Advanced GIS - Class Notes on Weight of Evidence

Eugenio Arima

November 30, 2018

1 Analysis with Binary Maps

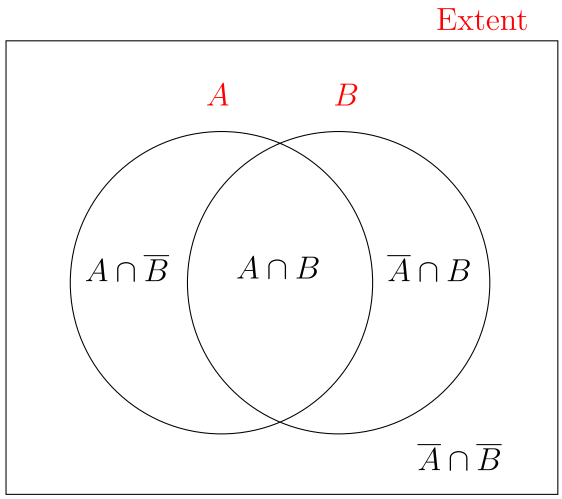
Note: This material comes largely from Bonham-Carter [1994] book. I highly recommend it for those interested in these types of methods.

Suppose you have two binary maps A and B. Cross-tabulate both maps to obtain a 2×2 table.

	A	$\overline{A}$	
B	T_{11}	T_{10}	$T_{1.}$
$\overline{B}$	T_{01}	T_{00}	$T_{0.}$
	$T_{.1}$	$T_{.0}$	$T_{..}$

The subscript 1 denotes the presence of a certain characteristic and 0 its absence (row, column). 'B' could be for example the epicenter of earthquakes and A a buffer distance from a fracking operation or from a geological fault. Let T_{11} denote the count of cells (or area) of characteristics that are present on both maps, T_{10} is the count of cells where B is present and A absent ($\overline{A}$), T_{01} where B is absent ($\overline{B}$) and A is present, and T_{00} where both are absent. T_1. is the total number of cells where B is present, T_0. total cells where B is absent, $T_{.1}$, total where A is present, and $T_{.0}$ total where A is absent; $T_{..}$ is the total number of cells or extent area. If all cells are of same size, which is often the case in GIS, then the math below can be done either by area or by cell count because this is a multiplying constant that vanishes in all ratios.

The Venn Diagram looks as follows:



In set notation, we have these relationships translated into areas as follows:

$$\begin{aligned} Area(A \cap B) &= T_{11} \\ Area(A \cap \overline{B}) &= T_{01} \\ Area(\overline{A} \cap B) &= T_{10} \\ Area(\overline{A} \cap \overline{B}) &= T_{00} \end{aligned}$$

and of course, $T_{1.}$ is set B, $T_{.1}$ is set A, and the extent is $T_{..}$.

2 Conditional Probabilities and Odds

The conditional probability of B occurring given the presence of A is written as $P(B|A)$. Conditional probability is defined as $P(B \cap A)/P(A)$. This can be expressed in terms of cross-tabulation areas as:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{p_{11}}{p_{.1}} = \frac{T_{11}}{T_{.1}}$$

where the little p is for probability. The last equality basically says that these probabilities can be expressed in terms of areas.

We can express the relationship between A and B in terms of conditional odds (probability of occurrence over probability of non-occurrence). Since these are binary maps, these values can be easily determined as follows:

$$O(B|A) = \frac{P(B|A)}{1 - P(B|A)} = \frac{P(B|A)}{P(\overline{B}|A)}$$

This can also be expressed in terms of area:

$$O(B|A) = \frac{p_{11}/p_{.1}}{p_{01}/p_{.1}} = \frac{p_{11}}{p_{01}} = \frac{T_{11}}{T_{01}}$$

Similarly, we can calculate the conditional odds of B given the *absence* of A, $O(B|\overline{A})$:

$$O(B|\overline{A}) = \frac{p_{10}/p_{.0}}{p_{00}/p_{.0}} = \frac{p_{10}}{p_{00}} = \frac{T_{10}}{T_{00}}$$

Combining the two conditional odds expressions we obtain a measure of association between the two binary patterns known as the **odds ratio** O_R, defined as:

$$O_R = \frac{O(B|A)}{O(B|\overline{A})} = \frac{T_{11}T_{00}}{T_{10}T_{01}}$$

If we take the natural log of this expression, we convert the odds ratio to a logit scale. This new index is called *contrast*, C_W.

$$C_W = \ln O(B|A) - \ln O(B|\overline{A})$$

With a little bit of manipulation using Bayes's law (see note), you can also show that (Bonham-Carter, pg. 307):

$$O(B|A) = O(B)\frac{P(A|B)}{P(A|\overline{B})}$$

where the term $\frac{P(A|B)}{P(A|\overline{B})}$ is known as the *sufficiency ratio*.

Likewise, we can write:

$$O(B|\overline{A}) = O(B)\frac{P(\overline{A}|B)}{P(\overline{A}|\overline{B})}$$

and the term $\frac{P(\overline{A}|B)}{P(\overline{A}|\overline{B})}$ is called *necessity ratio*.

3 Weights of Evidence

The contrast however is usually expressed as the difference between the weights in "weights of evidence." A pair of weights W^+ and W^- are defined as the difference between the unconditional or 'naive' and the conditional or posterior logits. In other words, what is the "gain" or difference in our knowledge about event B before and after having information A?

$$W^+ = \ln O(B|A) - \ln O(B) = \ln\left[\frac{O(B|A)}{O(B)}\right] = \ln\left[\frac{T_{11}/T_{01}}{T_{1.}/T_{0.}}\right] = \ln\left[\frac{T_{11}T_{0.}}{T_{01}T_{1.}}\right]$$

and

$$W^- = \ln O(B|\overline{A}) - \ln O(B) = \ln\left[\frac{O(B|\overline{A})}{O(B)}\right] = \ln\left[\frac{T_{10}/T_{00}}{T_{1.}/T_{0.}}\right] = \ln\left[\frac{T_{10}T_{0.}}{T_{00}T_{1.}}\right]$$

The above equalities follow from the fact that $O(B|A) = P(B|A)/P(\overline{B}|A)$, which in area becomes T_{11}/T_{01}. And $O(B|\overline{A}) = P(B|\overline{A})/P(\overline{B}|\overline{A})$, which becomes T_{10}/T_{00}.

The *contrast* is then $C_W = W^+ - W^-$. The magnitude of contrast reflects the overall strength of the spatial association between factors A and B (could think of causality between buffer distance to faults and earthquakes).

From those relationships, we can produce the *posterior* logit:

$$\ln O(B|A) = \ln O(B) + W^+$$

and

$$\ln O(B|\overline{A}) = \ln O(B) + W^-$$

You can think of $O(B)$ as the *naive odds* (in a Bayesian sense, see below). Let the proportion of B in the extent of the study area be δ, then the 'naive' odds is $\delta/(1-\delta)$. This δ is what is known as a "naive" probability because it is essentially the proportion of the event B in the area.

Suppose you have five different binary maps $A_1, A_2, ..., A_5$. Then the *posterior logit* when all factors are present is,

$$\ln O(B|A_1, A_2, ..., A_5) = \ln O(B) + \sum_{i=1}^{5} W_i^+ \equiv \zeta$$

and the *posterior odds* are $\exp(\zeta) \equiv \rho$. We can finally convert that back to *posterior probabilities* $\pi = \rho/(1+\rho)$. This follows from the definition of odds $\rho = \pi/(1-\pi)$, where π is the posterior probability. Note that you can use any combination of weights depending on whether each one of the factors are present or absent (see Excel spreadsheet) and calculate its posterior probability given the factors. A general posterior logit equation would be:

$$\ln O(B|\widehat{A}_1, ..., \widehat{A}_n) = \ln O(B) + \sum_{i=1}^{n} W_i^{+,-}$$

where the $(+,-)$ superscript indicate whether the weight is for presence or absence (symbol$\hat{}$) of the i^{th} factor.

4 A note on Bayesian theory (skip)

You probably have heard about Bayesian theory. It all starts with the definition of conditional probabilities. We know that

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Likewise,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

It turns out that $P(B \cap A) = P(A \cap B)$. Check the Venn diagram and see if you agree with this statement. Setting those two equalities and rearranging terms, we get:

$$\begin{array}{ll} P(B|A)P(A) & = P(A|B)P(B) \\ P(B|A) & = P(A|B)P(B)/P(A) \\ or & \\ P(A|B) & = P(B|A)P(A)/P(B) \end{array}$$

Either one of the last two equations is the famous Bayes' law. This is really a device for "inverting" probabilities. If we take the last equation for instance, $P(A|B)$ is called the posterior distribution of A given the data B. $P(A)$ is called the prior distribution and $P(B|A)$ is called the likelihood. In Bayesian analysis we are typically interested in obtaining a set of parameters, say Θ, given a set of data X. Bayes rules can be written as:

$$p(\Theta|X) = \frac{p(X|\Theta)p(\Theta)}{p(X)}$$

and we typically "ignore" the marginal probability $p(X)$ because it does not depend on Θ and, for fixed X, is just a constant. This term is also difficult to obtain algebraically because it involves multidimensional integration over a set of parameter values. $p(X)$ essentially normalizes the probabilities to one but we can do some tricks to get over this difficult problem (i.e use conjugate priors or computational methods, see below). Thus, Bayesian analysis uses:

$$p(\Theta|X) \propto p(X|\Theta)p(\Theta)$$

where the term $\propto$ means "proportional to." This means that if we know the likelihood that we get a particular set of data given some (prior) parameters and an idea of the prior probability, we can get the posterior given the data.

The intuition behind Bayesian analysis is quite simple. Of course, implementation of it is much more complicated. It was often the case to use "conjugate priors," which are distributions of the same family as the likelihood. Therefore, once those two were multiplied, we would know the distribution of the posterior, which made the whole process more attainable but it did constrain our choices for likelihoods and priors distributions. A breakthrough occurred with a paper by Gelfand and Smith [1990] who showed that we could use computers and algorithms (e.g. Metropolis-Hastings, Gibbs sampler) to implement what is know as Markov Chain Monte Carlo approach to obtain a distribution very close to the posterior by drawing sequentially from each distribution (prior and likelihood) one at a time, hundreds or thousands of times. The theory behind this paper was laid out much earlier (e.g. Fundamental Theorem of Markov Chains, algorithms) but it did make Bayesian analysis more popular among many practitioners (myself included).

References

Graeme F Bonham-Carter. *Geographic information systems for geoscientists: modelling with GIS*, volume 13. Elsevier, 1994. ISBN 1483144941.

A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410): 398–409, 1990.

Last Modified: November 30, 2018