The Place of Probability Combination Schemes

Sahyun Hong and Clayton V. Deutsch

Data integration requires the multivariate distribution among the considered data sources. Probability combination schemes including permanence of ratios, the tau model, the nu model and the lamda model have received some attention recently. These approaches involve the combination of each calibrated probability conditioned to individual data source in order to approximate the joint probability, which is termed probability combination method. In many cases, direct density estimation is preferred; however probability combination approach is more useful in certain cases. Also, probability combination methods are applicable for information integration (different from data integration) to get consensus prediction. Probability combination model is overviewed in this paper. The non-convexity characteristic is shown with numerical example.

Introduction

Multivariate modeling in geostatistics is common. Various data sources with varying degrees of quality are often available and even a single source such as geophysical provides different attributes. Geostatistical data integration requires the building of a probability model of properties in terms of optimal use of these all data sources.

Diverse data includes local drilling samples, geophysical measurements, conceptual geology and analogue data. Well drilling data and geophysical data are numeric data with different spatial coverage. Conceptual geology and analogue data are often captured and quantified by training images (TIs). These sources of information are aimed at characterizing the variable of primary interest despite of varying degree of spatial extent, quality, support and correlation. The purpose of data integration is to build joint probability distribution conditioned to all these diverse data simultaneously. Thus, the built probabilistic model must account for redundancy among these data.

There are two ways to construct joint probability of interest: indirect estimation method is termed probability combination schemes (PCS) and direct density estimation method. Direct density estimation is to construct multivariate probability directly: multivariate pdf is directly built and the conditional probability of interest is extracted from the multivariate distribution. Indirect probability combination method is to approximate the target joint probability by combining univariate probability. Various names of PCS model exist depending on how to measure and calibrate the data redundancy. CCG proposed the lamda model as one of probability combination scheme last year and proposed a direct density estimation algorithm this year (see paper-101/102 in this volume). Direct estimation method is mathematically robust and transparent in terms of accuracy and explicitly quantifying data redundancy. Thus this method should be dominant and a standard in data integration study. However, indirect method is sometimes necessary and chosen as an alternative to direct estimation method in case of integrating non-numeric data and qualitative information. Besides, PCS is used to aggregate information that is integrated probability at higher level than data source integration level.

This paper overviews probability combination approach and the place of PCS will be discussed in the light of both data integration and information integration level.

Data Integration by Probability Combination Models

Probability combination scheme (PCS) has been developed independently in many research areas in order to find a consensus probability using several single source derived probabilities. Main principle of PCS is to approximate the target joint probability through linking the individual probability that is already computed using each datum. For the nominal expression, let us set data sources to be considered as $(D_1,\ldots,D_m)$ and the event of primary variable as A. Now, the joint probability is approximated by PCS model generically denoted as $\Phi[\cdot]$,

$$\frac{p(A\,|\,D_1,\ldots,D_m)}{p(A)} \approx \frac{1}{p(D_1,\ldots,D_m)}\Phi\left[\frac{p(A\,|\,D_1)\,p(D_1)}{p(A)},\ldots,\frac{p(A\,|\,D_m)\,p(D_m)}{p(A)}\right] \qquad (1)$$

Equation (1) is derived by Bayes law. Event A of primary variable can be interpreted as anything depending on the purpose of study. For example, event A can be occurrence of the moving target in intelligent navigation and data sources ($D_1,\ldots,D_m$) can be images from thermal camera, radar camera and so on. In geologic applications, the event A can be interpreted as being either 'ore' or 'waste' in mining engineering, and either 'net facies' or 'non-net facies' in petroleum engineering. The joint probability of interest is a probability of event A given all data, and it is estimated at every visited cell location $\mathbf{u}$, $\mathbf{u}\in$ A. p(A) is a global probability of event A. Univariate conditional probabilities $p(A|D_i)$, i=1,…,m are obtained from calibrating of each datum D_i with respect to event A. Figure-1 illustrates schematic diagram of probability combination method.

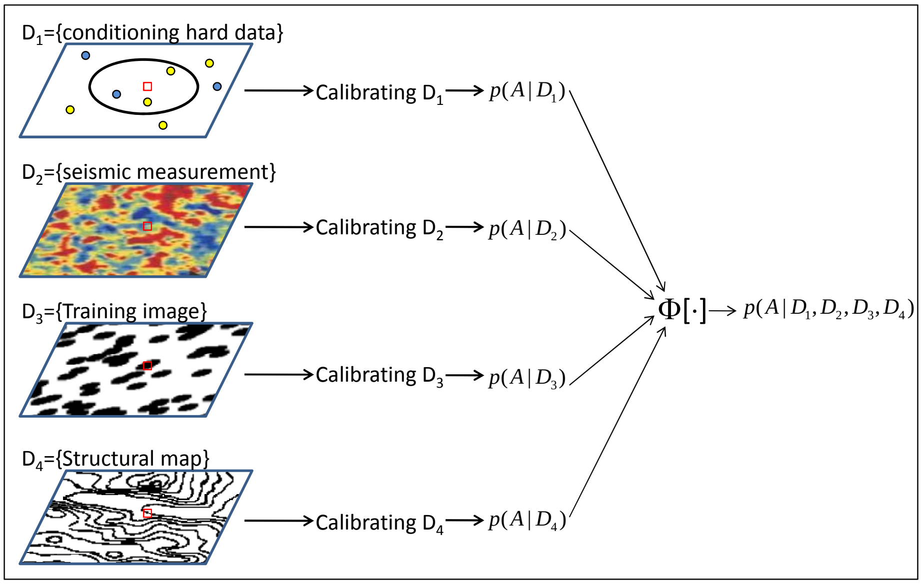

Figure-1: Schematic illustration of probability combination scheme.

Function $\Phi[\cdot]$ is a combination model and it has various names depending on how to measure data redundancy among data sources ($D_1,\ldots,D_m$). The simple model is to assume conditional independence (assuming no redundancy) among data sources, which is called permanence of ratios (PR-model). PR-model assumes that data sources are independent each other conditioned to geologic event A. Tau-model is an advanced model and it mitigates independence assumption by imposing exponential weights (weights are denoted as τ) on each calibrated probability. Lamda model is another name of tau-model but it quantifies data redundancy through calibrating the approximated probability and the hard data at well locations. Equation (1) now can be expressed as one term such as,

$$\frac{p(A|D_1,\ldots,D_m)}{p(A)} \approx \frac{1}{p(D_1,\ldots,D_m)} \prod_{i=1}^{m} \left(\frac{p(A|D_i)\,p(D_i)}{p(A)} \right)^{\lambda_i} \qquad (2)$$

The joint probability is approximated by each calibrated term and exponential redundancy weights expressed as $\lambda_i \in [0,1]$, i=1,…,m. Weights λ_i, i=1,…,m control the influence of each calibrated probability. They are associated with inherent data redundancy; however, the role of exponential weights is not clearly understood. The combination model converges to PR if λ_i, i=1,…,m are set as 1. Several weights selection methods to measure redundancy have been proposed, but it is not easy to get optimal weights since data redundancy is related with data ($D_1,\ldots,D_m$) and event A at the same time. In PCS approach, data redundancy is implicitly accounted for in terms of alternative criterion and/or objective function (see paper-105 in report 9). This makes PCS less attractive because the meaning of calibrated weights is not transparent and untraceable. When finding the redundancy weights another constraint should be considered. Basic probability requirements are followings:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(A|D_1,\ldots,D_m)\,p(D_1,\ldots,D_m)\,dD_1,\ldots,dD_m = p(A) \qquad (3\text{-}1)$$

$$\sum_{A \in All\ Events} p(A \mid D_1,...,D_m)\, p(D_1,...,D_m) = p(D_1,...,D_m) \tag{3-2}$$

When inserting the approximated term instead of original $p(A|D_1,\ldots,D_m)$ then

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{p(A \mid D_1)\, p(D_1)}{p(A)} \right)^{\lambda_1} \times \cdots \times \left(\frac{p(A \mid D_m)}{p(A)} \right)^{\lambda_m} dD_1,...,dD_m = 1 \tag{4-1}$$

$$\sum_{A \in All\ Events} \left(\frac{p(A \mid D_1)\, p(D_1)}{p(A)} \right)^{\lambda_1} \times \cdots \times \left(\frac{p(A \mid D_m)\, p(D_m)}{p(A)} \right)^{\lambda_m} p(A) = p(D_1,....,D_m) \tag{4-2}$$

Above relations are marginality constraints that should be met. Exponential weights, thus, must be calibrated based on both the certain optimal measure of redundancy and marginal constraints. Direct density estimation is an alternative to PCS in the light of its mathematical rigorousness and transparency: the method builds multivariate density constrained with marginal conditions and avoids the complicated step of weight selection procedure. The details about direct density estimation are not discussed in this paper. Its theoretical background and applications are demonstrated in other papers (see paper-101/102 in this volume). But it should be noted that direct density estimation method explicitly account for data redundancy between data sources and meets the marginality property.

Despite of direct estimation method's advantages, probability combination scheme is sometimes to be considered as data integration technique. For example, geologic interpretation and modern analogue could be interpreted by expert geologist and those qualitative data will be quantified as facies probability. Deterministic model by expert system is critical and it is often enough for geologic modeling. Even if this probability assignment does not show variations at small scale it could provide important large scale heterogeneity.

Non-Convexity of Data Integration

The probability combination scheme is the approximation way of the joint probability of interest when the joint modeling of data sources is not applicable. Let us see the approximation equation (1) again and let us assume conditional independence (λ_i=1,i=1,…,m) then,

$$\frac{p(A \mid D_1,...,D_m)}{p(A)} \approx \frac{p(A \mid D_1)}{p(A)} \times \cdots \times \frac{p(A \mid D_m)}{p(A)} \times C(D_1,...,D_m) \tag{5}$$

The term $p(D_1) \times\cdots\times p(D_m)/p(D_1,\ldots,D_m)$ in eq. (1) is summarized as $C(D_1,\ldots,D_m)$ that is independent of event A only function of data sources $(D_1,\ldots,D_m)$. $p(A)$ is a priori probability or global proportion of event A. The below table demonstrates one example of non-convexity of the integrated probability. Binary facies case (A=1 or 0) with two data source denoted as (D_1,D_2) is prepared.

p(A=1)	$p(A=1\|D_1)$	$p(A=1\|D_2)$
0.6	0.7	0.88

Probabilities of A=1 conditioned to D_1, and D_2 are both higher than the global probability 0.6. Equation (5) is replaced by the above numeric value then,

$$\frac{p(A=1 \mid D_1,D_2)}{0.6} \approx \frac{0.7}{0.6} \times \frac{0.88}{0.6} \times C \Rightarrow p(A=1 \mid D_1,D_2) = 1.0267 \times C$$

$$\frac{p(A=0 \mid D_1,D_2)}{0.4} \approx \frac{0.3}{0.4} \times \frac{0.12}{0.4} \times C \Rightarrow p(A=0 \mid D_1,D_2) = 0.09 \times C$$

Term C is regardless of event A (=1 or 0) so it is canceled by normalizing $p(A=1|D_1,D_2)+p(A=0| D_1,D_2)=1$. Integrated probability $p(A=1| D_1,D_2)$ is, thus, calculated as 0.919 which is much higher than p(A=1), $p(A=1|D_1)$ and $p(A=1|D_2)$.

p(A=1)	p(A=1|D_1)	p(A=1|D_2)	Esimated p(A=1|D_1,D_2)
0.6	0.7	0.88	0.919

The reverse case is also shown in the below. $p(A|D_1)$ and $p(A|D_2)$ are below global probability, which results in much lower integrated probability.

p(A=1)	p(A=1|D_1)	p(A=1|D_2)	Estimated p(A=1|D_1,D_2)
0.6	0.4	0.55	0.352

This property is called a non-convexity of data integration. Data integration amplifies the impact of data sources if they represent possibility in the same direction: higher probability or lower probability than global probability. This non-convexity is very natural. In equation (5), if $p(A|D_i)$ is larger than $p(A)$ then $p(A|D_i)/p(A)$ gets larger than 1, which makes $p(A|D_1,\ldots,D_m)$ to be much larger. In consideration of weighted PCS model, this non-convexity is preserved as well. Weighted combination model is shown in the following,

$$\frac{p(A|D_1,D_2)}{p(A)} \approx \left[\frac{p(A|D_1)}{p(A)}\right]^{\lambda_1} \times \left[\frac{p(A|D_2)}{p(A)}\right]^{\lambda_2} \times C(D_1,D_2,\lambda_1,\lambda_2) \tag{6}$$

The term $p(D_1)^{\lambda 1} \times p(D_2)^{\lambda 2}/p(D_1,D_2)$ is summarized as $C(D_1,D_2,\lambda_1,\lambda_2)$. We assumed redundancy weights λ_1 and λ_2 are only depending on different data source regardless of event A. Above simple numeric example is used again. Redundancy weights λ_1 and λ_2 are arbitrarily chosen as 0.7 and 0.8. High weights being close to 1 means data source D1 and D2 are less redundant.

$$\frac{p(A=1|D_1,D_2)}{0.6} \approx \left[\frac{0.7}{0.6}\right]^{0.7} \times \left[\frac{0.88}{0.6}\right]^{0.8} \times C \Rightarrow p(A=1|D_1,D_2) = 0.908 \times C$$

$$\frac{p(A=0|D_1,D_2)}{0.4} \approx \left[\frac{0.3}{0.4}\right]^{0.7} \times \left[\frac{0.12}{0.4}\right]^{0.8} \times C \Rightarrow p(A=0|D_1,D_2) = 0.1248 \times C$$

Integrated probability $p(A=1|D_1,D_2)$ with redundancy weights is calculated as 0.879 which is still higher than $p(A=1)$, $p(A=1|D_1)$, and it is close to $p(A=1|D_2)$.

p(A=1)	p(A=1|D_1)	p(A=1|D_2)	Estimated p(A=1|D_1,D_2) with λ_1=0.7 and λ_2=0.8
0.6	0.7	0.88	0.879

The degree of non-convexity is reduced when comparing with no redundancy weighting model. Figure-2 demonstrates the non-convexity of the integrated probability. X axis indicates the term $p(A|D_i)/p(A)$ which is within $[0,\infty]$. X-axis value being less than 1 represents data source D_i predict lower probability than global probability $p(A)$. This makes much lower integrated probability $p(A|D_1,\ldots,D_m)$. When adopting exponential weights, the effect of non-convexity is reduced. The curve (weighted model with λ=0.3 and 0.7) goes below the straight line (PR-model) if x-axis value is greater than 1, and the curve goes above the straight line if x-axis value is between 0 and 1. This arithmetic behavior of PCS model can be interpreted such that no redundant data D_i, i=1,…,m are totally new information (no information overlapping) so the integrated probability will be far away from the used several estimators. No redundant data should be combined by PR-model maximizing the non-convexity. If data sources are somewhat redundant those are to be interpreted as being partially new informative. Thus, weighted PCS model that reduces non-convexity should be considered. Integrated probability will not be far away from the used estimators, $p(A|D_i)$, i=1,…,m.

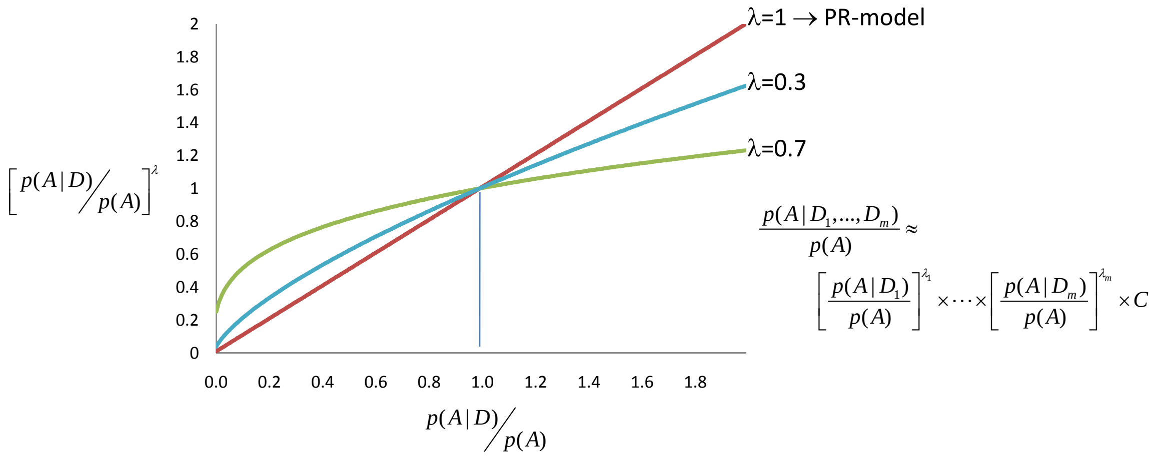

Figure-2: Non-convexity property of the integrated probability

Information Integration by PCS

Another potential application of probability combination scheme is to aggregate the information that is obtained from the data integration process. Probability combination at this level should be differentiated from PCS at data integration level. *Data integration* is referred to as assimilating the available data such as well data, seismic data, TI and others. Integrated results are to be various and different based on the choice of integration methodology. Different integrated results sometimes need to be considered together as integration procedure often requires the subjective user-input parameters and algorithm settings. Evaluating of multiple results will ensure the production of the most satisfactory possible. Pooling different integration results, conditioned to the same data sources, involves procedure with the goal of combining results to find consensus estimates. This is termed *information integration*. Figure-3 illustrates data integration and information integration.

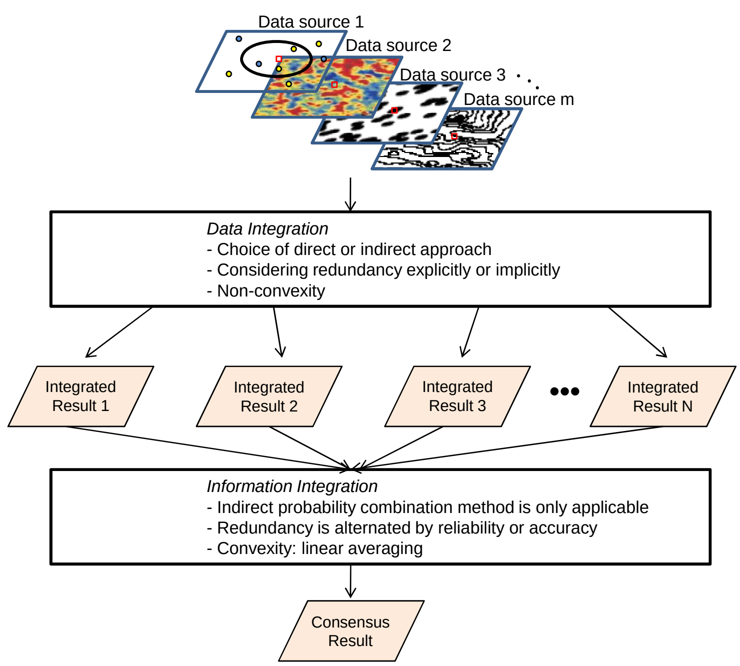

Figure-3: Data integration and information integration

Data integration step builds the joint probability given all available data by either indirect or direct estimation method. Direct density estimation is preferred when data redundancy can be explicitly quantified, and probability combination method is adopted when data redundancy is quantified only by implicit manner. Non-convexity is a significant characteristic in this stage. Various integrated predictions are resulted from different choice of integration model or parameters with the same data. Information integration aggregates those decisions to provide consensus predictions. The consensus must be found by weighted linear averaging: convexity is a desirable property in information integration,

$$p_{\text{consensus}}(A \mid D_1,...,D_m) = \frac{1}{\sum_{i=1}^{N}\alpha_i}\sum_{i=1}^{N}\alpha_i p_i^{\text{integrated}}(A \mid D_1,...,D_m)$$

Weights $\alpha_i \in [0,1]$, i=1,…,N might be selected in terms of integration model reliability, accuracy or other criterion. Linear averaging lets consensus probability exist between the integrated results.

Discussion

CCG proposed two approaches to assimilate multiple data sources in geostatistical reservoir modeling: indirect probability combination scheme (PCS) and direct density estimation. Both methods are aiming at construct probabilistic model conditioned to all data sources; however, direct density estimation accounts for data redundancy in explicit manner and PCS infer optimal redundancy weights via a certain calibration algorithm. Direct estimation is preferred in most cases but PCS is more appropriate when the data to be integrated is non-numeric and qualitative. Non-convexity is that the integrated probability exists outside individual data source derived probability. This is a very natural phenomenon by probability integration equation. Weighted combination models maintain non-convexity but the degree of non-convexity is controlled by exponential weights.

Another application of PCS is to pool integrated results, which is information integration. Data integration procedure made a full use of data sources with accounting for data redundancy and non-convexity. Information integration is to aggregate those multiple integrated results in order to obtain consensus predictions. Convexity is ensured by linear averaging in information integration.

References

McConway, K. J., 1981, Marginalization and Linear Opinion Pools, *Journal of American Statistical Association*, Vol. 76, No. 374.

Benediktsson, J. A. and Swain, P. H., 1992, Consensus Theoretic Classification Methods, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, No. 4.

Journel, A. G., 2002, Combining Knowledge from Diverse Sources: An Alternative to Traditional Data Independence Hypotheses, *Mathematical Geology*, Vol. 34, No. 5.

Hong, S. and Deutsch, C. V., 2007, Methods for Integrating Conditional Probabilities for Geostatistical Modeling, *Centre for Computational Geostatistics In Report 9.*

Hong, S. and Deutsch, C. V., 2008, Data Integration with Direct Multivariate Density Estimation, *Centre for Computational Geostatistics In Report 10.*

Hong, S. and Deutsch, C. V., 2008, Integrating Secondary Data and Training Images with Direct Density Estimation, *Centre for Computational Geostatistics In Report 10.*