Applied Interdisciplinary Theory in Health Informatics
P. Scott et al. (Eds.)

doi:10.3233/SHTI190108

Information Theory and Medical Decision Making

Paul KRAUSE[a,1]
[a] *Department of Computer Science, University of Surrey, United Kingdom*

Abstract. Information theory has gained application in a wide range of disciplines, including statistical inference, natural language processing, cryptography and molecular biology. However, its usage is less pronounced in medical science. In this chapter, we illustrate a number of approaches that have been taken to applying concepts from information theory to enhance medical decision making. We start with an introduction to information theory itself, and the foundational concepts of information content and entropy. We then illustrate how relative entropy can be used to identify the most informative test at a particular stage in a diagnosis. In the case of a binary outcome from a test, Shannon entropy can be used to identify the range of values of test results over which that test provides useful information about the patient's state. This, of course, is not the only method that is available, but it can provide an easily interpretable visualization. The chapter then moves on to introduce the more advanced concepts of conditional entropy and mutual information and shows how these can be used to prioritise and identify redundancies in clinical tests. Finally, we discuss the experience gained so far and conclude that there is value in providing an informed foundation for the broad application of information theory to medical decision making.

Keywords. Shannon entropy; Relative entropy; Conditional entropy; Mutual information; Medical diagnosis

Learning objectives

After reading this chapter, the reader will be able to:

1. Understand the basic concepts of information theory: information content; Shannon entropy; relative entropy.
2. Understand how these concepts can be applied to medical decision making at a general level.
3. Understand how the more advanced concepts, conditional entropy and mutual information, could provide deeper insights into the potential redundancies in laboratory tests.

1. Introduction to Information Theory

Information theory has gained application in a wide range of disciplines, including statistical inference, natural language processing, cryptography and molecular biology. It covers the study of the transmission, processing, extraction, and utilization of information at a foundational, mathematical level. A fundamental goal of information

[1] Corresponding Author: Paul Krause; E-mail: p.krause@surrey.ac.uk

theory is to provide a sound basis for optimising the amount of information that can be extracted from a specific situation. Many of the outcomes from the study of information theory have been reduced to engineering practice in a wide range of disciplines, from Artificial Intelligence and Machine Learning, to cybernetics and complexity science. So, the question naturally arises: could it be used to inform the practice of medicine. This is the topic of the current chapter.

Central to Information Theory is the study of situations where one agent (the *transmitter*) conveys some message over a *channel* to another agent (the *receiver*). This is typically performed by having the transmitter send a series of partial messages. In the case of the Internet, for example, the Transmission Control Protocol (TCP) defines how a message may be broken down into packets before sending, enabling the resulting packets to be reassembled in the correct order by the receiver. Each of these partial messages can be thought of as resolving some measure of uncertainty in the receiver as to the content of the original message. The measure of uncertainty resolved by a partial message is its *information content*.

Let us start with a schematic of a general communication system, redrawn after Shannon's original paper [9].

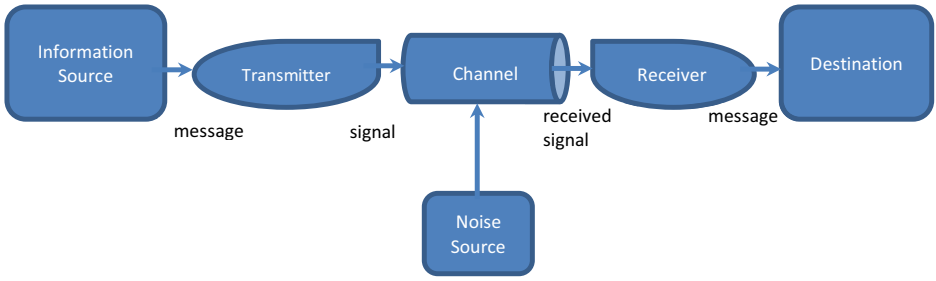

Figure 1. Schematic of a general information system.

We start with an *information source*, which generates a *message* or sequence of messages which are intended to be communicated to a *destination*. The destination is assumed to be remote from the information source. Hence, the message needs to be converted by a *transmitter* into a *signal* that is in a suitable form to be transmitted through some channel, after which the *received signal* is converted back into a suitable format by a *receiver* to enable it to be easily interpreted at the *destination*.

The challenge of communication theory is to understand how information that is transmitted from the source can be completely and correctly received and interpreted by the destination. This is a challenge because in general any communication channel will have an associated *noise source* that may corrupt, to a greater or lesser extent, the transmitted signal before it is received (by altering or even losing components of the signal). Furthermore, we cannot be certain that the transmitter and the receiver are perfect converters of message to signal, and signal to message, respectively.

In this chapter, we will show how viewing diagnosis as embedded within a communication system can lead to an information theoretic perspective on medical diagnosis. Each test or intervention can be seen as a partial message leading towards the desired complete message that provides sufficient information to confirm a diagnosis. At any stage in an investigation, one would then select the next test as the one that would

maximise the information gained. The important point about this perspective is that it provides a rational basis for identifying which test to perform at each stage of an engagement with a patient.

It is quite straightforward to map the model in Figure 1 onto the specifics of telegraphy, radio or television broadcasting, and the internet, for example. However, it will be of more interest in the current context if we instantiate the model within a clinical setting.

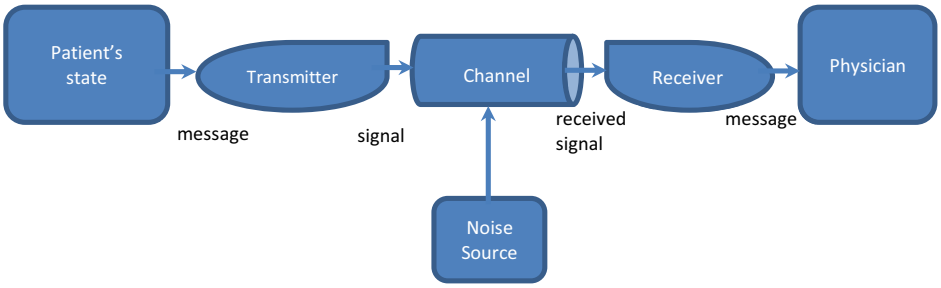

Figure 2. The communication system in a clinical setting.

We have actually changed very little in Figure 2 compared with Figure 1: the information source has become the patient's state; the destination has become the physician who has responsibility for performing the diagnosis of the patient's state.

We do still need to keep the model quite general. The transmitter might be, for example:

- The patient themselves, in the context of a consultation;
- A measurement on the patient;
- A trained physician or nurse examining the patient;
- The result of a test performed on the patient.

Do note the distinction between a message from the patient's condition that is some deviation from what is normal, and the signal that may be transmitted by the patient themselves or by a physician or nurse examining the patient. We cannot guarantee that the signal is an accurate representation of the original message (or that the transmitter has not missed a message, or even invented a message).

The channel might be, for example:

- A verbal utterance if transmitter and receiver are in the same room;
- A telephone line;
- An internet link;
- A written communication.

The task of the receiver is to transcribe the signal into an electronic or written record. Finally, the physician is the end point for a sequence of messages that will progressively inform a medical diagnosis.

There are two key points to keep in mind:

1. Any stage in the communication system may lead to loss or distortion of the information in a message as it is transmitted from its source to the physician;
2. Each message will contain a certain amount of *information* that will inform candidate diagnoses.

If we can maximise the amount of information in each message, then we should be able to minimise the number of messages needed in order to reach a confirmed diagnosis.

Performing such a minimisation, of course, presupposes that we do have available some measure of information content. This is the topic of the next section.

2. Information content and entropy

Before introducing a measure of information content, let us first explore a simple example to motivate the precise choice of measure. This is a necessarily brief introduction. A next step for the interested reader may be to read a more extended tutorial such as that provided in [11].

Consider an array of N binary switches, where N could be any integer greater than zero. For each switch, we have two possible states. One could think of these as "on" and "off". Correspondingly, we have two possible messages: one indicating the switch is in state "on" and one indicating the switch is in state "off".

With just one switch, we can store 1 "bit" of information: we just need to receive one message in order to determine the state of that switch. We can store 2 bits of information with an array of two switches, and we will require 2 messages each of 1 bit or one message of 2 bits to determine the state of that array.

Note that our measure of information content is additive; that is, the information content of a single message from an array of 2 binary switches is simply the sum of the information content of single messages sent from each of those switches separately.

In general, with an array of N switches, we will need a message that is N bits long in order to determine the internal state of that array of switches. Now, let us also look at the total number of states of an array of binary switches.

For two switches, we have four possible states: {on, on}; {on, off}; {off, on}; {off, off}. Correspondingly, we will have four possible messages that will tell us the state of the array in a single message.

In the general case of N switches in an array, we have 2^N possible states and 2^N possible messages.

Now, if we assume that each switch acts independently, and that each of these possible messages is equally likely, then any one message has a probability $p = 1/(2^N)$ of occurrence.

Consider an outcome in which a message m of length N is received. The above discussion motivates a requirement for a measure of information content that is additive. In addition, we have seen that the number of states in a system tends to increase exponentially. This suggests the use of a logarithmic function such as that of equation 1:

Eq 1. $$h(x) = -log_2 p(x)$$

Substituting our message m with probability of occurrence $p(m) = 1/(2^N)$ into Equation 1, we get:

$$h(m) = -log_2\left(1/_{2^N}\right) = log_2(2^N) = N$$

Equation 1 is thus returning us our informally proposed measure of information content; it is in fact the definition of the *Shannon information content of an outcome*.

We now need to generalise this. When a patient presents, that patient's state is not known with certainty. Thus, the possible messages that may be received form an ensemble M, with each message $m \in M$ having a probability of occurrence, $p(m)$. We use the word ensemble here in a statistical sense. Writing this out more formally, M is a

triple (m, A_M, P_M) where m is a "random variable" that can take on one of a number of possible values from an alphabet (a set of legal characters) $A_M = \{m_1, m_2, \ldots, m_K\}$ with respective probabilities $P_M = \{p_1, p_2, \ldots, p_K\}$. That is to say, the probability that $m = m_k$ for some $1 \leq k \leq I$ is p_k. We also require that $p_k \geq 0$ for all k, and $\sum_{k=1}^{I} p_k = 1$.

A measure $H(M)$ on the ensemble M can then be defined which is the average Shannon information content of an outcome:

Eq 2. $$H(M) \equiv -\sum_{k=1}^{K} p_k log_2 p_k$$

Strictly, this is simply providing us with the expected value of the information content in a message m that has been received from the ensemble M. However, the form of equation 2 is identical (apart from a constant) to the definition of entropy in the statistical mechanics model of thermodynamics:

$$S = -k_B \sum_i p_i \ln(p_i)$$

Here p_i represents the probability of a certain microstate of the thermodynamic system under consideration, and the sum is over all possible microstates. The natural logarithm is used in thermodynamics, but essentially the different base of the logarithm together with the use of Boltzmann's constant k_B simply provides a scaling between S and H.

By analogy with the form of this version of Boltzmann's equation, and the fact that the ensemble M in some sense represents the possible states of the system (a person in our case) under observation, $H(M)$ is referred to as the (Shannon) entropy of that ensemble. As with the Shannon information content, it also has the unit of bits (when using logarithm to the base 2).

Let us look at a couple of general-purpose examples to gain a little more intuition about how Equation 2 might be used before moving back to a diagnostic setting.

Consider an ensemble M in which an outcome is simply a character drawn at random from an English document. That is, the random variable m will be instantiated by selecting at random a character from an English document where $A_M = \{$a, b, c, d, e, …, x, y, z, _$\}$. We will not distinguish upper- and lower-case letters, but we do include the use of a space character, "_". $P_M = \{.0575, .0128, .0263, .0285, .0913, \ldots, .0007, .1928\}$[2] are the respective p_is for $1 \leq i \leq 27$.

Using the figures provided, it can be calculated that the outcome m = "z" has Shannon information content 10.4 bits, while the outcome m = "e" has information content of 3.5 bits. Overall, our English language document has an entropy of 4.1 bits. The full table of probabilities and corresponding measures of information content can be found in [7].

Let us examine this a little more. Providing a clear semantics to Shannon entropy is still a matter of debate (see, for example, p. 65 of [8]). Although it has the same form of thermodynamic entropy, it does not for example have the same units, as we have discussed; equation 2 has units of bits, whilst Boltzmann's entropy has units of Joules

[2] These values were estimated by the late David Mackay for use in his *Information Theory, Inference and Learning Algorithms* text book, Cambridge, 2003. His choice of text from which to estimate the probabilities, *The Frequently Asked Questions Manual for Linux*, of course means that these probabilities are conditional on the assumption that this text is representative of the distribution of letters in an English language document.

per Kelvin. However, although it would be wrong to say that Shannon entropy is "the same thing" as "entropy", it would be equally wrong to say they are unrelated: the two equations only differ by a constant (which defines the scale of measurement), and one can begin to reconcile the two if one relates the probabilities of the microstates of the system under consideration with the probabilities of the symbols generated by that system. Indeed, Jaynes argued in depth that the information theoretic view of entropy was a generalisation of thermodynamic entropy [3][4]. We implicitly advocate the same position in the context of medical diagnosis.

Going back to our document example. If we take a new document, pick a character at random and that character turns out to be a "z", a character with one of the lowest probabilities of occurrence in a typical English document, then that is providing us with more information about it (relative to a "normal" document) than if we had received an "e".

Two general properties are also worth noting. Firstly, if only one outcome in an ensemble M has a non-zero probability of occurring (in which case, its probability must be 1), then:

Property 1: $H(M) = 0$

(By convention, if $p(m_k) = 0$, then $0 \times log_2 0 \equiv 0$).

At the other end of the scale, the $H(M)$ is maximized if all of the outcomes are equally likely. An expression for the value for this is quite easy to derive. Let our ensemble M_e have K possible outcomes. Then we must have for all k, $p(m_k) = 1/K$. Substituting this into Equation 2, we get:

$$H(M_e) = -\sum_{k=1}^{K} \frac{1}{K} log_2 \frac{1}{K} = \frac{1}{K} log_2(K) \sum_{k=1}^{K} 1 = log_2(K)$$

(Noting that $log(1/K) = -log(K)$ and that $\log(K)$ is a constant and so can be factored to the outside of the summation). So,

Property 2: $H(M_e) = log_2(K)$ if all K outcomes are equally likely

In the case of our English document example, if the characters were uniformly distributed, then we would have $H(M_{uniform}) = \log_2(27) = 4.76$ bits. This is slightly higher than that for our representative English language document (4.1 bits).

Returning to the application of this to medical diagnosis, we can interpret these two situations as follows:

- $H(\ ) = 0$ if only one message/positive test result is possible. That is, a specific diagnosis has been confirmed.
- H is at its maximum when all messages are equally possible. That is, we are at a state of complete ignorance about the patient's internal state.

From this we can see that the challenge of diagnosis is to reduce the entropy to as close to zero as possible, and to select tests so that the result of each test (what we are calling "messages" here) maximises the reduction of entropy.

Two points should be emphasised here before we move on:

1. We are equating the probability of occurrence of messages with the probability of microstates of the patient under examination, to justify the usage of the term "entropy";

2. We are ignoring uncertainties in the veracity of a message that might be introduced by the communication pathway of Figure 2.

Point 2 is critically important when taking this into a real clinical setting. However, for simplicity of exposition we will continue to ignore this issue until the concluding section.

3. Relative entropy and diagnostic tests

Let us phrase the diagnostic strategy a little more formally. A patient's specific internal state has an ensemble of messages $M = (m, A_M, P_M)$ associated with it. A message will normally be triggered by a specific "interrogation" being performed on the patient. An interrogation may be, for example: a question asked of the patient; a test performed on the patient; an inspection performed by a nurse.

Prior to an interrogation the alphabet A_M of messages will have a probability distribution Q_M over it. Receipt of a message m_k (a positive test result, for example) will result in a posterior probability P_M over the alphabet of messages.

To measure the change in entropy, we use the *relative entropy*, or *Kullback-Leibler divergence*, between the two probability distributions [7]:

Eq 3. $$D_{KL}(P_M||Q_M) = \sum_{k=1}^{K} p_k log_2 \frac{p_k}{q_k}$$

It is worth noting two properties of relative entropy. Firstly, it satisfies what is known as Gibbs' inequality, with equality if and only if $P_M = Q_M$:

Eq 4. $$D_{KL}(P_M||Q_M) \geq 0$$

Secondly, in general it is not symmetric under interchange of the two probability distributions. That is, $D_{KL}(P_M||Q_M) \neq D_{KL}(Q_M||P_M)$. Consequently, relative entropy/Kullback-Leibler divergence does not formally qualify as a measure (hence the use of the term "divergence").

Expressed in terms of Bayesian inference, $D_{KL}(P||Q)$ is a measure of the information gained when a physician's beliefs are revised from a prior Q to a posterior P following some investigation.

We will use a hypothetical example adapted from [2] to illustrate the approach so far. We hypothesise a population of patients with arthritis, framed with a prior probability distribution over four possible syndromes. We have two diagnostic tests that are available to us, t_1 and t_2. Table 1 provides the pre-test probabilities and the respective post-test probabilities following a positive outcome from each of the two tests. Which of the two tests provides the greater information gain?

Table 1. Hypothetical Example (adapted from [2]).

Candidate Diagnosis	Pre-test Probability (t_0)	Post-test Probability (t_1)	Post-test Probability (t_2)
Gout	0.25	0.5	0.4
Osteoarthritis	0.5	0	0.1
Pseudogout	0.125	0.5	0.4
Other possibilities	0.125	0	0.1

Using Equation 3, it is straightforward to calculate that the information gain from t_1 is 1.5 bits, whereas the information gain had we chosen to perform t_2 would have been 0.68 bits (to 2 d.p.). Note again, that we make the assumption that the probabilities are continuous and so $0.log_2(0) = 0$. So, in the first case we have:

$D_{KL}(t_1||t_0) = 0.5 \times log_2(0.5/0.25) + 0.5 \times log_2(0.5/0.125) = 0.5 \times 1 + 0.5 \times 2 = 1.5$

In the second case we have:

$$D_{KL}(t_2||t_0) = 0.4 \times log_2(0.4/0.25) + 0.1 \times log_2(0.1/0.5) + 0.4 \times log_2(0.4/0.125) + 0.1 \times log_2(0.1/0.125)$$
$$= 0.4 \times 0.6781 + 0.1 \times (-2.322) + 0.4 \times 1.678 + 0.1 \times (-0.322)$$
$$= 0.68 \text{ (to 2 d.p.)}$$

The question naturally arises: why use relative entropy and not merely the difference of the pre-test and post-test entropies as measured using Equation 2. The latter was indeed proposed in early discussions on the use of entropy in medical decision making. However, Asch, Patton and Hershey concluded that it "fails to capture reasonable intuitions about the quantity of information provided by diagnostic tests" [1]. This point was reiterated in [2], which shows that relative entropy captures those intuitions more effectively. Kullback and Leibler [5], of course, provide a more formal justification of what we are calling relative entropy, as a sufficient statistic for discriminating between two probability distributions.

Let us now take a look at how these concepts from information theory might act as aids in medical decision making.

4. Shannon entropy and binary outcomes

Many laboratory tests are designed to assess the presence or absence of a disease state; a binary outcome. We can take a coin flip as a reference point, with the outcomes being heads or tails. Now, consider a collection of coins that are biased to some extent. That is, each coin will have a probability p that the outcome is a heads, with p varying over the collection between 0 and 1.

For a given coin C, from Equation 2 noting that the probability of a tails will then be $(1 - p)$, entropy is:

Eq 5. $$H(C) = -p \times log_2(p) - (1 - p)log_2(1 - p)$$

We can see that the entropy varies between 0 and 1, with a maximum at 1 when $p = 0.5$ (see figure 3).

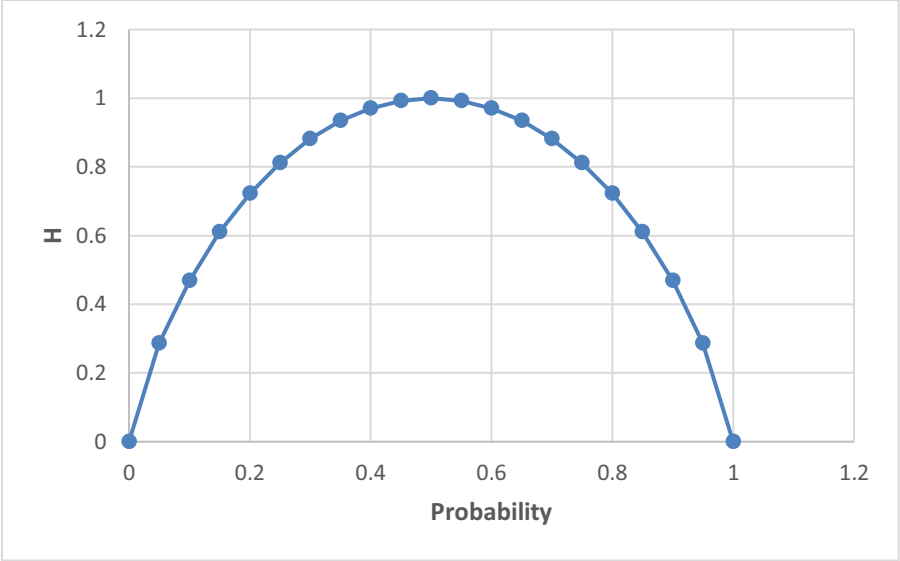

Figure 3. Variation of Entropy vs probability for a biased coin.

Think of the coin flip as a test on the internal state of a patient. A "heads" says the patient may have the disease, a "tails" says the disease is not present. If the coin is unbiased then as a test it is not helping us; all internal states are equally possible. We need a test where the entropy is close to 0 or 1 in order for us to be able to gain anything informative about the internal state of the patient.

Vollmer [13] explored the use of entropy to analyse the information content of a number of laboratory tests. He demonstrated how the concepts from information theory can be used as an aid to evaluating and understanding laboratory test results. We will use just one example to illustrate the point by using Figure 3 as a reference point.

Stadelmann et al [10] reported that the probability of 10-year mortality for malignant melanoma could be estimated from tumour thickness t using the following formula:

$$p = 1 - .966 \times e^{-(0.2016t)}$$

In Figure 4, we plot H vs tumour thickness t, using Equation 5.

We can see that over quite a wide range of values, with median $\approx 3.5mm$, tumour thickness provides limited information about the outcome.

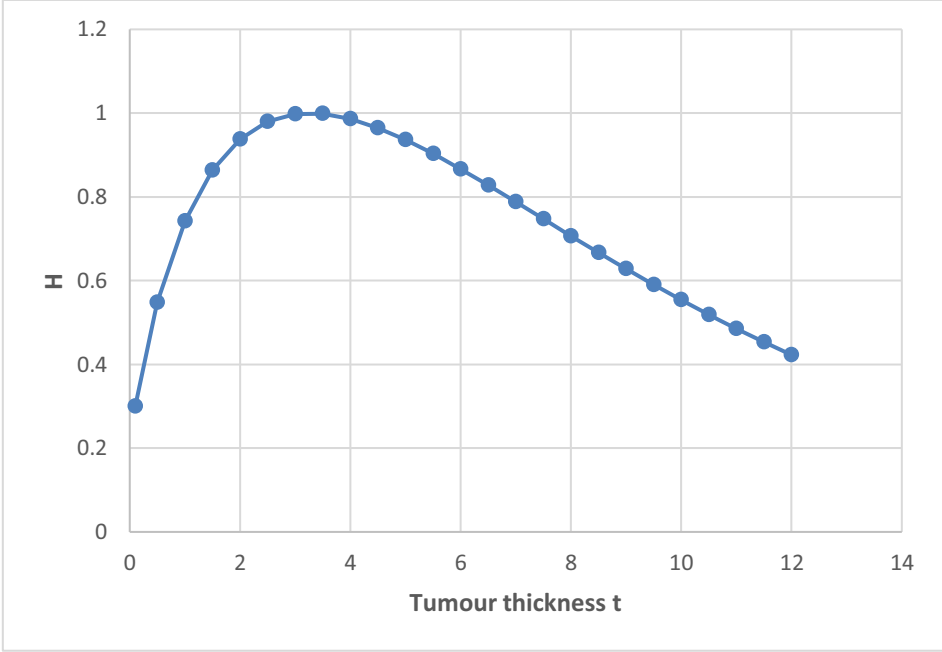

Figure 4. Entropy as a function of tumour thickness *t* (in mm) as a test for 10-year fatality from malignant melanoma.

5. Using information theory to prioritise laboratory tests

We are beginning to build up a number of approaches to using techniques for assisting with the choice of diagnostic tests in a clinical setting. First, we used relative entropy to discriminate between two candidate tests. Then we looked at the variation of entropy with the outcome of some diagnostic test (for a binary outcome) to identify the range of outcomes for which the test is informative. It should be emphasised that these are not the only tools available, but published work to date has argued for their value as additional tools that may aid in decision making.

Lee and Maslove [6] made the case that an information theoretic approach could have particular value in the case of identifying redundancy in tests in an intensive care unit (ICU). Parsimony in the case of blood tests is particularly important in the ICU context as repeated bloodwork can:

Cause anaemia and increase the need for blood transfusions;
Cause patient discomfort;
Disrupt sleep;
Lead to delirium.

The challenge, then, is to identify which blood tests are the most informative at a system wide level. A key issue here is that there may be some level of *redundancy* between laboratory tests; that is, especially over time, some tests may add little information over previously conducted tests. In cases where we can identify that there is a high degree of mutual information between tests (either through the same test being repeated too frequently, of for two different tests having too strong a dependency on

common information), then we have an objective basis for reducing the number of tests performed.

Two additional concepts were used in Lee and Maslove [6]. The first of these was the *conditional entropy* of X given Y. This measures the average uncertainty about a random variable x that remains when y is known (x and y being the respective random variables for the ensembles X and Y). It is defined as:

Eq 6. $$H(X|Y) = \sum_{xy \in A_X A_Y} p(x,y) log \frac{1}{p(x|y)}$$

Referring back to the definition of an ensemble, A_X is the alphabet of the ensemble X; that is, the set of legal values of the random variable x. Similarly, A_Y is the alphabet of the ensemble Y.

A related concept is the *mutual information* between X and Y. This measures the amount of information that x conveys about y, and is defined as:

Eq 7. $$I(X;Y) \equiv H(X) - H(X|Y)$$

Note that we have followed the definitions as given in Mackay [7].

Lee and Maslove extracted laboratory test results from MIMIC II, a fully anonymised public database. They analysed a total of 29,149 ICU admissions, investigating the following laboratory tests: haematocrit; platelet count; white blood cell count (WBC); glucose; HCO_3; potassium, sodium; chloride; BUN (Blood Urea Nitrogen); creatinine; and, lactate. Overall, their findings strongly supported the view that a significant amount of the bloodwork performed in ICUs is unnecessary. This had previously been discussed in [14], but Lee and Maslove were able to quantify the level of redundant information content. As a specific example, they found a high level of redundancy in information between the tests for BUN and creatinine; suggesting that if one is known, the other can be inferred with reasonable confidence. Furthermore, their analysis indicated that given the choice, it would be better to prefer BUN over creatinine.

Of course, clinical judgement will always be needed but this information theoretic approach does provide an objective foundation to an informed choice.

6. Discussion

We have shown in this chapter that information theory can have value in informing medical decision making. We have drawn on a number of studies in order to illustrate this. However, there is one area where we do beg to differ with most of those studies. Many of them bring in additional terminology to try and provide an intuitive semantics to some of the concepts in information theory; notions of “surprise”, “closeness to certainty”, perhaps a tendency to try and equate entropy to uncertainty. One can understand this. Within classical thermodynamics, entropy is perhaps one of the hardest concepts to gain a feeling for. However, we have been careful to refer only to measures of information and entropy. We have briefly alluded to an equivalence between Shannon entropy and entropy from statistical mechanics through associating the messages that can be potentially received from a patient with the internal microstates of that patient.

The paper by Tribus and McIrvine [12] is a good starting point for a deeper study of the conceptualisation of Shannon entropy, and could perhaps be read after the tutorial [11] that was mentioned earlier and the papers by Jaynes [3][4].

We do believe that the time is ripe to propagate a deeper understanding of information theory through practitioners of health informatics. The potential for significant enhancements in the rigour of medical decision making is waiting to be realised. However, some stronger guidelines do need to be developed for its usage. We have described a number of different strategies and there would be real value in documenting a common foundation that would inform all of these. In addition, we would emphasise the need to explicitly model the "noise" that is inherent in the communication model. We have included this in the communication model at the beginning of the chapter, and it is an important but often neglected factor in the risk of misdiagnosis of a patient.

Teaching questions for reflection

1. Can you think of a clinical setting from your own experience where information theory might have usefully informed your choices?
2. Do any of the examples provided in this chapter have more mainstream statistical methods of achieving the same result?
3. What do you feel are the barriers to the adoption of information theoretic approaches in the wider community?

References

[1] D.A. Asch, J.P. Patton and J.C. Hershey, Prognostic information versus accuracy: once more with meaning, *Medical Decision Making* **11** (1991), 45-47.
[2] W.A.Benish, Relative Entropy as a Measure of Diagnostic Information, *Medical Decision Making* **19** (1999), 202-206.
[3] E.T. Jaynes, Information Theory and Statistical Mechanics, *Physical Review* **106** (1957), 620-630.
[4] E.T. Jaynes, Information Theory and Statistical Mechanics, *Physical Review* **108** (1957), 171-190.
[5] S. Kullback and R.A. Leibler, On information and sufficiency, *Ann Math Stat* **11** (1951), 79-86.
[6] J. Lee and D.M. Maslove, Using information theory to identify redundancy in common laboratory tests in the intensive care unit, *BMC Medical Informatics and Decision Making* **15:59** (2015), 1-8.
[7] D.J.C. Mackay, *Information Theory, Inference and Learning Algorithms*, CUP, Cambridge UK, 2003.
[8] L. Pismen, *The Swings of Science – From Complexity to Simplicity and Back*, Springer Nature, Switzerland, 2018.
[9] C.E. Shannon, A Mathematical Theory of Communication, *The Bell System Technical Journal* **27** (1948), 379-423.
[10] W.K. Stadelmann, D.P. Rapaport, S-J. Soong, Prognostic factors that influence melanoma outcome. In: Balch CM, Houghton AN, Sober AJ, et al, eds. *Cutaneous Melanoma.* 3rd ed. St Louis, MO: Quality Medical Publishing; 1998:11-35.
[11] J.V. Stone, *Information Theory – A Tutorial Introduction*, Sebtel Press, 2015.
[12] M. Tribus and E.C. McIrvine, Energy and Information, *Scientific American* **225** (1971), 179-190.
[13] R.T. Vollmer, Entropy and Information Content of Laboratory Test Results, *Am J Clin Pathol* **127** (2007) 60-65.
[14] C. van Walraven and C.D. Naylor, Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits, *JAMA* **280** (1998), 550-558.