. If \( x_k \geq x_{j-1} \), assign \( k \) to treatment \( j \).
   ii. If \( x_k < x_{j-1} \) and treatment \( j \) has not received \( N_j \) examinees, then choose from the following rules the one with the larger posterior expected utility: (a) assign \( k \) to \( j \); (b) assign \( k \) to \( j-1 \), reassigning the lowest scoring examinee in \( j+1 \) to \( j \) (and, if \( j + 1 \) then has fewer than \( N_{j+1} \) examinees, the lowest scoring one in \( j + 2 \) to \( j + 1 \); etc.).
   iii. If \( x_k < x_{j-1} \) and \( j \) has received \( N_j \) examinees, then assign \( k \) to \( j-1 \).

3. Again suppose the \((k-1)\)th examinee has been assigned to treatment \( j \).
   i. If \( x_k < x_{j-1} \), then assign \( k \) to the treatment in \{ \( j - 1, j - 2, \ldots, 1 \) \} with the largest posterior expected utility.
   ii. If \( x_k \geq x_{j-1} \) and treatment \( j \) has received \( N_j \) examinees, then choose from the following rules the one with the larger posterior expected utility: (a) assign \( k \) to \( j - 1 \); (b) assign \( k \) to \( j \) reassigning the highest scoring examinee in \( j \) to \( j + 1 \) (and, if \( j + 1 \) had already \( N_{j+1} \) examinees, the highest scoring one in \( j + 1 \) to \( j + 2 \); etc.).
   iii. If \( x_k \geq x_{j-1} \) and \( j \) has not received \( N_j \) examinees, then assign \( k \) to \( j \).

Classification Decisions with Threshold Utility

The same notation as for the problem of the placement decision will be used. However, because every treatment now has its own criterion, the criterion variable is treatment dependent and will be denoted as \( Y_j \).
Suppose that for each criterion a success threshold \( d_j \) can be defined. Then the following threshold utility function may be a proper choice:

\[
u_j(y_j) = \begin{cases} w_j & \text{for } y_j \geq d_j \\ v_j & \text{for } y_j < d_j \end{cases}
\]  

(26)

with

\[ w_j > v_j, \quad j = 1, \ldots, m. \]

The last condition simply states that the utility of a success on treatment \( j \) is larger than the one of a failure.

Let \( \Phi_j(d_j|x) \equiv \int_{d_j}^{\infty} g_j(y_j|x)dy_j \), where \( g_j(y_j|x) \) is the probability density function of \( Y_j \) given \( X = x \) and treatment \( j \). This quantity defines the probability of a failure on criterion \( Y_j \) after treatment \( j \). Van der Linden (1987) shows that the following conditions are sufficient for a monotone Bayes rule in a classification problem with threshold utility:

\[ w_{j-1} - v_{j-1} \leq w_j - v_j, \quad j = 2, \ldots, m \]  

(27)

\[ \Omega_j(d_j|x) \text{ is decreasing in } x, \quad j = 1, \ldots, m \]  

(28)

\[ \Omega_{j-1}(d_{j-1}|x) - \Omega_j(d_j|x) \text{ is increasing in } x, \quad j = 2, \ldots, m. \]  

(29)

The condition in equation (27) indicates that the relevant order of the treatments is with respect to \( w_j - v_j \). The other two conditions state that the probability of a failure should be decreasing in the test score, but that the difference between probabilities for two adjacent treatments should increase.

The optimal cutoff score for a decision between treatments \( j \) and \( j + 1 \) is found by comparing posterior expected utilities. If an individual with test score \( X = x \) is assigned to treatment \( j \), the posterior expected utility is equal to

\[ E_j(u_j(Y_j)|x) = \Omega_j(d_j|x)v_j + [1 - \Omega_j(d_j|x)]w_j, \]  

(30)

The interest is in the value of \( x \) at which \( E_j(u_j(Y_j)|x) \) and \( E_{j+1}(u_{j+1}(Y_{j+1})|x) \) cross, that is, the value of \( x \) for which

\[
\begin{align*}
\Omega_j(d_j|x)v_j + [1 - \Omega_j(d_j|x)]w_j \\
= \Omega_{j+1}(d_{j+1}|x)v_{j+1} + [1 - \Omega_{j+1}(d_{j+1}|x)]w_{j+1}
\end{align*}
\]

or

\[
(w_j - v_j)\Omega_j(d_j|x) - (w_{j+1} - v_{j+1})\Omega_{j+1}(d_{j+1}|x) \\
+ w_{j+1} - w_j = 0.
\]  

(31)
However, since $X$ is discrete, the optimal cutoff score is the smallest value of $x$ for which the left-hand side of equation (31) is positive. The conditions in equations (27) to (29) guarantee that this expression is an increasing function of $x$. If it takes the value of zero outside the range of possible test scores, the cutoff score is set at its corresponding border.

An interesting case arises if the utility parameters $w_j$ and $v_j$ do not vary across treatments. Then it follows that $w_j - v_j = w_{j+1} - v_{j+1}$ and $w_{j+1} - w_j = 0$, and the left-hand side of equation (31) reduces to

$$
\Omega_j(d_j|x) - \Omega_{j+1}(d_{j+1}|x).
$$

(32)

This expression reminds us of equation (21). Analogous to the argument following equation (21), it can be shown that for the choice of a normal distribution function for $\Omega_j(\cdot|x)$ together with linear regression functions $E_j(Y|x) = \beta_j x + \alpha_j$ and homoscedasticity, an optimal value of $x_j$ equal to the smallest value of $x$ larger than

$$
\frac{(d_{j+1} - \alpha_{j+1})[\text{Var}_j(Y,X)]^{1/2} - (d_j - \alpha_j)[\text{Var}_{j+1}(Y,X)]^{1/2}}{\beta_{j+1}[\text{Var}_j(Y,X)]^{1/2} - \beta_j[\text{Var}_{j+1}(Y,X)]^{1/2}}
$$

(33)

is obtained (van der Linden, 1987). This quantity can easily be calculated if estimates of the regression parameters and the pooled variances of $Y$ given $X = x$ are available.

**Multivariate Test Scores**

The use of a test battery instead of a single test was mentioned earlier as a possible refinement of test-based decision making. How Bayes rules can be derived in such cases will now be shown for the problem of placement decisions.

It is assumed that the preceding theory of placement decisions with threshold utility holds. However, in addition to test score $X$, a second test score $Z$ with information about the criterion variables $Y_j(j = 1, \ldots, m)$ is given. $Z$ is also assumed to be defined by number-right scoring and may take the values 0, 1, \ldots, $r$. Let $k_j(x, y_j, z)$ be the joint probability function of test scores $X$ and $Z$ and criterion score $Y_j$, while $p(z)$ is the probability function of $Z$. Furthermore, let $S_j$ denote the set of ordered pairs $(x, z)$ for which treatment $j$ is assigned. The expected utility when sampling from the population is defined as

$$
R(S_1, \ldots, S_m) = \sum_{j=1}^m \sum_{S_j} \int u_j(y_j) k_j(x, y_j, z) dy_j.
$$

(34)
It follows immediately that

$$R(S_1, \ldots, S_m) = \sum_{j=1}^{m} \sum_{s_j} p(z) \int u_j(y_j) q(x, y_j | z) dy_j,$$

(35)

where \(q(x, y_j | z)\) is the probability function of \((X, Y_j)\) given \(Z = z\).

Suppose now that equations (28) and (29) hold for each value of \(z\). Then, for each possible \(z\)-coordinate, equation (35) can be maximized following equations (30) and (31) with respect to the \(x\)-coordinates in \(S_j\). Thus, the optimal sets \(S_j\) are defined by a series of \((r + 1) (m - 1)\) cutoff scores, namely, \(m - 1\) scores for each of the \(r + 1\) possible values of \(z\).

By symmetry, it holds that the \(z\)-coordinates if the optimal sets \(S_j\) can also be defined by a series of cutoff scores \(z_1, \ldots, z_{m-1}\) of the failure probabilities \(\Omega_{2}(d_j | z)\) have the properties in equations (28) and (29) for each value of \(x\). In the testing literature, classification rules that are monotone in two distinct test scores are known as "conjunctive rules." For classification into a given treatment, such rules require that the examinee pass a certain cutoff score on \(x\) as well as on \(y\). In other words, it is impossible to compensate a failure on one test by a high score on the other.

**Combination of Basic Decisions**

As noted before, a well-known example of combination of basic decisions in education is an individualized instruction system. Figure 5–8 shows a simple system in which a selection decision is followed by a module with a mastery decision after which a placement decision guides the students through two possible sequences of mastery decisions. Real-life systems often have a more involved structure.

In systems of more than one decision point, it is possible to optimize decision rules simultaneously. In doing so, more efficient use of the data in

![Figure 5–8. Example of an Individualized Study System](image-url)
Figure 5–9. A System of One Placement and One Mastery Decision

the system can be made. Also, more realistic utility structures can be used.

In order to illustrate the procedure, the simple case of a system with one placement and one mastery decision (van der Linden & Vos, 1986; Vos & van der Linden, 1987) will be dealt with (see figure 5–9). The placement test score is denoted as $X(x = 0, 1, \ldots, m)$. the observed mastery test score as $Y(y = 0, 1, \ldots, n)$, and the true mastery test score as $T(t \in [0, 1])$.

Furthermore, the following definitions are made: $x_r$ is a cutoff score on $x$, $y_r$ is a cutoff score on $y$, $t_r$ is the mastery threshold on $t$, $j = 1, 2$ are the two possible treatments, and $i = 0, 1$ are the possible states of nonmastery ($T < t_r$) and mastery ($T \geq t_r$), respectively. Without bothering about conditions for optimality, a monotone rule with maximum utility is looked for; that is, from the pairs of cutoff scores $(x_r, y_r)$, the one with the largest Bayes utility will be chosen. The following threshold utility function is adopted:

$$u_j(t) = \begin{cases} 
\nu_{11} + w_j & \text{for } t < d, y < c \\
\nu_{10} + w_j & \text{for } t \geq d, y < c \\
\nu_{01} + w_j & \text{for } t < d, y \geq c \\
\nu_{00} + w_j & \text{for } t \geq d, y \geq c 
\end{cases} \quad (36)$$

The parameter $\nu$ is dependent both on the mastery state and the mastery decision, but it is independent of the treatment. In addition, the treatment-dependent parameter $w$ can be used to allow for differences in, for instance, treatment costs. Although this function could be made more realistic by replacing $\nu$ by continuous functions of the true score $t$, it nicely demonstrates how in a simultaneous approach a utility function defined on the ultimate criterion of the system (mastery-nonmastery state) can be brought into a previous decision (placement decision).
The Bayes utility for assignment to treatment 1 and a correct mastery decision is equal to

$$\sum_{x=0}^{x_c-1} \sum_{y=0}^{y_c-1} \int_0^{t_r} (v_{00} + v_{11}) \phi(x, y, t)dt,$$

(37)

where $\phi(x, y, t)$ is the probability function of $(X, Y, T)$. Combining the two possible treatments, two possible mastery decisions and two possible mastery states, the total expected utility when sampling from the population of individuals is equal to the sum of eight expressions like equation (37). From van der Linden and Vos (1986) it can be verified that this sum can be reduced to the following posterior form:

$$R(x_c, y_c) = \text{constant} + \sum_{y=y_c}^{y} \left[ (-v_{00} - v_{11}) \int_0^{t_r} \phi_0(t|y) + v_{11} \right] \kappa_0(y)$$

$$+ \sum_{x=x_c}^{x} \left\{ v_{00} \int_0^{t_r} (\rho_1(t|x) - \rho_0(t|x))dt + w_1 - w_0 \right\} \lambda(x)$$

$$+ \sum_{x=x_c}^{x} \sum_{y=y_c}^{y} \left\{ (v_{00} + v_{11}) \int_0^{t_r} \pi_0(t|x, y)dt - v_{11} \right\} \eta_0(x, y)$$

$$- \left[ (v_{00} + v_{11}) \int_0^{t_r} \pi_1(t|x, y)dt - v_{11} \right] \eta_1(x, y) \right\},$$

(38)

with $\phi_0(t|y)$, $\rho(t|x)$, $\pi(t|x, y)$ being the posterior probability density functions of $T$ given $Y = y$, $X = x$ and $(Y = y, X = x)$, respectively, and where $\eta_0(x, y)$, $\lambda(y)$, and $\kappa_0(y)$ are the marginal probability functions.

To obtain the optimal values of $x_c$ and $y_c$, equation (38) has to be evaluated numerically for all possible pairs of values, whereupon the pair for which equation (38) is maximal can be determined. This can be done as soon as accurate estimates for the three different posterior probabilities of a failure in equation (38) are available. To obtain these estimates, a test theory model is needed.

**Conclusion**

A review of the applications of Bayesian decision theory to test-based decision making was given. Four basic types of decisions were distinguished, each of which may be subject to further restrictions or generalizations. Decisions may also be made separately or in combination. The test theory literature contains many more results for selection and mastery decisions than were presented in this chapter. The placement and classification decisions as well as some of the possible generalizations and re-
strictions and the case of combinations of decisions have been largely unexplored. Of paramount importance, however, is research on realistic utility functions for test-based decisions. Such functions should not only yield robust decision rules (Vijn & Molenaar, 1981), but also be supported by empirical evidence of their appropriateness (Vrijhof, Mellenbergh, & van den Brink, 1983). Developments in this field will be decisive for the applicability of Bayesian theory to test-based decision making.

References


