Permanence of Ratios

Fábio Pereira[1] and Clayton V. Deutsch[2]

[1]University of Alberta
[2]University of Alberta

Learning Objectives

- Understand the Permanence of Ratios method to integrate multiple data.
- Derive Permanence of Ratios equations starting with Bayes' law.
- Build intuition on Permanence of Ratios through example.

1 Introduction

Numerical subsurface models estimate resources and quantify uncertainty. Direct sampling is usually sparse and geological variability occurs in several scales. Predicting spatial uncertainty is straightforward when a probability model of the unsampled locations can be inferred. For univariate cases the procedure to estimate and quantify uncertainty are well established in the literature. However, it is common to have several secondary variables sampled in the domain. These secondary variables may provide information about the resource being modeled.

The context of this Lesson is prediction of a binary categorical variable, such as the presence of a rock type or facies, with multiple secondary data sources. The secondary variables could be spatially distributed data of the same variable, geophysical remote sensing or other geological data. The variable being predicted will be denoted A. Consider two secondary data sources denoted B and C. The prior information about A in the stationary domain is calculated as $P(A)$. The information content of B and C about A are calculated as conditional probabilities $P(A|B)$ and $P(A|C)$. A detailed explanation and examples of how to calculate these quantities can be found in (Deutsch & Deutsch, 2018). Therefore, to predict the binary categorical variable utilizing all information sources, it is necessary to infer $P(A|B,C)$

The probabilistic relation between A, B and C simultaneously is not well understood due to incomplete sampling and redundant information. The likelihood of the secondary variables given the primary may be infeasible to calculate. Therefore, a full probabilistic model of the joint, marginal and conditional probabilities is unknown from the data. This lesson explores the Permanence of Ratios model, a methodology to combine the information provided by each variable into a single probability value.

2 Background

Using Bayes' theorem, the combined conditional probability can be written as:

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

B and C can be switched in the decomposition. Being able to calculate the conditional probability at each location allows simulation to proceed and the joint spatial uncertainty to be characterized. Even though Bayes' theorem provides an analytical solution to the problem, estimating the quantities $P(C|A,B)$ and $P(B,C)$ from the data is difficult (Journel, 2002). Some form of Co-kriging could be used, however, besides being computationally intensive, the approach relies on a generalized linear regression model, which may be inadequate when combining categorical and secondary continuous variables (Hong & Deutsch, 2009). Probability combination schemes may be used to estimate the desired conditional probability. These methods, developed independently in several different research areas, combine the primary prior probability with conditional probabilities calculated from secondary variables into a single conditional probability.

Two simple methods to calculate the combined conditional probabilities is to assume full independence or conditional independence. This two assumptions can be written as:

Full Independence, assuming $P(B,C) = P(B)P(C)$ and using Bayes' Inversion:

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

However, this hypotheses is not robust and may be inconsistent (Journel, 2002). For example, if $P(A) = 0.3$, $P(A|B) = 0.8$ and $P(A|C) = 0.6$ would entail a combined probability $P(A|B,C) = 1.6$ (Journel, 2002). Since a probability cannot exceed the value of one, this result shows that numerical instability may arise with this hypothesis.

Conditional Independence, assuming $P(B|A,C) = P(B|A)$ and $P(C|A,B) = P(C|A)$:

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

Using Bayes' inversion:

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

$$P(C|A) = \frac{P(C)P(A|C)}{P(A)}$$

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

This hypothesis, even though less restrictive and more robust than full independence (Ortiz, 2002), does not eliminate the joint and marginal probabilities of B and C. In earth sciences, these quantities may be hard to get from data. Often each variable is sampled at different locations, directly or indirectly, and the dependency between B and C can be complex. This non-trivial relationship may lead to numerical instability and inconsistency when using a naive independence assumption (Journel, 2002; Krishnan, 2008).

The Permanence of Ratios formalism attempts to build a more robust scheme to integrate data from information from multiple sources.

3 Theory of Permanence of Ratios

First, using Journel's (2002) notation, the ratios are defined as:

$$a = \frac{1-P(A)}{P(A)}, b = \frac{1-P(A|B)}{P(A|B)}, c = \frac{1-P(A|C)}{P(A|C)}$$

$$x = \frac{1-P(A|B,C)}{P(A|B,C)}$$

Permanence of Ratios assumes that the marginal gain in information of integrating data event C to the knowledge of A is the same before or after knowing B (Journel, 2002). Therefore, Permanence of Ratios model can be stated as:

$$\frac{x}{b} = \frac{c}{a} \quad or \quad \frac{x}{c} = \frac{b}{a}$$

$$x = \frac{bc}{a}$$

Using the above assumption, it is possible to solve for the desired probability as:

$$P(A|B,C) = \frac{1}{1+x} = \frac{a}{a+bc} = \frac{\frac{1-P(A)}{P(A)}}{\frac{1-P(A)}{PA} + \frac{1-P(A|B)}{P(A|B)}\frac{1-P(A|C)}{P(A|C)}}$$

The above equation eliminates the terms $P(B,C)$, $P(B)$ and $P(C)$ and gives values of $P(A|B,C)$ that are inside the interval $[0,1]$ (Journel, 2002). The model also provides a clear and simple equation to calculate the combined conditional probability.

Interpretation

The ratios are the inverse of the odds of an event. They can be interpreted as probability distance (Caers, Avseth, & Mukerji, 2001; Journel, 2002). For example, as $P(A)$ gets closer to one, a converges to zero: the probability distance shrinks to zero as A is sure to happen.

Also, it can be shown that dividing two ratios of probability is the same as updating the odds ratio of an event by the knowledge brought by a secondary variable. For example, take $\frac{c}{a}$:

$$\frac{c}{a} = \frac{1-P(A|C)}{P(A|C)}\frac{P(A)}{1-P(A)} = OR(A)\frac{1-P(A|C)}{P(A|C)}$$

Where $OR(A)$ is the odds ratio of the event A. Therefore, the ratio is the incremental gain of information after updating the prior distance by the knowledge of the event C. This update can increase or decrease the probability of A happening. As $P(A)$ approaches zero, the distance goes to infinity.

The model can be extended to any number of secondary variables; the derivation can be found in (Journel, 2002).

4 Derivation from Bayes' Theorem

Hong and Deutsch (2007) showed that the conditional probability estimated using conditional independence and Permanence of Ratios model is the same. Connecting the theory proposed by Journel (2002) to a classic result in probability may improve understanding the assumptions and drawbacks of Permanence of Ratios.

Definitions

In order to make the derivation less cumbersome, the complement of an event will be written as:

$$1 - P(A) = P(A^c)$$

$$1 - P(A|.) = P(A^c|.)$$

Following Journel's paper, the ratios can be written in the following form:

$$\frac{c}{a} = \frac{P(C|A^c)}{P(C|A)} \quad and \quad \frac{b}{a} = \frac{P(B|A^c)}{P(B|A)}$$

Derivation

Starting with Bayes Law:

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

Assuming conditional independence amounts to:

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

To get to the same result assuming the Permanence of Ratios hypothesis it is necessary to assume that $P(B,C)$ can be marginalized using Bayes' law, then consider total probability along with conditional independence:

$$P(B,C) = \sum_{A^*=A,A^c} P(A^*)P(B|A^*)P(C|B,A^*)$$

$$P(B,C) = \sum_{A^*=A,A^c} P(A^*)P(B|A^*)P(C|A^*)$$

$$P(B,C) = P(A^c)P(B|A^c)P(C|A^c) + P(A)P(B|A)P(C|A)$$

Therefore:

$$P(A|B,C) = \frac{P(A)P(B|A)P(C|A)}{\sum_{A^*=A,A^c} P(A^*)P(B|A^*)P(C|B,A^*)}$$

Manipulating the above equation:

$$P(C|A) = \frac{P(A|B,C)P(A)P(B|A)P(C|A)}{P(A)P(B|A)} + \frac{P(A|B,C)P(A^c)P(B|A^c)P(C|A^c)}{P(A)P(B|A)}$$

$$P(C|A) - P(A|B,C)P(C|A) = P(A|B,C)\frac{P(A^c)}{P(A)}\frac{P(B|A^c)}{P(B|A)}P(C|A^c)$$

$$\frac{1-P(A|B,C)}{P(A|B,C)} = \frac{P(A^c)}{P(A)}\frac{P(B|A^c)}{P(B|A)}\frac{P(C|A^c)}{P(C|A)}$$

From the definitions, this is equal to:

$$x = \frac{bc}{a}$$

Therefore, the Permanence of Ratios is equivalent to Bayes' theorem where conditional independence is assumed along with the hypothesis that the relation between B and C can be retrieved from the law of total probability considering A and A^c.

5 Practical aspects of Permanence of Ratios

Even though the theory provides a clear and simple equation to combine conditional probability, it is worth noting a somewhat non-intuitive result. The combined probability may not be inside the range of the components.

Consider the case of $P(A) = 0.5$. Given conditional probabilities $P(A|B) = 0.5$ and $P(A|C) = 0.5$ (the center of the plot), all data provides the same information and the combined probability is also 0.5. However, if $P(A) = 0.7$; $P(A|B,C)$ is 0.3 given $P(A|B) = 0.5$, $P(A|C) = 0.5$ (center of the plot). This result may seem counter-intuitive since the prior probability of A is greater than the non-informative scenario of 0.5. As mentioned, the 'ratio of a ratio' can be interpreted as an update of the odds ratio of the event A. So, in this case, the information provided by the events B and C agree that the probability of A occurring is less then the prior suggests. Therefore the updated probability is less than expected.

In a similar way, setting $P(A) = 0.3$; $P(A|B,C)$ at the center of the figure would be 0.7. Similar considerations apply in this case. Starting with a non-informative prior and updating it with agreeing probabilities provides a combined probability that is greater than the marginal ones. This property is called non-convexity, i.e, the updated probability $P(A|B,C)$ doesn't necessarily lies between the priors $P(A)$, $P(A|B)$ and $P(A|C)$.

However, if the prior probability of event A is non-informative, and the additional piece of information provided by B and C disagree, e.g $P(A|B) = 0.8$ and $P(A|C) = 0.2$, the updated conditional probability of A would still be 0.5 as expected.

Real world application

Caers, Avseth and Mukerji (2001) used Permanence of Ratios to integrate data and create a fine-scale reservoir model for turbidite system constrained by prestacked seismic and well-log data. The model was used to integrate the different sources of information and use the calculated conditional probability in a multi point statistic simulation of shale and sand facies.

Deutsch and Deutsch (2018) discuss Bayes' theorem for geostatistical mapping. The text provides an example of a map generated by the Permanence of Ratios model. A more detailed description of a geostatistical work-flow that incorporates secondary data can also be found.

6 Drawbacks of the Permanence of Ratios model

Permanence of Ratios provides a clear and simple analytical solution to the problem of integrating information from different sources. However, some care must be taken when utilizing it. First, the calculation using the model is only consistent when A is binary. Otherwise the resultant probability may be greater than one. Hong (2010) provides a simple and straightforward numerical example of this feature.

A key step in the derivation above is the binary marginalization of the joint probability $P(B,C)$ using the events A and A^c. Bayes' Law of Total Probability is only consistent when the variables used to decompose the probability provide a proper partition of the sample space, i.e they are disjoint and their union forms the sample space. Even though the events A and A^c are disjoint, i.e their intersection equals zero, their sum is only equal to the sample space when the variable is binary.

The model does not account, at least directly, for any known information on the relationship of B and C, e.g $P(B|C)$. To address the bias that conditional independence may generate, a weighted combination scheme might be used. This approach allows for dependency between secondary variables by using weighted combinations of the elementary probabilities (Hong & Deutsch, 2007; Journel, 2002; Krishnan, 2008).

7 Summary

Permanence of Ratios provides a clear analytical solution for the data integration problem. If the primary variable is binary, Permanence of Ratios seems more robust then a simple conditional independence assumption. The simultaneous use of categorical and continuous secondary variables at the same time makes the model useful in classification problems. Permanence of ratios could be extended to any number of secondary variables.

8 References

Caers, J., Avseth, P., & Mukerji, T. (2001). Geostatistical integration of rock physics, seismic amplitudes, and geologic models in north sea turbidite systems. *The Leading Edge*, *20*(3), 308–312.

Deutsch, C. V., & Deutsch, J. (2018). An application of bayes theorem to geostatistical mapping. *Geostatistics Lessons. Retrieved from Http:// Geostatisticslessons. Com/-Lessons/Bayesmapping*.

Hong, S., & Deutsch, C. V. (2007). Methods for integrating conditional probabilities for geostatistical modeling. *Centre for Computatinal Geostatistics (CCG) Annual Report, 9(105)*.

Hong, S., & Deutsch, C. V. (2009). On secondary data integration. *Center for Computational Geostatistics (CCG) Annual Report, 11(101)*.

Journel, A. (2002). Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses. *Mathematical Geology*, *34*(5), 573–596.

Krishnan, S. (2008). The tau model for data redundancy and information combination in earth sciences: Theory and application. *Mathematical Geosciences*, *40*(6), 705.

Ortiz, J. M. (2002). Short note: Naive bayes classifiers and permanence of ratios. *Center for Computational Geostatistics (CCG) Annual Report, 4*, 12.

Citation

Pereira, F. L., & Deutsch, C.V (2020). Permanence of Ratios. In J. L. Deutsch (Ed.), Geostatistics Lessons. Retrieved from
http://geostatisticslessons.com/lessons/permanenceofratios