

7.1.1.2 Decision Errors

There are two possible errors that can occur:

Type I error A Type I error occurs when the null hypothesis is true, but is rejected nonetheless. This is also known as an α error. For biometric verification, the result of a Type I error is falsely rejecting a genuine match, and is therefore known as a *false non-match*.²

Type II error A Type II error occurs when the null hypothesis is false, but is accepted. This is sometimes referred to as a β error. For biometric verification, the result of a Type II error is falsely accepting an impostor match as genuine, thus resulting in a *false match*.

The probabilities of false match and false non-match errors can be computed from the probability density functions for H_0 and H_1 . The probability of a false non-match is the false non-match rate (FNMR), which is the area under the H_0 distribution to the left of the match threshold. This area is indicated in Fig. 7.1 as the area shaded with horizontal lines. The probability of a false match is the false match rate (FMR), which is the area under the H_1 distribution, to the right of the match threshold. In Fig. 7.1, this is the area shaded with vertical lines.

7.1.2 Performance Rates

The previous section presented a method for making verification decisions (accept or reject) based probability distributions for the null hypothesis (i.e. the user is telling the truth) and the alternative hypothesis (i.e. the user is lying). The determination of these prior probabilities is now examined, and in particular how they are used to compute the standard performance measures.

7.1.2.1 Representing Match Score Distributions

Traditional hypothesis testing typically uses a family of test statistics whose distributions are known a priori (such as the t-test or z-test). However, in the case of biometric systems, the distributions of the genuine and impostor test statistics must be determined empirically. In other words, the distributions are estimated by computing the test statistic for a large sample of genuine and impostor matches.

² In statistics literature, a Type I error is often referred to as a *false positive*, since the alternative hypothesis is falsely accepted. This terminology may cause confusion in the case of biometrics, because the outcome of *falsely accepting* the alternative hypothesis (i.e. that the samples belong to different people) is *falsely rejecting* a genuine match. The confusion arises from the conflicting use of the term “acceptance”, which is applied to the claim of identity in one case, and the alternative hypothesis in the other. The convention used here is that “acceptance” refers to the claim of identity [15].

Collecting sample match scores is a vital step for system evaluation, and it is important that steps are taken to ensure robust data collection. The following are some important caveats for sample score collection (also see Chap. 5):

- The subjects (and their templates) used for generating the sample match scores should be representative of the target population.
- Scores should be sampled from the full range of possible values. For example, some biometric engines only output the top ranked scores by default in the matching logs. Without samples from the entire range of scores, it is not possible to calculate verification error rates.
- As many genuine and impostor samples as practical should be collected in order to maximize the precision of the estimates (see Sect. 7.3).
- It is very important *not* to assume a parametric form of the match score distribution, such as normality. Such assumptions are rarely justified, and may severely bias error rate computations.

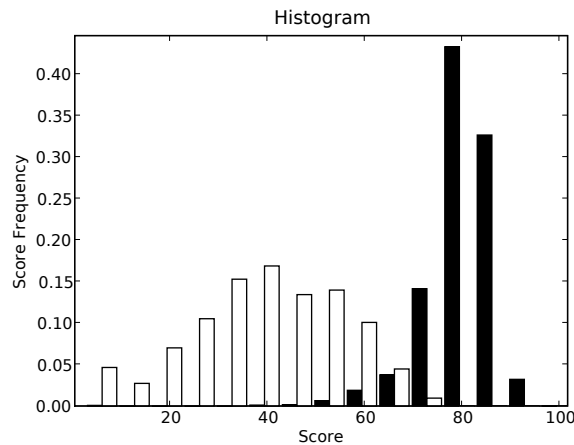


Fig. 7.2 A match score histogram. The black bars represent genuine scores, and the white bars are impostors.

After a sample of genuine and impostor scores have been collected, they are used to estimate the match score distributions. Note that a finite number of sample scores are used to estimate the true (unknown) distributions, which are continuous. *Histograms* tabulate the number of samples in disjoint ranges of scores (bins), and are commonly used in statistics to graphically represent discrete distributions. Figure 7.2 contains an example of a match score histogram.

The distance between, or overlap of, the genuine and impostor match score distributions has an important interpretation for biometric data (see Fig. 7.3). Generally, the greater the distance between these distributions, the lower the system's error rates. This can be demonstrated by considering an extreme example where there is

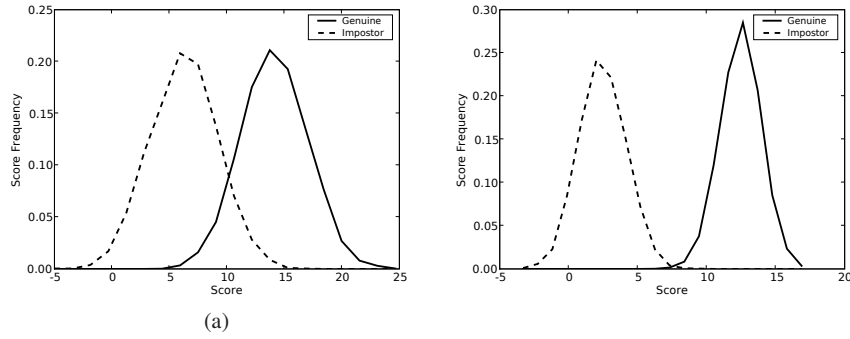


Fig. 7.3 The relationship between score distributions and accuracy. (a) The overlap between the genuine and impostor score distributions is large, therefore there is a range of scores (8-12) where both genuine and impostor scores are likely. This will lead to false match or false non-match errors, depending on the match threshold. (b) There is little overlap between the genuine and impostor score distributions. If the match threshold is set to 7.5, false match and false non-match errors will be rare.

no overlap between the two distribution. In this case, one can set the match threshold anywhere in the area between the distributions. At this threshold, all scores above the threshold are genuine, all below are impostors, and therefore no false match or false non-match errors will occur.

7.1.2.2 The False Match, False Non-Match and Equal Error Rates

Section 7.1.1.2 introduced the two primary error types for verification systems: the false match and the false non-match. The false match rate (FMR) is the proportion of impostor matches that obtain scores above the match threshold, and the false non-match rate (FNMR) is the proportion of genuine matches that obtain scores below the match threshold.³ The following is a procedure for computing these rates:

1. Select a match score threshold t
2. Sort the genuine (impostor) match score in increasing order
3. Divide the number of genuine (impostor) scores $< t$ ($\geq t$) by the total number of genuine (impostor) scores to find the false non-match (false match) rate

Note that the FMR and FNMR are functions of t . Denote $\text{FNMR}(t)$ and $\text{FMR}(t)$ to be the corresponding rates at the given threshold.

³ Recall that a false match is a Type II error. The *power* of a hypothesis test is the probability of rejecting a false null hypothesis. Intuitively, this is the probability of correctly non-matching an impostor (1-FMR). The Type I error (false non-match) rate of a system can be reduced to an arbitrarily small level by lowering the match threshold. However, for a system to have statistical power, it must still be able to reject impostors.

Generally, one would like to know how a system performs over a range of thresholds, so the above procedure is repeated many times for different values of t . The thresholds at which to compute the error rates can be chosen in a variety of ways. For example, they can be selected at equally spaced intervals throughout the range of possible match score values. Alternatively, one can compute the error rates for each unique score that has been observed.

The $FMR(t)$ and $FNMR(t)$ are plotted together in order to see their relation to each other. Figure 7.4 (b) contains these plots for the error rates corresponding to the score distributions of Fig. 7.4 (a). As the threshold value increases, the FMR varies from 100% (every impostor match is accepted) to 0% (every impostor match is rejected). Conversely, the FNMR varies from 0% (every genuine match is accepted) to 100% (every genuine match is rejected) as the threshold increases.

The point where the FMR curve and FNMR curve intersect is known as the *equal error rate* (EER). It is called this because at this threshold, the FMR and FNMR are equal. With respect to the score distributions, the EER occurs at the threshold t where the area under (i.e. the integration of) the genuine distribution $< t$ equals the area under the impostor distribution $\geq t$. Note that this is generally not the same as the threshold at which $p(s|Genuine)=p(s|Impostor)$ for a match score s (i.e. at the crossover of the probability density functions). This can be observed in Fig. 7.4 (a). The EER threshold is indicated by a vertical line, and does not correspond to the crossover of the probability density functions. However, the EER threshold does correspond to the crossover point of the FMR and FNMR curves in Fig. 7.4 (b).

Reporting the EER

The EER is often used as a simple way to summarize the performance of a biometric system. This due to its convenience as a single value with an intuitive interpretation. However, it should be kept in mind that there is no other special distinction of the value, and systems in production are rarely configured to operate at the EER. Furthermore, systems should not usually be compared based on this value alone. Simply because system A has a lower EER than system B, this does not necessarily imply that system A will outperform system B over all thresholds.

7.1.2.3 Selecting an EER Point *

The true (unknown) distributions for the genuine and impostor scores are continuous, yet are estimated using a finite number of samples. Therefore, the FMR and FNMR curves constructed are discrete. This raises difficulties for the computation of the EER, as an exact point of crossover may not exist (especially for small data sets). For example, in Fig. 7.4 (b) the FMR and FNMR curves appear smooth. However,