On the Behaviour of Information Measures for Test Selection

Danielle Sent[1] and Linda C. van der Gaag[2]

[1] Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
danielle.sent@utwente.nl
[2] Department of Information and Computing Sciences, Utrecht University, P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
linda@cs.uu.nl

Abstract. In diagnostic decision-support systems, a test-selection facility serves to select tests that are expected to yield the largest decrease in the uncertainty about a patient's diagnosis. For capturing diagnostic uncertainty, often an information measure is used. In this paper, we study the Shannon entropy, the Gini index, and the misclassification error for this purpose. We argue that for a large range of values, the first derivative of the Gini index can be regarded as an approximation of the first derivative of the Shannon entropy. We also argue that the differences between the derivative functions outside this range can explain different test sequences in practice. We further argue that the misclassification error is less suited for test-selection purposes as it is likely to show a tendency to select tests arbitrarily. Experimental results from using the measures with a real-life probabilistic network in oncology support our observations.

Keywords: Shannon entropy, Gini index, misclassification error, test selection.

1 Introduction

In many fields of medicine, physicians have to establish a diagnosis and have to decide upon an appropriate therapy in relative uncertainty about a patient's true condition. To assist physicians in their complex reasoning processes, sophisticated decision-support systems are being developed. Such a system is often equipped with a test-selection facility that serves to indicate which tests had best be performed to decrease the uncertainty about the patient's diagnosis [1,2]. The two most commonly used measures for capturing diagnostic uncertainty in decision-support systems, are the Shannon entropy and the Gini index [6]; in other contexts, also the misclassification error is used for measuring uncertainty [5]. The three measures are defined for a probability distribution over a designated diagnostic variable and express the expected amount of information that is required to establish the value of this variable with certainty.

The Shannon entropy and the Gini index are generally considered to behave alike for test-selection purposes, in particular for diagnostic variables with a small number of values [3]. In fact, common knowledge has it that the two measures are interchangeable in practice. In this paper, we compare the Shannon entropy, the Gini index and

R. Bellazzi, A. Abu-Hanna, and J. Hunter (Eds.): AIME 2007, LNAI 4594, pp. 316–325, 2007.

the misclassification error from a fundamental perspective. By studying the first derivatives of the three functions, we argue that for a large range of probability distributions over the main diagnostic variable, the Shannon entropy and the Gini index are indeed expected to behave alike. For the more extreme probability distributions, however, the two measures are expected to result in different test sequences. We further argue that the misclassification error is less suited for test-selection purposes as it is likely to show a tendency to select tests randomly.

We studied the Shannon entropy and the Gini index also from an experimental perspective. For this purpose, we implemented the two measures in a decision-support system for the domain of oesophageal cancer and performed test selection for 162 real patients. Upon analysing the sequences of tests yielded, we found that for 71% of the patients, already the first or second test selected differed between the two measures. In contrast with common knowledge, therefore, the Shannon entropy and the Gini index gave rise to quite different test-selection behaviour. All differences could be explained, however, from the insights that we had gained from our more fundamental analysis of the Shannon entropy, the Gini index, and their first derivatives.

The paper is organised as follows. Section 2 reviews the Shannon entropy, the Gini index, and the misclassification error, and details how these measures are used for test selection. Section 3 summarises our fundamental analysis of the three measures, and of their first derivatives more specifically. In Section 4 we report on the experimental results obtained with the Shannon entropy and the Gini index, and explain the observed differences. The paper ends with our conclusions in Section 5.

2 Information Measures and Test Selection

In a diagnostic decision-support system, test selection generally amounts to selecting tests that are expected to yield the largest decrease in the uncertainty about a patient's diagnosis. For capturing diagnostic uncertainty, typically an information measure is used. The three most commonly used measures are the Shannon entropy, the Gini index, and the misclassification error. These measures are defined for a probability distribution Pr over a set of stochastic variables. We distinguish a diagnostic variable D, modelling the diagnoses of interest; the possible values of D are denoted d_j, $j = 1,\ldots,m, m \geq 2$. We further distinguish $n \geq 2$ test variables T_i, modelling diagnostic tests whose results can influence the uncertainty in D; the results of a test T_i are denoted t_i^k, $k = 1,\ldots,m_i$, $m_i \geq 2$. Each of the three measures attains its maximum when the uncertainty about the value of the diagnostic variable is the largest, that is, when the probability distribution over this variable is a uniform distribution. For a distribution with $\Pr(d_i) = 1$ for some value d_i of D and $\Pr(d_j) = 0$ for all $d_j \neq d_i$, the uncertainty about the value of the diagnostic variable is resolved and the measures yield their minimum value of 0.

The *Shannon entropy* $H(\Pr(D))$ of the probability distribution Pr over the diagnostic variable D is the expected amount of information that is required to establish the value of D with certainty; more formally, the entropy is defined as

$$H(\Pr(D)) = -\sum_{j=1,\ldots,m} \Pr(D = d_j) \cdot {}^2\log \Pr(D = d_j)$$

where $0 \cdot {}^2\log 0$ is taken to be 0. Now suppose that some diagnostic test T_i is performed and that the result t_i^k is yielded. Because of this additional information, the probability distribution over D will change from the prior distribution to the posterior distribution given $T_i = t_i^k$. The entropy of the distribution over D will then change as well, to the entropy of the posterior distribution:

$$H(\Pr(D \mid T_i = t_i^k)) = -\sum_{j=1,\ldots,m} \Pr(D = d_j \mid T_i = t_i^k) \cdot {}^2\log \Pr(D = d_j \mid T_i = t_i^k)$$

Prior to performing the test T_i, however, we do not know for certain which result will be obtained: each possible result t_i^k is yielded with a probability $\Pr(T_i = t_i^k)$. Before actually performing the test, therefore, we expect the entropy of the posterior probability distribution over D to be

$$H(\Pr(D \mid T_i)) = \sum_{k=1,\ldots,m_i} H(\Pr(D \mid T_i = t_i^k)) \cdot \Pr(T_i = t_i^k)$$

We now have that the decrease in uncertainty in the diagnostic variable D by performing the test T_i is expected to be $\widetilde{H}(T_i) = H(\Pr(D)) - H(\Pr(D \mid T_i))$. A test that maximises $\widetilde{H}$ thus is the best test to perform. We assume that upon selecting a test that maximises the expected decrease in uncertainty, ties are broken at random.

The *Gini index* $G(\Pr(D))$ of the probability distribution Pr over the variable D is defined as

$$G(\Pr(D)) = 1 - \sum_{j=1,\ldots,m} \Pr(D = d_j)^2$$

The expected Gini index $G(\Pr(D \mid T_i))$ after performing a test T_i is defined as the expected value of the Gini index where the expectation is taken over all possible results:

$$G(\Pr(D \mid T_i)) = \sum_{k=1,\ldots,m_i} G(\Pr(D \mid T_i = t_i^k)) \cdot \Pr(T_i = t_i^k)$$

The best test to perform again is a test that is expected to result in the largest decrease in diagnostic uncertainty, that is, a test that maximises $\widetilde{G}(T_i) = G(\Pr(D)) - G(\Pr(D \mid T_i))$.

Occasionally also the misclassification error is used for capturing uncertainty [5]; in the sequel we will argue that this measure is less suited for the purpose of test selection, however. The *misclassification error* $M(\Pr(D))$ of the probability distribution Pr over the diagnostic variable D captures the difference between the probability of a certain diagnosis, that is, a probability equal to 1, and the probability of the most likely diagnosis; more formally, it is defined as

$$M(\Pr(D)) = 1 - \max\{\Pr(D = d_j) \mid j = 1, \ldots, m\}$$

The expected misclassification error after performing a diagnostic test T_i is

$$M(\Pr(D \mid T_i)) = \sum_{k=1,\ldots,m_i} M(\Pr(D \mid T_i = t_i^k)) \cdot \Pr(T_i = t_i^k)$$

The decrease in uncertainty in D by performing the test T_i thus is expected to be $\widetilde{M}(T_i) = M(\Pr(D)) - M(\Pr(D \mid T_i))$. A test that maximises $\widetilde{M}$ again is the best test to perform.

3 The Measures from a Fundamental Perspective

To provide for predicting the test-selection behaviour of the Shannon entropy, the Gini index and the misclassification error, we study the three measures from a fundamental perspective. Upon doing so, we focus on a binary diagnostic variable only; our considerations, however, also hold for non-binary variables. For a binary diagnostic variable D, with values d_1 and d_2, the Shannon entropy, the Gini index and the misclassification error can be written as

$$\begin{aligned} H(\Pr(D)) &= -\sum_{j=1,2} \Pr(D=d_j)\cdot {}^2\log \Pr(D=d_j) = \\ &= -x\cdot {}^2\log x - (1-x)\cdot {}^2\log(1-x) \end{aligned}$$

$$\begin{aligned} G(\Pr(D)) &= 1 - \sum_{j=1,2} \Pr(D=d_j)^2 = \\ &= 2x - 2x^2 \end{aligned}$$

$$\begin{aligned} M(\Pr(D)) &= 1 - \max\{\Pr(D=d_j \mid j=1,2\} = \\ &= \begin{cases} x & , \text{if } x \in [0,\frac{1}{2}] \\ 1-x & , \text{if } x \in \langle \frac{1}{2}, 1] \end{cases} \end{aligned}$$

where $x = \Pr(D=d_1)$; the functions are shown in Figure 1(a), (b) and (c) respectively.

From Figure 1(a) and (b), we observe that the Shannon entropy has a higher value than the Gini index. To formally support this observation, we consider the second derivatives of the two functions:

$$\begin{aligned} H''(x) &= -\frac{1}{x\cdot \ln 2} - \frac{1}{(1-x)\cdot \ln 2} \\ G''(x) &= -4 \end{aligned}$$

We observe that $H''(x) < G''(x)$ for all $0 < x < 1$. Since both measures attain their maximum at $x = \frac{1}{2}$, we thus have that in the interval $\langle 0, \frac{1}{2}\rangle$ the ascent of the Shannon entropy is steeper than that of the Gini index; in the interval $\langle \frac{1}{2}, 1\rangle$, the Shannon entropy shows a steeper descent than the Gini index. We further observe that the two functions attain the same value at $x = 0$ and at $x = 1$. We conclude that the two functions do not otherwise intersect and, hence, that $H(x) > G(x)$ for all $0 < x < 1$. We also compare the misclassification error and the Gini index. Within the interval $[0, \frac{1}{2}]$, we have that

$$G(x) = 2x - 2x^2 \geq x$$

since from $0 \leq x \leq \frac{1}{2}$ we can conclude that $2x^2 \leq x$. The Gini index, therefore, lies above the misclassification error. Within the interval $\langle \frac{1}{2}, 1]$, we find that

$$G(x) = 2x - 2x^2 = 2x\cdot(1-x) > 1-x$$

since from $\frac{1}{2} < x \leq 1$ we can conclude that $2x > 1$. Again, the Gini index lies above the misclassification error. We thus conclude that $G(x) \geq M(x)$. In fact, the misclassification error can be looked upon as a piece-wise linear interpolation of the three points $G(0), G(\frac{1}{2})$ and $G(1)$ of the Gini index.

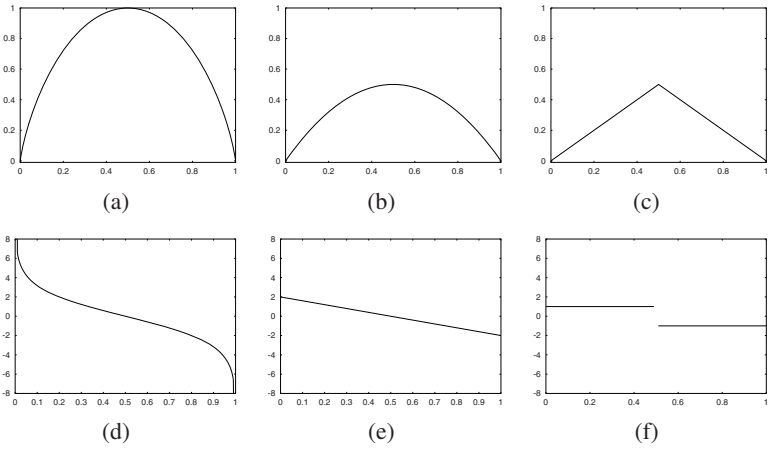

Fig. 1. The Shannon entropy (a), the Gini index (b), and the misclassification error (c) of a distribution over a binary variable, and their first derivatives (d), (e) and (f)

Now, for test-selection purposes, we are not so much interested in the precise values that the Shannon entropy, the Gini index and the misclassification error attain for a specific probability distribution over the diagnostic variable D. We are more interested in the way they value a *shift* in the distribution that is occasioned by a test result. We therefore also study the first derivatives of the three functions:

$$H'(x) = -^2\log x + ^2\log(1-x)$$

$$G'(x) = 2-4x$$

$$M'(x) = \begin{cases} 1 & , \text{if } x \in [0, \frac{1}{2}\rangle \\ -1 & , \text{if } x \in \langle \frac{1}{2}, 1] \end{cases}$$

These derivative functions are depicted in the Figures 1(d), (e) and (f) respectively.

The first derivative of the Gini index can be regarded as an approximation of the first derivative of the Shannon entropy for a large range of values of x. To support this observation, we consider the first three terms of the Taylor expansion of $H'(x)$ around $x = \frac{1}{2}$, divided by $G'(x)$. For the quotient, we find that

$$\frac{H'(x)}{G'(x)} = 1.44 + R$$

where the rest term R equals

$$R = 2.85 \cdot (x - \tfrac{1}{2})^2$$

We observe that the rest term is dependent upon the value of x at which we compare the two derivatives. Within the interval $[0.37, 0.63]$, for example, the rest term is smaller than 0.05. Within this interval, therefore, we have that the Taylor approximation of the first derivative of the Shannon entropy differs from the first derivative of the Gini index by a multiplicative factor only, with an error of at most 0.05. This finding is supported

by Figure 1(d), from which we observe that the first derivative of the Shannon entropy approximates the linear derivative function of the Gini index in the middle part of the $[0,1]$-interval. From the figure, we further observe that this property no longer holds for the more extreme values. From the rest term R, we find, for example, that for the value $x = 0.3$ the approximation error is less than 0.19 while for the value $x = 0.25$ it has grown to 0.49. For x approaching the extremes, therefore, the quotient $H'(x)/G'(x)$ grows excessively in favour of $H'(x)$.

To compare the first derivatives of the Gini index and of the misclassification error, we begin by observing that G' is a linear function and M' is a piecewise constant function. We further observe that $G'(\frac{1}{4}) = M'(\frac{1}{4})$ and $G'(\frac{3}{4}) = M'(\frac{3}{4})$. We conclude that the first derivative of the misclassification error is a two-point approximation of the first derivative of the Gini index. We now briefly address the suitability of the misclassification error for the purpose of test selection. We observe that within the interval $x \in [0, \frac{1}{2}]$, the misclassification error for the probability distribution over the diagnostic variable D equals $M(\Pr(D)) = \Pr(D = d_1) = x$. Now suppose that for a test variable T_i, we have that $\Pr(D = d_1 \mid T_i = t_i^k) \in [0, \frac{1}{2}]$ for all possible results t_i^k of T_i. We then find that the expected value of the misclassification error after performing the test equals

$$M(\Pr(D \mid T_i)) = \sum_{k=1,\ldots,m_i} \Pr(D = d_1 \mid T_i = t_i^k) \cdot \Pr(T_i = t_i^k) = \Pr(D = d_1) = x$$

The expected misclassification error $M(\Pr(D \mid T_i))$ of the posterior distribution thus equals the misclassification error $M(\Pr(D))$ of the prior distribution, and the expected decrease in uncertainty in D by performing the test T_i is $\widetilde{M}(T_i) = M(\Pr(D)) - M(\Pr(D \mid T_i)) = 0$. Similar observations hold for $\Pr(D = d_1) = x \in \langle \frac{1}{2}, 1]$. Only if the posterior probabilities of a diagnosis given the possible results t_i^k of T_i, are distributed over both intervals can the expected decrease in diagnostic uncertainty $\widetilde{M}(T_i)$ be larger than 0. Now, if at some stage in the test-selection process, for all remaining diagnostic tests the expected decrease in diagnostic uncertainty equals 0, the misclassification error will select a test at random. Since the probability distribution over the diagnostic variable is likely to become less uniform as the test-selection process progresses, the probability that a test will induce a shift to the other interval decreases. The misclassification error will then show a tendency to select tests rather arbitrarily; this tendency has been noted before by Breiman et al. [4]. We note that the tendency of the misclassification error to select tests at random may be quite undesirable for real-life decision-support systems.

We conclude by reviewing the implications of our findings for the test-selection behaviour of the Gini index and the Shannon entropy. The two measures value a test based upon the shifts that its results induce in the probability distribution over the diagnostic variable and upon the probabilities with which these results are expected to be found. Tests that induce a large shift in the probability distribution with a high probability, are valued as more informative than tests that result in a minimal shift with a high probability or in a large shift with just a small probability. Since the first derivative of the Gini index is a decreasing linear function, we find that it values a shift in distribution concavely by a constant factor. Since the first derivative of the Shannon entropy approximates a linear function within the interval $[0.37, 0.63]$, it values a shift in a distribution where x stays within this interval in the same way as the Gini index. We conclude that

the Shannon entropy and the Gini index will yield the same diagnostic test upon test selection as long as the tests under consideration are unlikely to result in a rather extreme distribution over the diagnostic variable. Since the Shannon entropy values a shift to an extreme distribution disproportionally more than the Gini index, the two measures may select different tests if a test is likely to result in such an extreme distribution. We note that several other researchers [4,6] also described this difference in behaviour between the Gini index and the Shannon entropy. Glasziou and Hilden for example argue that the Shannon entropy overestimates the gain in information for shifts in an already extreme probability distribution.

4 The Experimental Results

We formulated, in the previous section, the differences to be expected in the test-selection behaviour of the three measures. Based upon our findings, we concluded that the misclassification error is not as suitable for test selection as the other two measures. In this section we therefore focus on the Shannon entropy and the Gini index. To study the differences between the two measures in a practical setting, we conducted a test-selection experiment using the measures in the context of a real-life decision-support system in oncology. We briefly introduce the system that we used for our experiment before presenting the results that we obtained.

With the help of two experts in gastrointestinal oncology from the Netherlands Cancer Institute, we developed a decision-support system for the staging of cancer of the oesophagus [7]. The kernel of the system is a probabilistic network that models the various presentation characteristics of an oesophageal tumour and the pathological processes involved in its growth. The network currently includes 42 statistical variables, for which almost 1000 probabilities are specified. The main diagnostic variable of the network is the variable *Stage* that summarises the depth of invasion of the primary tumour and the extent of its metastasis; this variable has six possible values. The oesophageal cancer network further includes 25 variables to represent the results of diagnostic tests. For the staging of a patient's oesophageal cancer, typically a number of tests are performed. The various tests differ considerably in their reliability characteristics.

To study the behaviour of the Shannon entropy and the Gini index in the context of the oesophageal cancer network, we extended our decision-support system with a sequential test-selection facility. With the facility, we conducted two experiments. For the first experiment, we extended the oesophageal cancer network with a new binary variable *Operable* that summarises the six possible values of the original diagnostic variable *Stage* by classifying a patient's oesophageal cancer as operable or inoperable. For the second experiment, we used the test-selection facility with the original six-valued diagnostic variable. For our experiments, we had available the medical records of 162 patients diagnosed with cancer of the oesophagus. To simulate a realistic setting, we entered, for each patient, the results of a gastroscopic examination into the network prior to using the facility; in daily practice, the physicians also start selecting tests based upon the initial findings from this standard test.

As an illustration of the results that we found from our first experiment, we discuss the test-selection behaviour of the two measures for a specific patient. When the test

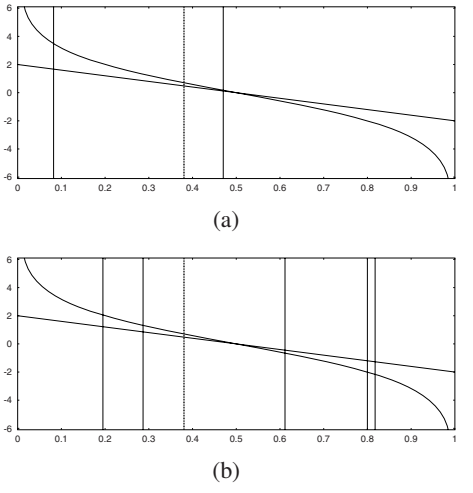

Fig. 2. The effects of the results of a CT-scan of the liver (a) and of an endosonography of the oesophageal wall (b), against the first derivatives of the Gini index and the Shannon entropy

selection is started, the probability of the cancer of this patient being operable, equals 0.38; the Gini index of the distribution over the variable *Operable* equals 0.471 and the Shannon entropy equals 0.958. For the next test to perform, the Gini index and the Shannon entropy suggest different tests. The Shannon entropy indicates that a CT-scan of the liver is expected to result in the largest decrease in diagnostic uncertainty, whereas the Gini index selects an endosonography of the oesophageal wall. More specifically, the expected Shannon entropy is computed to be 0.862 for the CT-scan and 0.899 for the endosonography; the expected values of the Gini index are 0.418 and 0.412 respectively.

To explain the observed difference in behaviour between the two measures, we study the shifts in the probability distribution over the diagnostic variable *Operable* that are occasioned by the various test results. Figure 2(a) shows, by means of vertical lines, the shifts in distribution that are yielded by the two possible results of a CT-scan of the liver; the shifts occasioned by the five different values of the endosonography of the oesophageal wall are shown in Figure 2(b). The prior probability of the patient's cancer being operable is indicated by a bold vertical line in both figures. From Figure 2(a), we observe that the leftmost vertical line, indicating the probability 0.082 of the patient's tumour being operable given that the result of the CT-scan of the liver is *yes*, is well within the range in which the first derivative of the Shannon entropy no longer approximates a linear function. The shift in the probability distribution over the variable *Operable* that is occasioned by this test result, therefore, is valued much higher by the Shannon entropy than by the Gini index. The result moreover is relatively likely to be found, with a probability of 0.231. The result *no* of the CT-scan is valued more or less the same by both measures. Two results of the endosonography, on the other hand, are valued more or less concavely by a constant factor by the Shannon entropy as well as by the Gini index. The other three results of the endosonography are not within the range where they are valued more or less the same by the two measures. These three

Table 1. The step, in the test-selection process, at which the Shannon entropy and the Gini index select different tests

Step	*Frequency*	*Step*	*Frequency*	*Step*	*Frequency*	*Step*	*Frequency*	*Step*	*Frequency*
1	24	5	4	9	1	13	0	none	3
2	90	6	6	10	1	14	0		
3	11	7	1	11	1	15	0		
4	19	8	1	12	0	16	3		

results have very low probabilities, of 0.034, 0.097 and 0.005, however. The result that serves to shift the probability of interest to 0.612, on the other hand, has a probability of 0.252, whereas the result that serves to yield a shift to 0.287 has a probability of 0.612. Note that although the probability 0.287 is not within $[0.37, 0.63]$, it is quite close to this interval. The shift to this probability is therefore valued more by the Shannon entropy than by the Gini index, yet not to a large extent. Since the shift occasioned by the endosonography is expected to result in a larger decrease of the uncertainty involved than that occasioned by the CT-scan of the liver, the Gini index selects the endosonography as the best test to perform. The expected decrease in diagnostic uncertainty by the CT-scan, however, is disproportionally larger with the Shannon entropy than with the Gini index, thereby explaining the Shannon entropy selecting the CT-scan. Note that these findings are conform the expectations from our fundamental analysis.

So far, we discussed in detail the differences in test-selection behaviour of the two measures under study for a binary diagnostic variable. We also studied the differences in behaviour for the original six-valued diagnostic variable *Stage*. Table 1 summarises, over all patients, the step in the test-selection process at which the Shannon entropy and the Gini index first selected a different diagnostic test. From the table we observe that for 24 patients (15%), already the first test differed. For 90 patients (56%), the measures selected the same diagnostic test for the first one to be performed, yet chose different tests for the second one. The range of tests selected in the first two steps was quite limited, however. The Shannon entropy selected the endosonography of the local region of the primary tumour for 44% of the patients as the most informative test, the endosonography of the oesophageal wall for 21% of the patients, and the CT-scan of the liver for 28% of the patients. The Gini index selected the endosonography of the local region of the primary tumour and of the oesophageal wall respectively, for 44% and 40% of the patients as the most informative test.

Since the Shannon entropy and the Gini index are commonly taken to be interchangeable for practical purposes, it is remarkable that for just three patients the two measures selected the same tests in the same order. The analysis from the previous section serves to explain why the two measures can select different tests. To explain the large number of differences found, we recall that, before the test-selection process is started for a patient, we entered the results from the gastroscopic examination into the network. Since these results tend not to influence the probability distribution over the diagnostic variable much, the test-selection process was started with a rather similar probability distribution for many patients. The example patient discussed in the previous section in fact belongs to this large group of similar patients.

5 Conclusions

In diagnostic decision-support systems, test selection amounts to selecting tests that are expected to yield the largest decrease in the uncertainty about a patient's diagnosis. For capturing this uncertainty, often an information measure is used. In this paper, we studied the Shannon entropy, the Gini index, and the misclassification error for this purpose. We argued that the first derivative of the Gini index can be regarded as an approximation of the first derivative of the Shannon entropy for a large range of values. We observed that, although a shift in many probability distributions over the diagnostic variable is valued similarly by the Gini index and the Shannon entropy, a shift to rather extreme distributions is valued much higher by the Shannon entropy than by the Gini index. Based upon this observation, the two measures are expected, at least occasionally, to select different tests. We feel that, despite their possible differences in behaviour, both measures are equally suited for use in a decision-support system. We furthermore concluded that the misclassification error should not be used for test-selection purposes due to its tendency to select tests randomly when all possible shifts in the probability distribution over the diagnostic variable are within the same half of the $[0, 1]$-interval.

We conducted an experiment to study the behaviour of the Shannon entropy and the Gini index in a real-life setting. The results from our experiment served to corroborate the differences in behaviour expected from our more fundamental analysis. A sensitivity analysis with respect to the selection of tests based upon the two information measures, moreover, showed that test selection based on the Shannon entropy and the Gini index is quite robust. Since our analysis of the two measures is independent of our application domain, we feel that the differences in test-selection behaviour observed in our experiments in the domain of oesophageal cancer, are likely to show in other domains as well.

References

1. Andreassen, S.: Planning of therapy and tests in causal probabilistic networks. Artificial Intelligence in Medicine 4, 227–241 (1992)
2. Ben-Bassat, M.: Myopic policies in sequential classification. IEEE Transactions on Computers 27(2), 170–174 (1978)
3. Breiman, L.: Technical note: some properties of splitting criteria. Machine Learning 24, 41–47 (1996)
4. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees. Wadworth & Brooks, Pacific Grove (1984)
5. Hastie, T., Tibshirani, R., Friedman, J.H.: The Elements of Statistical Learning. Springer, New York (2001)
6. Hilden, J., Glasziou, P.: Test selection measures. Medical Decision Making 9, 133–141 (1989)
7. Van der Gaag, L.C., Renooij, S., Witteman, C.L.M., Aleman, B.M.P., Taal, B.G.: Probabilities for a probabilistic network: A case-study in oesophageal cancer. Artificial Intelligence in Medicine 25(2), 123–148 (2002)