

Jonathan L. Wilson and Kristin A. Duncan

AP0912



A Multinomial-Dirichlet Model for Analysis of Competing Hypotheses

Jonathan L. Wilson^{*,1}, Kristin A. Duncan² *Corresponding author: jwilson@sciences.sdsu.edu

Abstract: Analysis of Competing Hypothesis, a method for evaluating explanations of observed evidence, is used in numerous fields including counterterrorism, psychology, and intelligence analysis. We propose a Bayesian extension of the methodology, posing the problem in terms of a multinomial-Dirichlet hierarchical model. The yet-to-be observed true hypothesis is regarded as a multinomial random variable and the evaluation of the evidence is treated as a structured elicitation of a prior distribution on the probabilities of the hypotheses. This model provides the user with measures of uncertainty for the probabilities of the hypotheses. We discuss inference such as point and interval estimates of hypothesis probabilities, ratios of hypothesis probabilities, and Bayes factors. A simple example involving the stadium relocation of the San Diego Chargers is used to illustrate the method. We also present several extensions of the model that enable it to handle special types of evidence including evidence that is irrelevant to one or more hypotheses, evidence against hypotheses, and evidence that is subject to deception.

Keywords: Analysis of competing hypotheses, Bayesian updating, data fusion, uncertainty

1 Introduction

Analysis of competing hypotheses (ACH) is a method for systematically comparing the likelihoods of competing hypotheses based on the available evidence. Richards Heuer [4] developed the procedure for the CIA in the 1970s, primarily for use in intelligence matters. Heuer's work did not provide a mathematical basis for drawing conclusions, though he did note that formalizing the process would be desirable. A software program from PARC [6] attempts to make ACH quantitative by reporting an inconsistency score for each hypothesis. Paté-Cornell [7] and McLaughlin [5] both present a Bayesian method for updating beliefs in hypotheses as evidence is obtained. The model we present here is richer than that presented in Paté-Cornell, though it is based on a similar idea.

2 Multinomial-Dirichlet Hierarchical Model

The multinomial-Dirichlet model is a generalization of the beta-binomial model in which there are N categories rather than the two categories success and failure. A draw from a multinomial distribution is represented by a vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, where each x_j is the count of observations that fall into category j. Let $\mathbf{p} = (p_1, p_2, \ldots, p_N)$ denote the probabilities of belonging to the N categories so that $\sum_{j=1}^{N} p_j = 1$ and let $n = \sum_{j=1}^{N} x_j$. A Dirichlet distribution is placed on \mathbf{p} ,

A Dirichlet distribution is placed on \mathbf{p} , the category probabilities of the multinomial distribution. The parameter for the Dirichlet distribution is $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$, a vector with nonnegative entries. Letting $\alpha_0 = \sum_{j=1}^N \alpha_j$, the marginal distribution of each p_j is a beta distribution.

The multinomial and Dirichlet distributions are conjugate to one another, meaning that when we start with a Dirichlet prior distribution on the category probabilities and update our knowledge with multinomial data, the resulting posterior distribution on the category (or hypothesis) probabilities is Dirichlet. Specifically, when the prior distribution on **p** is Dirichlet(α), and multinomial data **x** is observed, the posterior distribution on **p**, conditioned on this data, is Dirichlet($\alpha_1 + x_1, \alpha_2 + x_2, \dots, \alpha_N + x_N$).

To pose ACH in terms of multinomial and Dirichlet distributions, we consider the final outcome, or determination of the true hypothesis to be a single draw from a multinomial distribution. Evaluating the evidence provides information about \mathbf{p} , the parameter of this multinomial distribution, which gives the probabilities of the hypotheses.

2.1 Algorithm

The algorithm for implementing our procedure is as follows;

- 1. Construct the framework for the ACH matrix. List hypotheses, H_j , $j = 1, \ldots, N$ as column headings and evidence items, E_i , $i = 1, \ldots, M$ as row headings. Include prior beliefs or "other evidence" as the first evidence item.
- 2. Assign evidence weights. Determine the equivalent prior sample size (n_{ess}) of the evidence as a whole. Assign weights, w_i , to the evidence items indicating their strength or relative importance. Scale these weights so that $\sum_{i=1}^{M} w_i = n_{ess}$.
- 3. Relate evidence to hypotheses. Proceeding one row at a time, rate the relative likelihood of E_i conditioned on each hypothesis by filling in the matrix with values x_{ij} . One may begin by assigning $x_{il} = 1$ where H_l is the hypothesis under which we are least likely to observe E_i . Continue by assigning the other x_{ij} values relative to x_{il} . If E_i is twice as likely to be observed when H_j is true compared to when H_l is true, then $x_{ij} = 2$. After the initial assignment, scale these values so that $\sum_{j=1}^{N} x_{ij} = w_i$.
- 4. Compute the posterior. The posterior distribution of **p**, the probabilities of the hypotheses, is Dirichlet with parameter α given by $\alpha_j = \sum_{i=1}^{M} x_{ij}$, for $j = 1, \ldots, N$. The marginal posterior distributions of the individual p_j parameters are beta $(\alpha_j, \alpha_0 \alpha_j)$.

Step 1 corresponds to steps 1, 2, and 3 of Heuer's method. Step 2 is a formalization of Heuer's call to "Analyze the diagnosticity of the evidence and arguments". The method of assigning x_{ij} values in Step 3 is posed as in McLaughlin [5], though we note here that 0 values are permitted for x_{ij} . An alternative way to elicit x_{ij} values is to interpret it as the number of "observations" of H_j with which E_i can be associated. Written out formally, we have

$$y \sim Multinomial(1, \mathbf{p})$$

 $\mathbf{p} \sim Dirichlet(\alpha)$
 $\alpha_j = \sum_{i=1}^M x_{ij} \quad j = 1..N,$ (1)

where y is the outcome we are trying to predict.

2.2 Inference

All inference is conducted using the posterior distribution. The mean of the posterior Dirichlet(α) distribution is

$$\hat{\mathbf{p}} = \left(\frac{\alpha_1}{\alpha_0}, \dots, \frac{\alpha_N}{\alpha_0}\right).$$

Interval estimates provide a measure of uncertainty for the point estimates above, but numerical methods are required to find quantiles of beta and Dirichlet distributions. A 95% equal-tail credible set for a single probability p_j is obtained by finding the .025 and .975 quantiles of the beta $(\alpha_j, \alpha_0 - \alpha_j)$ distribution. Highest posterior density sets [2] will be slightly narrower than the equaltail sets. The width of these interval estimates is strongly affected by the choice of n_{ess} .

3 Extensions

The multinomial-Dirichlet ACH model, as presented thus far yields a Dirichlet posterior distribution on the hypotheses that is easy to compute and easy to use for conducting inference. However, it is missing the ability to handle items that give evidence against one or more hypotheses, items that are not relevant to all hypotheses, and items which may be subject to deception. Unfortunately these extensions result in posterior distributions that are not Dirichlet and this loss of conjugacy makes inference more difficult. The use of Monte Carlo methods, however, makes inference computationally feasible, and samples from the posterior can be obtained fairly quickly.

3.1 Evidence Against

Evidence against hypothesis H_k is expressed as seeing observations that do not fall into category k. In a beta-binomial setting, where there is only one category that is "not H_k ", this is easy to handle. In a Dirichlet model, the observations need to be allocated amongst all the categories that are "not H_k " without changing the relative probabilities of these other hypotheses and without changing our certainty about the relative probabilities of these other hypotheses. This is accomplished by treating the evidence against as a binomial random variable with probability of success $1-p_k$. We know the observations belong to the "not H_k " categories, but we do not know how many observations fall into each. If an evidence item E_i is against a set of hypotheses, rather than just a single hypothesis H_k , it is "for" the complement of this set which we will refer to as \mathbf{F}_i . Then the evidence is treated as a binomial random variable with probability of success $\sum_{j \in \mathbf{F}_i} p_j.$

Let our current knowledge of \mathbf{p} be given by a proper prior distribution $\pi(\mathbf{p})$, and let E_i provide evidence for a set of hypotheses F_i , with the strength of this evidence represented by w_i . Then, the posterior distribution for \mathbf{p} , updated with E_i is given by

$$\pi(\mathbf{p}|E_i) \propto \left(\sum_{j \in \mathbf{F}_i} p_j\right)^{w_i} \pi(\mathbf{p}).$$
 (2)

In practice, an analyst would input plus and minus signs in the matrix to indicate evidence for and against hypotheses and use w_i to represent $\sum_{i \in \mathbf{F}_i} x_{ij}$.

3.2 Irrelevance

When a piece of evidence is irrelevant to hypothesis H_k , it is as though the data associated with H_k is missing. It is not appropriate to assign $x_{ik} = 0$ when E_i is not relevant to H_k as this would reduce the likelihood of H_k relative to the other hypotheses when it should be held constant.

Let our current knowledge of \mathbf{p} again be given by the distribution $\pi(\mathbf{p})$, and let E_i be relevant only to hypotheses in \mathcal{R}_i . Then, the posterior distribution for \mathbf{p} , updated with E_i is

$$\pi(\mathbf{p}|E_i) \propto \left[\prod_{j \in \mathcal{R}_i} \frac{p_j^{x_{ij}}}{\sum_{k \in \mathcal{R}_i} p_k}\right] \pi(\mathbf{p}). \quad (3)$$

This posterior distribution is guaranteed to be proper if the prior distribution is proper.

3.3 Combining Different Types of Evidence

When a user decides to enter an evidence item as either for, against, or irrelevant to a hypothesis, the posterior distribution for **p** is no longer in closed form. In order to express the posterior, we partition the evidence by type into three sets. Evidence items with only numerical values will belong to set **A**. Evidence items that contain irrelevant hypotheses (NA's) will be labeled as set **B**. Evidence items that are just for (+) and against (-) will belong to set **C**. Items that contain both an NA, and + or - will be placed into set **C**. Evidence in set **C** may not contain numeric values.

Using Equations (1),(2), and (3) the posterior distribution for any set of evidence is given by

$$\pi(\mathbf{p}|\mathbf{E}) \propto \prod_{i \in \mathbf{A}} \prod_{j=1}^{N} p_j^{x_{ij}}$$
$$\times \prod_{i \in \mathbf{B}} \prod_{j \in \mathcal{R}_i} \frac{p_j^{x_{ij}}}{\sum_{k \in \mathcal{R}_i} p_k} \times \prod_{i \in \mathbf{C}} \left(\sum_{j \in \mathbf{F}_i} p_j\right)^{w_i}$$
(4)

where N is the number of hypotheses, \mathcal{R}_i is the set of relevant hypotheses for evidence item i, \mathbf{F}_i is the set of hypotheses that item i provides evidence for (+), and w_i is the scaled weight for evidence item i. Importance sampling is used to sample from the distribution.

3.4 Stadium Example with Extensions

The city of San Diego in 2006 declined to provide the owners of the San Diego Chargers football team with the support they were seeking to build a new stadium and redevelop the site of Qualcomm Stadium. The Chargers organization then stated that they will definitely be moving from Qualcomm Stadium, with their contract requiring them to stay only through the end of the 2008 season. Other cities in the Southwest such as Las Vegas and San Antonio have been proposed as new locations for the Chargers. However, the Chargers have stated that they would like to stay in the San Diego area and have considered several other sites in San Diego County; Oceanside, National City,

and Chula Vista. National City dropped from the running in the spring of 2007. In this section we apply the multinomial-Dirichlet ACH model to the problem of predicting the new stadium site for the Chargers based on available evidence. The evidence was obtained from news articles in the San Diego Union Tribune through the spring of 2007. The three hypotheses are Oceanside, Chula Vista, and Other.

Table 1 shows the scaled input of an analyst who used the extensions proposed. n_{ess} was again chosen to be 10. There are four evidence items with complete numeric values belonging to set **A** (E_3, E_5, E_6, E_7) , one with an irrelevant hypothesis belonging to set **B** (E_4) , and one with evidence for and against belonging to set **C** (E_2) . The posterior distribution for **p** is $\pi(\mathbf{p}|\mathbf{E}) \propto p_1^{3.45} p_2^{3.02} p_3^{1.53} (p_1 + p_2)$.

Computing $\sum_{i \in \mathbf{A}} x_{ij}$, the column sums for evidence items in \mathbf{A} , gives $\alpha =$ (2.78, 2.69, 1.53). Next, we update α with evidence in \mathbf{B} to obtain α^* . $E_4 =$ (.67, .33, NA) gives the updated value $\alpha^* =$ (3.30, 2.95, 1.75). Finally, we update α^* with evidence in \mathbf{C} to obtain α^{**} . $E_2 = (+, +, -)$ gives $\alpha^{**} = (4.36, 3.89, 1.75)$.

We now have a close Dirichlet distribution to use as an importance function for importance sampling. Table 2 summarizes the results of the posterior containing estimates of the mean, 95% HPD intervals, and p_{max} with a Monte Carlo sample size of 100,000. Figure 1 displays the true marginal posteriors along with the importance function. Results are quite similar to the previous assignment of matrix values, with Oceanside having a slightly higher posterior mean in this assignment. Since either of these special types of evidence can be interpreted as missing data, there will generally be more variance in the posterior distribution when "NA", "+", and "-" are used than when all matrix entries are numeric.

3.5 Deception

A third type of evidence that can be incorporated into this model is evidence with the potential for deception. The analyst can assign θ_i values which indicate the probability of deception, or continue to add hyperpriors to the set-up, perhaps placing a beta distribution on the θ_i . Once again, the posterior distribution cannot be obtained in closed form, but Monte Carlo methods make inference possible.

4 Discussion

The aim of this paper was to give a probabilistic framework for ACH that is sophisticated enough to provide measures of certainty while at the same time providing a simple interface and easy to interpret output for users who may not have a great deal of training in probability theory. We did not delve into the psychology of reasoning or human logic, seeking only to give a statistical formalization of an existing method for assessing evidence. Our approach involves viewing ACH as a structured elicitation of a prior distribution by equating evidence to multinomial observations.

The ability to weight evidence items by importance or diagnosticity is a great advantage of our model over simpler approaches such Paté-Cornell's. In the simpler set-up, updating with a single signal or evidence item E results in the same posterior distribution on the hypotheses whether the prior for H is noninformative or based on hundreds of other evidence items.

We admit that the assignment of weights to hypotheses can seem overly subjective as can the assignment of an equivalent sample size n_{ess} . One can always examine the sensitivity of conclusions to these choices, though care should be taken that these parameters are not manipulated to obtain a preconceived conclusion. Weed [9] provides a review of the concept of "weight of evidence".

We conclude with a comment on diagnosticity and discrimination. Some ACH guides call for discarding evidence with little diagnostic power. Heuer refers to the diagnosticity of a piece of evidence as its helpfulness in judging the relative likelihoods of the hypotheses. Unfortunately, this can be misinterpreted to mean that an item of evidence should be discarded if it does not help you choose one hypothesis or set of hypotheses over another hypothesis or set of hypotheses; i.e. if it does not push the probabilities of the hypotheses closer to 0 and 1. Such a piece of evidence does not help to discriminate between hypotheses, but it may still be diagnostic. When measuring the uncertainty of our estimates, we want to include relevant evidence even if it tells us that the relative likelihoods of two hypotheses become closer to one another. Excluding nondiscriminatory evidence will likely result in higher likelihood ratios for hypotheses. This is analogous to the problem arising from medical journals that only publish papers with results significant at the .05 level. Canvassing journals to perform meta-analysis on like studies results in overestimating treatment effects because studies with insignificant effects are excluded. Heuer cites psychological studies in which experts' certainty in their evaluations becomes higher with more evidence even though the quality of their evaluations remains the same. Including all relevant evidence, regardless of its discriminatory power could mediate this effect.

	w	Oceanside	Chula Vista	Other
E1: Prior beliefs or unlisted evidence	0.0			
E2: Chargers say they want to stay in San Diego	2.0	+	+	-
E3: Chargers want financial assistance (any city)	2.5	0.75	0.25	1.50
E4: Chargers like parking and transit in Oceanside	1.0	0.67	0.33	NA
E5: San Diego State University wants to be involved	0.5	0.14	0.35	0.01
E6: Chargers paid \$200K to study sites in Chula Vista	2.0	0.09	1.90	0.01
E7: Oceanside council set aside \$100K for consultants	2.0	1.80	0.19	0.01

Table 1: ACH Matrix for Stadium Example with Extensions

	Mean	Interval	p_{max}
OC	0.442	(0.156, 0.726)	0.535
CV	0.388	(0.120, 0.675)	0.406
Other	0.170	(0.004, 0.397)	0.059

Table 2: Results for Extensions Example



Figure 1: Marginal Posterior Distributions for **p** and Proposal Densities

References

[1] Saul Blumenthal, Multinomial sampling with partially categorized data, Journal of the American Statistical Association **63** (1968), no. 322, 542–551.

- [2] Bradley P. Carlin and Thomas A. Louis, Bayes and empirical bayes methods for data analysis, second ed., Champman and Hall/CRC, 2000.
- [3] Daniel F. Heitjan and Donald B. Rubin, Ignorability and coarse data, The Annals of Statistics 19 (1991), no. 4, 2244–2253.
- [4] Richards J. Heuer, Jr., Psychology of intelligence analysis, Center for the Study of Intelligence, Central Intelligence Agency, 1999.
- [5] Jessica McLaughlin, A Bayesian updating model for intelligence analysis: A case study of Iraq's nuclear weapons program, 2005.
- [6] Palo Alto Research Center, Ach : Version 2.0.3, 2006.
- [7] M. E. Paté-Cornell, Fusion of intelligence information: A Bayesian approach, Risk Analysis 22 (2002), no. 3, 445–454.
- [8] Brian D. Ripley, Stochastic simulation, J. Wiley, 1987.
- [9] Douglas L. Weed, Weight of evidence: A review of concept and methods, Risk Analysis 25 (2005), no. 6, 1545–1557.