Information Graphs for Epidemiological Applications of the Kullback-Leibler Divergence

Dear Editor,

The topic addressed by this brief communication is the quantification of diagnostic information and, in particular, a method of illustrating graphically the quantification of diagnostic information for binary tests. To begin, consider the following outline procedure for development of such a test. An appropriate indicator variable that will serve as a proxy for the actual variable of interest, often the disease status of a subject, is identified. Two mutually exclusive groups are formed, one of definitively diseased subjects; the other of definitively non-diseased subjects. The classification of subjects into diseased and non-diseased groups is made by a gold standard method, independent of the putative indicator variable.

The value of this indicator variable is recorded for all subjects in both groups, leading to separate distributions of indicator scores for the diseased and non-diseased groups. The indicator variable is usually calibrated in such a way that on average, diseased subjects tend to have larger indicator scores than non-diseased subjects. Typically, we find that the two distributions of indicator scores overlap and, in such cases, any choice of a particular threshold indicator score will result in imperfect discrimination between the diseased and non-diseased groups on the basis of the indicator variable. Within this setting, we will consider two information theoretic analyses of diagnostic information [1, 2].

The Kullback-Leibler divergence, also referred to as the Kullback-Leibler distance and denoted as the Kullback-Leibler distance and of non-diseased subjects. Lee described the Kullback-Leibler distance as “an abstract concept arising from statistics and information theory” [1]. Here, restricting our attention to binary diagnostic tests, we construct a diagrammatic interpretation of the Kullback-Leibler distance as used in Lee’s application. We refer to this as an ‘information graph’, following Benish [4]. Information graphs provide a visual basis for the evaluation and comparison of binary diagnostic tests. This communication is motivated by the hope that a diagrammatic interpretation of the Kullback-Leibler distance will make it seem less of an abstract concept, and so make its application more accessible to epidemiologists and diagnosticians.

Some analysis and notation is needed to describe the basis for our information graph and its correspondence with Lee’s analysis, although once that is done with, the resulting graph is straightforward to construct. As far as possible we will use Lee’s original notation. No re-interpretation of Lee’s analysis is required; the objective is solely to provide a new diagrammatic format to present such analysis. For a binary diagnostic test, the test outcomes of diseased and of non-diseased subjects are Bernoulli distributed. Generically, the Bernoulli distribution is a discrete probability distribution with two possible outcomes denoted \( x \in \{0, 1\} \), in which \( x = 1 \) occurs with probability \( \theta_1 \) and \( x = 0 \) occurs with probability \( \theta_2 = 1 - \theta_1 \). We write: \( \text{Pr}(x) = \theta_1^x \theta_2^{1-x}; \) \( x \in \{0, 1\}; \theta_1 + \theta_2 = 1; 0 < \theta_1 < 1. \) Then the Kullback-Leibler distances between two Bernoulli distributions \( f(x) = \theta_1^x \theta_2^{1-x} \) (describing test outcomes of diseased subjects) and \( g(x) = \theta_2^x \theta_1^{1-x} \) (describing test outcomes of non-diseased subjects) as calculated by Lee are:

\[
D(f \parallel g) = f_1 \ln \left( \frac{f_1}{g_1} \right) + f_2 \ln \left( \frac{f_2}{g_2} \right)
\]

when \( g(x) \) is the reference distribution and \( f(x) \) is the comparison distribution. When

\[
D(g \parallel f) = g_1 \ln \left( \frac{g_1}{f_1} \right) + g_2 \ln \left( \frac{g_2}{f_2} \right)
\]

when \( f(x) \) is the reference distribution and \( g(x) \) is the comparison distribution [5]. The base of the logarithm can be chosen to suit the application. Lee works in natural logarithms, in which case the information quantities calculated have units of nits [6]. \( D(f \parallel g) \) and \( D(g \parallel f) \) are not necessarily equal, a property which Lee exploits in developing his application [1]. In the usual terminology associated with binary diagnostic tests, we identify \( f_1 \) as ‘sensitivity’ (denoted \( Se \)) with \( f_2 \) as its complement; similarly, \( g_2 \) is ‘specificity’ (\( Sp \)) with \( g_1 \) its complement. Kullback-Leibler distances are always \( \geq 0 \), and only \( = 0 \) when the two distributions are identical (the latter not being a scenario of interest in the context of Lee’s analysis).

Construction of the information graph requires representation of the Kullback-Leibler distance as a Bregman divergence [7]. Bregman divergences are properties of convex functions and in particular, the Kullback-Leibler distance is the Bregman divergence associated with:

\[
f(p) = p \cdot \ln(p) + (1 - p) \cdot \ln(1 - p)
\]

which is the negative of the Shannon entropy function [8]. Generically, a tangent to the curve \( f(p) \), gradient \( f'(p) \), is drawn at point \( p_1 \) (the reference point); then we calculate the Bregman divergence between the tangent and the curve at point \( p_2 \). For scalar arguments, this is:

\[
B(p_2\|p_1) = f(p_2) - f(p_1) - (p_2 - p_1) \cdot f'(p_1)
\]

and for the particular \( f(p) \) that is the negative Shannon entropy, this is a Kullback-Leibler divergence.

For example, Lee describes a binary diagnostic test for which \( Se = 0.8, Sp = 0.9 \). Then \( D(f \parallel g) = 1.36 \text{nits and } D(g \parallel f) = 1.15 \text{nits} \) (using natural logarithms and calculating to 2 dp) [1]. First consider the tangent to the curve \( f(p) \) at \( p_1 = Se \) (the reference point); this has slope = 1.386 and intercept = -1.609. Then \( B(p_2\|p_1) \) with \( p_2 = 1- Sp \) is:

\[
f(0.1) - f(0.8) = -0.1 \cdot 1.386 = -1.386 \text{nits}
\]

The distance between the curve \( f(p) \) and the tangent to \( f(p) \) drawn at \( p_1 = Se \), calculated at \( p_2 = 1 - Sp \), is equal to \( D(g \parallel f) \). Next consider the tangent to \( f(p) \) at \( p_1 = 1 - Sp \) (\( p_1 \) is again the reference point); this has slope = -2.197 and intercept = -0.105. Then \( B(p_2\|p_1) \) with \( p_2 = Se \) is:

\[
f(0.8) - f(0.1) = -0.8 \cdot (-2.197) = 1.36 \text{nits}
\]

This distance between the curve

---

**Methods Inf Med 2013; 52: IV–VI**

**Correspondence to:**  
Dr. Gareth Hughes  
SRUC  
The King's Buildings  
West Mains Road  
Edinburgh EH9 3JG  
UK  
E-mail: gareth.hughes@sruc.ac.uk

© Schattauer 2014
Methods Inf Med 1/2014 © Schattauer 2014

\( f(p) \) and the tangent to \( f(p) \) drawn at \( p_1 = 1 - Sp \), calculated at \( p_2 = Se \), is equal to \( D(f||g) \).

The information graph resulting from this analysis is shown in Figure 1. The Kullback-Leibler distances are represented by vertical lines between the negative Shannon entropy curve and the two tangents to the curve, one at \( Se \) and the other at \( 1 - Sp \). Note that the convexity of \( f(p) \) means that a tangent touches the curve at the reference point and otherwise lies below it, guaranteeing \( B(p_2||p_1) \geq 0 \) and so the non-negativity of the Kullback-Leibler distance. Particularly noteworthy is the way that the information graph gives us a simple visual interpretation of the asymmetry of the distances \( D(g||f) \) and \( D(f||g) \).

Appearing almost simultaneously with Lee’s epidemiological application of Kullback-Leibler distance was another, by Benish [2]. In order to describe Benish’s application, we need a little more notation. Again restricting our attention to binary tests (and using Benish’s original notation as far as possible), we have the Bernoulli distribution \( a(x) = a_1^{-x}a_2^x \), describing the pre-test probabilities \( a_1 \) (actual status diseased) and \( a_2 \) (actual status non-diseased). The Bernoulli distribution \( b(x) = b_1^{-x}b_2^x \) describes the post-test probabilities \( b_1 \) (actual status diseased given a positive test outcome) and \( b_2 \) (actual status non-diseased given a positive test outcome). The Bernoulli distribution \( c(x) = c_1^{-x}c_2^x \) describes the post-test probabilities \( c_1 \) (actual status diseased given a negative test outcome) and \( c_2 \) (actual status non-diseased given a negative test outcome). Then for Benish’s application the Kullback-Leibler distance we have:

\[
D(b||a) = b_1 \cdot \ln \left( \frac{b_1}{a_1} \right) + b_2 \cdot \ln \left( \frac{b_2}{a_2} \right)
\]

and

\[
D(c||a) = c_1 \cdot \ln \left( \frac{c_1}{a_1} \right) + c_2 \cdot \ln \left( \frac{c_2}{a_2} \right)
\]

Benish worked in base 2 logarithms, in which case the information quantities calculated have units of \textit{bits}. Here we will continue to work in natural logarithms. To convert results from nits to bits, divide the quantity calculated in nits by \( \ln(2) \).

For example, Benish considers a binary diagnostic test (for mammography given pre-test probability of actual status diseased \( a_1 = 0.2 \), post-test probability of actual status diseased given a positive test outcome \( b_1 = 0.67 \), and post-test probability of actual status diseased given a negative test outcome \( c_1 = 0.05 \). The long-dashed line shows the negative Shannon entropy curve \( f(p) = p \cdot \ln(p) + (1-p) \cdot \ln(1-p) \). The solid line shows the tangent to the curve at \( a_1 = 0.2 \) (slope = -1.386, intercept = -0.223). The short-dashed lines show the corresponding Bregman divergences, respectively \( B(p_2||p_1) = 0.52 \) nits with \( p_1 = a_1, p_2 = b_1 \); and \( B(p_2||p_1) = 0.09 \) nits with \( p_1 = a_1, p_2 = c_1 \). These Bregman divergences are, respectively, equivalent to the Kullback-Leibler distances \( D(b||a) \) and \( D(c||a) \) as calculated in bits by Benish [2].

Figure 1 An information graph for an example of a binary diagnostic test with \( Se = 0.8 \) and \( Sp = 0.9 \). The long-dashed line shows the negative Shannon entropy curve \( f(p) = p \cdot \ln(p) + (1-p) \cdot \ln(1-p) \). The solid lines show the tangents to the curve at \( Se = 0.8 \) (slope = 1.386, intercept = -1.609) and at \( 1 - Sp = 0.1 \) (slope = -2.197, intercept = -0.105). The short-dashed lines show the corresponding Bregman divergences, respectively \( B(p_2||p_1) = 1.15 \) nits with \( p_2 = Se, p_2 = 1 - Sp \); and \( B(p_2||p_1) = 1.36 \) nits with \( p_2 = 1 - Sp, p_2 = Se \). These Bregman divergences are, respectively, the Kullback-Leibler distances \( D(g||f) \) and \( D(f||g) \) as calculated by Lee [1].

Figure 2 An information graph for an example of a binary diagnostic test with \( Se = 0.8 \) and \( Sp = 0.9 \), pre-test probability of actual status diseased \( a_1 = 0.2 \), post-test probability of actual status diseased given a positive test outcome \( b_1 = 0.67 \), and post-test probability of actual status diseased given a negative test outcome \( c_1 = 0.05 \). The long-dashed line shows the negative Shannon entropy curve \( f(p) = p \cdot \ln(p) + (1-p) \cdot \ln(1-p) \). The solid line shows the tangent to the curve at \( a_1 = 0.2 \) (slope = -1.386, intercept = -0.223). The short-dashed lines show the corresponding Bregman divergences, respectively \( B(p_2||p_1) = 0.52 \) nits with \( p_1 = a_1, p_2 = b_1 \); and \( B(p_2||p_1) = 0.09 \) nits with \( p_1 = a_1, p_2 = c_1 \). These Bregman divergences are, respectively, equivalent to the Kullback-Leibler distances \( D(b||a) \) and \( D(c||a) \) as calculated in bits by Benish [2].
palpable breast mass) with \( Se = 0.8, Sp = 0.9 \). We have \( a_1 = 0.2, a_2 = 0.8; b_1 = 0.67, b_2 = 0.33; \) and \( c_1 = 0.05, c_2 = 0.95 \) [2]. Then 
\[ D(b \| a) = 0.52 \text{ nits and } D(c \| a) = 0.09 \text{ nits.} \]
Now, consider the tangent to the curve \( f(p) \) at 
\( p_1 = a_1 \) (the reference point); this has slope \( -1.386 \) and intercept \( -0.223 \). Then the Bregman divergence 
\[ B(p_2 \| p_1) \text{ with } p_2 = b_1, \]
is: 
\[ f(0.67) - f(0.20) - (0.67 - 0.20) \cdot (-1.386) = 0.52 \text{ nits.} \]
The distance between the curve \( f(p) \) and the tangent to \( f(p) \) drawn at \( p_1 = a_1 \), calculated at \( p_2 = a_1 \), is equal to \( D(b \| a) \). The Bregman divergence 
\[ B(p_2 \| p_1) \text{ with } p_2 = c_1 \]
is: 
\[ f(0.05) - f(0.20) - (0.05 - 0.20) \cdot (-1.386) = 0.09 \text{ nits.} \]
The distance between the curve \( f(p) \) and the tangent to \( f(p) \) drawn at \( p_1 = a_1 \), calculated at \( p_2 = c_1 \), is equal to \( D(c \| a) \). The information graph resulting from this analysis is shown in Figure 2. The Kullback-Leibler distances are represented by vertical lines (at post-test probabilities \( b_1 \) and \( c_1 \)) between the negative Shannon entropy curve and the tangent to the curve at the pre-test probability \( a_1 \).

Particularly noteworthy is the way that the information graph gives us a simple visual interpretation both of the post-test probability revisions and of the corresponding information properties of the test outcomes. No re-interpretation of Benish’s analysis is required; the objective here is to show how the new diagrammatic format we have described for binary diagnostic tests accommodates both Benish’s and Lee’s epidemiological applications of Kullback-Leibler distance.

As discussed by Lee, Kullback-Leibler divergence is a measure of the distance between the distributions of test outcomes of diseased and of non-diseased subjects [1]. Suppose we ask of a binary diagnostic test; how much does the test tell us, in information units, on average over both positive and negative test outcomes, about subjects who are actually diseased? The answer is provided by \( D(f \| g) \). And if we ask, of the same test; how much does the test tell us, in information units, on average over both positive and negative test outcomes, about subjects who are actually non-diseased? In this case the answer is provided by \( D(g \| f) \) [5]. In epidemiological terminology, \( D(f \| g) \) and \( D(g \| f) \) are average log-likelihood ratios, characteristic of a diagnostic test. The average log-likelihood ratio \( D(f \| g) \) describes a diagnostic test in terms of the expected weight of evidence available for discrimination in favour of diseased against non-diseased subjects. In contrast, \( D(g \| f) \) describes the expected weight of evidence available for discrimination in favour of non-diseased against diseased subjects. As discussed by Benish, Kullback-Leibler divergence is a measure of the distance between the distributions of pre-test probability and post-test probability for positive and for negative test outcomes [2]. Suppose we ask of a binary diagnostic test; how much does a positive test outcome tell us, in information units, on average over both actually diseased and actually non-diseased subjects? The answer is provided by \( D(b \| a) \). And if we ask, of the same test; how much does a negative test outcome tell us, in information units, on average over both actually diseased and actually non-diseased subjects? In this case the answer is provided by \( D(c \| a) \). \( D(b \| a) \) and \( D(c \| a) \) are average information contents of, respectively, positive and negative test outcomes. In conclusion, we have seen that information graphs can be used to illustrate graphically the Kullback-Leibler divergences as calculated by Lee and by Benish, characterizing different aspects of diagnostic information for a diagnostic test [1, 2].

Yours sincerely

Dr. Gareth Hughes

Acknowledgment
SRUC receives grant-in-aid from the Scottish Government.

References