Cause odf effects: an important evaluation in Forensic Science

Cause of effects: an important evaluation in Forensic Science

Cause degli Effetti: una rilevante valutazione nelle Scienze Forensi

Fabio Corradi and Monica Musio

Sommario *Causes of Effects* (COE) queries concern the assessment of causal relationship in individual cases by evaluating the *probability of causation.* However is not always clear how and whether, to usefully employ scientific data for this purpose. Given even a randomized sample we can typically only provide bounds for the *probability of causation* if some fundamental conditions, namely exogeneity, comparability and sufficiency [4] , are satisfied. In this work we make the fundamental conditions operative by means of a Bayesian model selection procedure.

Sommario *Le problematiche di Cause di Effetti (COE) riguardano la valutazione di relazioni causali individuali attraverso il calcolo della probabilità di causazione. In questo contesto non è sempre chiaro come utilizzare dati provienti da studi scientifici. Infatti anche in presenza di esperimenti randomizzati si possono solo calcolare degli intervalli per la probabilità di causazione, qualora siano soddisfatte le così dette condizioni fondamentali di esogeneità, comparabilità e sufficienza [4]. L'obiettivo di questo lavoro è di rendere operative queste condizioni tramite una procedura di selezione di modelli in ambito Bayesiano.*

Key words: Causes of Effects, Fundamental conditions, Bayesian Model Selection

1 Introduction

In causal inference it is important to distinguish two types of causal query which, although not entirely unconnected, are nevertheless very different, both in form and in the type of answer they require. The following example traces back to Holland,

Fabio Corradi
Dipartimento di Statistica, Informatica e Applicazioni, Università di Firenze, e-mail: corradi@disia.unifi.it

Monica Musio
Dipartimento di Matematica ed Informatica, Università di Cagliari e-mail: mmusio@unica.it

1986 [6] and is archetypical in causal literature.
Effects of Causes (EoC): " Ann has a headache. She is wondering whether to take aspirin. Will that cause her headache to disappear?"
Causes of Effects (CoE): "Ann had a headache and took aspirin. Her headache went away. Was that caused by the aspirin?"
EoC type questions have the form "Will A cause B ?" CoE questions have the form "Did A cause B ?", and are also known as problems of individual causation.

How to use scientific data on individual cases is a problem that is central to the courts of justice and, in forensic literature, is referred as the G2i ("Group-to-individual") problem Dawid *et al.*, (2014) [3]).
In the CoE case, Ann actually chose to take the aspirin ($E = 1$) and assumed the drug ($T = 1$), this produces her headache to disappear ($R = 1$).

In this work we concentrate on the general CoE setting and more explicitly on how to support the conditions required to evaluate the probability of causation.

2 Basic of COE

Let's consider again the CoE example about the aspirin: Ann had a headache and decided ($E = 1$) to took aspirin ($T = 1$). Her headache went away ($R = 1$). Was that caused by the aspirin? One possible answer, "Probability of Causation", PC, relies on potential responses (R_0, R_1) (where R_x denotes the value R takes when $T = x$, *i.e.* T is set to x). We know that, for Ann, $T = 1$ so that $R_1 = 1$: she took the aspirin and her headache disappeared. Now if in "counterfact" $R_0 = 1$, then Ann's headache would have disappeared even if she had not taken the aspirin, so we must conclude that it was not the aspirin that cured her. Conversely, if $R_0 = 0$ then we can indeed attribute her cure to having taken the aspirin. Everything must be also evaluated considering her choice to take the drug, $E = 1$, which could be informative of her health status or, in general, on her preference to assume the treatment. In this way our CoE causal question is formulated in terms of the contrast between the factual outcome R_1 and the counterfactual outcome R_0. Formally we can define the PC as the *conditional probability* (see, *e.g.*, Dawid *et al*., (2016) [4])

$$\mathrm{PC}_A = \Pr(R_0 = 0 \mid H^A, E = 1, R_1 = 1) = \frac{\Pr(R_0 = 0, R_1 = 1 \mid H^A, E = 1)}{\Pr(R_1 = 1 \mid H^A, E = 1)}, \quad (1)$$

where PC_A denotes the judge's probability distribution over attributes of Ann and H^A the background information, some relevant known information about Ann. We denote by H a set of variables $H = \{H_1, \ldots, H_K\}$ and with $H^A = \{H_1^A, \ldots, H_K^A\}$ their corresponding values for Ann. This formal approach does however leave open the questions of how to evaluate PC and what evidence can be used. The numerator of (1) is not estimable, since we can never observe both R_0 and R_1 for the same individual, hence we can never assess this dependence without making any further

assumptions. CoE questions may simply not have well-determined answers, but we can sometimes set bounds so long as some *fundamental conditions* are satisfied (Dawid *et al*., (2016) [4]). In such case it is possible to show that

$$\Pr(R_0 = 0 \mid H^A, E = 1, R_1 = 1) \geq \max\{0, 1 - \frac{1}{RR_A}\} \tag{2}$$

with

$$RR_A := \frac{\Pr(R_1 = 1 \mid H^A, E = 1)}{\Pr(R_0 = 1 \mid H^A, E = 1)}.$$

Whenever $RR_A > 2$ the Probability of Causation PC_A will exceed 50%. In a civil court this is often taken as the criterion to assign legal responsibility "on the balance of probabilities". How much information H^A to take into account is still an unsolved problem. The aim of this work is to evaluate how different specifications of H support the fundamental conditions. We pose the issue as a model selection problem within the class of models induced by the fundamental conditions, where each model specifies a particular choice for H. Models will be evaluated by considering their marginal likelihood and a prior on the model space.

3 Fundamental conditions

3.1 Exogeneity

The potential outcomes (R_0, R_1) have the same joint distribution among both treated and untreated study subjects having the same background information H^A as Ann.
This condition essentially assumes no confounding $(R_0, R_1) \perp\!\!\!\perp T \mid H^A$ for every specification of H. This assumption cannot be tested empirically since in the treated patients we only observe R_1 and in the untreated patients we only observe R_0. So any argument we make for exogeneity has to be justified because of the randomization and ignorance of the influence of the H characteristics on the response.

3.2 Comparability

Conditional on knowledge of the pre-treatment characteristics of Ann and the trial subjects we can regard Ann's potential responses as comparable with those of the treated subjects having characteristics H^A.
Comparability essentially says that we are not able to distinguish between Ann and the group of treated individuals as it concerns the uncertainty of their response to the treatment. This is nothing but exchangeability, a basic form of dependence introduced by de Finetti. If exchangeability holds, the joint distribution of the random variables considered for a problem are invariant to permutation, i.e. there is no way

to make a distinction among them. According to what characteristics are included in H, different individuals in the randomized sample will be compared with Ann. Then using the representation theorem we can evaluate the probability to observe Ann and the group of responses if exchangeability holds among them.

3.3 Sufficiency

Conditional on H^A, Ann's intention to take (or not) the treatment does not affect the distribution of her potential responses.
This condition requires that, given Ann's own background information H^A, her potential outcomes are not further influenced by her (known) desire to take the aspirin. We need first too address the condition required to evaluate the denominator of RR_A: *Conditional on knowledge of the pre-treatment characteristics of Ann and the trial subjects, we can regard Ann's potential responses as comparable with those of the* untreated *subjects having characteristics H^A*. This potential probability clearly concerns a counterfactual event, since Ann decided and actually took the aspirin; obviously we don't know the response if she had not taken the aspirin but we know her will to assume the drug. We assume that E is observed in the sample: for example, a patient entered into a randomized trial is conscious that he may not receive the treatment but only a placebo and nevertheless agrees to express the wish to take aspirin or not. The ambition is to evaluate $\Pr(R_0 = 1 \mid H^A, E = 1)$ using traditional experimental data for $T = 0$, without ignoring the decision of Ann to take the drug, which is *observational* in nature. $E = 1$ describes the Ann's desire to take the aspirin, information not included in H^A but possibly relevant to determine the probability of the response. If we can obtain reasonable support to the condition

$$R_0 \perp\!\!\!\perp E \mid H^A \qquad \text{i.e. if} \tag{3}$$

$$\Pr(R_0 = 1 \mid H^A, E = 1) = \Pr(R_0 = 1 \mid H^A, E = 0) \tag{4}$$

it would be possible to estimate $\Pr(R_0 = 1 \mid H^A, E = 1)$ by $\Pr(R = 1 \mid H^A, T = 0)$.

4 Model selection

We pose the problem of finding the group most fitting the fundamental conditions as one of model selection solved, as usual, by computing the marginal likelihood, based on the data, for each possible specifications of H. We have 2^K possible different choices for the characteristics that can be selected from H. Let J be one of these choices, that can be identified as a subset of $\{1, \ldots, K\}$. We introduce an equivalence relation $\sim$ on the set of treated individuals that agree to receive the treatment ($T = 1, E = 1$). Namely if I_1 and I_2 are two such individuals then

$$I_1 \sim I_2 \iff H_j^{I_1} = H_j^{I_2}, \quad \forall j \in J.$$

A similar definition can be given for the other subgroups ($T = 1, E = 0$), ($T = 0, E = 1$) and ($T = 0, E = 0$). Thus each equivalence relation determines a partition of one of the above four groups. These four partitions identify a *model* M_J. We denote each element of one quotient set by $X_{t,e}^s$, where $t \in \{0,1\}$, $e \in \{0,1\}$ and $s \in \{0,1,2\ldots,N\}$, $N-1$ being the number of elements of each partition. We reserve the notation $X_{t,e}^0$ to the sets of individuals who share all the characteristics with Ann. Using the same notation as before, we indicate by $\mathbf{r}_{t,e}$ the vector of responses in the main 4 groups, by $\mathbf{r}_{\{\setminus X\}}$ the vector of responses in the complementary of the set X, by $\mathbf{r}_{t,e}^s$ the vector of responses referred to a specific group of individuals and by r_A the Ann's response. The responses of the treated and untreated individuals are not considered exchangeable each other and are modeled separately using the conditional representation indexed by their own θs, which are a priori assumed independent. So the marginal likelihood we want to evaluate factorizes and we evaluate each contribution separately.

$$\begin{aligned}
&\Pr(r_A, \mathbf{r}_{1,1}, \mathbf{r}_{1,0}\mathbf{r}_{0,1}, \mathbf{r}_{0,0}|M_J) = \qquad (5)\\
&= \int_{\Theta_{T=1}} \int_{\Theta_{T=0}} \Pr(r_A, \mathbf{r}_{1,1}, \mathbf{r}_{1,0}\mathbf{r}_{0,1}, \mathbf{r}_{0,0}, \theta_{T=1}, \theta_{T=0}|M_J) d\theta_{T=1} d\theta_{T=0}\\
&= \int_{\Theta_{T=1}} \Pr(r_A, \mathbf{r}_{1,1}, \mathbf{r}_{1,0}, \theta_{T=1}|M_J) d\theta_{T=1} \int_{\Theta_{T=0}} \Pr(\mathbf{r}_{0,1}, \mathbf{r}_{0,0}, \theta_{T=0}|M_J) d\theta_{T=0}.
\end{aligned}$$

4.1 Marginal likelihood for comparability

Because of the sufficiency condition, we assume the same mixing parameter $\theta_{T=1}^s$ in the sets $X_{1,1}^s$ and $X_{1,0}^s$, whose prior is assumed a non-informative $Be(1,1)$, while for the comparability condition Ann is suppose to be exchengeable with the treated individuals sharing with Ann the same characteristics. The marginal likelihood is

$$\begin{aligned}
&\Pr(r_A, \mathbf{r}_{1,1}, \mathbf{r}_{1,0}|M_J) = \Pr\left(r_A, \mathbf{r}_{1,1}^0, \mathbf{r}_{1,0}^0, \mathbf{r}_{\setminus\{X_{1,1}^0, X_{1,0}^0\}}\right) \qquad (6)\\
&= \int_{\Theta_{T=1}^0} \Pr(r_A, \mathbf{r}_{1,1}^0, \mathbf{r}_{1,0}^0, \theta_{T=1}^0) d\theta_{T=1}^0 \cdot \prod_{s\neq 0} \int_{\Theta_{T=1}^s} \Pr(\mathbf{r}_{1,e}^s, \theta_{T=1}^s) d\theta_{T=1}^s =\\
&= \frac{x_{1,e}^0 + 1}{n_{1,e}^0 + 2} \prod_{s\neq 0} \frac{1}{n_{1,e}^s + 1}
\end{aligned}$$

where $x_{1,e}^0$ is the number of success in the group $X_{1,1}^0 \cup X_{1,0}^0$, $n_{1,e}^0 = |X_{1,1}^0| + |X_{1,0}^0|$ and $n_{1,e}^s = |X_{1,1}^s| + |X_{1,0}^s|$.

4.2 Marginal likelihood for sufficiency

The sufficiency assumption requires E to have not influence on the response for the untreated with the Ann's characteristics. So, if $R_0 \perp\!\!\!\perp E|H^A$ holds, then the two groups of r.v., $\mathbf{r}^0_{0,1}, \mathbf{r}^0_{0,0}$, share a common mixing distribution parameter $\theta^0_{T=0}$, whose prior is assumed a non-informative $Be(1,1)$. We denote by $\theta^s_{T=0,E=e}$ the mixing parameter in the quotient set $X^s_{0,e}$. The marginal likelihood is:

$$\Pr(\mathbf{r}_{0,1}, \mathbf{r}_{0,0}|M_J) = \Pr(\mathbf{r}^0_{0,1}, \mathbf{r}^0_{0,0}, \mathbf{r}_{\setminus\{X^0_{0,1}, X^0_{0,0}\}}) = \int_{\Theta^0_{T=0}} \Pr(\mathbf{r}^0_{0,1}, \mathbf{r}^0_{0,0}, \theta^0_{T=0}) d\theta^0_{T=0} \cdot$$

$$\prod_{s\neq 0} \prod_{e\in\{0,1\}} \int_{\Theta^s_{T=0,E=e}} \Pr(\mathbf{r}^s_{0,e}, \theta^s_{T=0,E=e}) d\theta^s_{T=0,E=e}$$

$$= \binom{n^0_{0,0}}{x^0_{0,0}} \binom{n^0_{0,1}}{x^0_{0,1}} \binom{n^0_{0,0}+n^0_{0,1}}{x^0_{0,0}+x^0_{0,1}}^{-1} \cdot \frac{1}{n^0_{0,0}+n^0_{0,1}+1} \prod_{s\neq 0} \prod_{e\in\{0,1\}} \frac{1}{n^s_{0,e}+1} \tag{7}$$

where we have used a similar notation as before.

Remark 1 (Marginal likelihood and the Irving-Fisher exact test). The marginal likelihood evaluated for a subset of H formally shares the hypergeometric part with the conditional Irving-Fisher exact test statistic used to evaluate differences among the rate of success of an event in two populations. The result is not surprising since we are looking for the set of H making E irrelevant, so supporting the so called H_0 hypothesis of no-difference between the success ratio in the two groups.

4.3 Prior and posterior in the model space

By (5), the required marginal likelihood is the product of (6) and (7). The goal is to evaluate the posterior of M_J given the responses observed on Ann and on the sample. To this end we introduce a prior over the space of models. The simpler choice is to consider an uniform distribution on this space. Another choice is this one proposed by Chen and Chen, (2008) [1]. They give the same prior probability (equal to $\frac{1}{k+1}$) to all models sharing the same number of characteristics k. In this way for the generic model M_J we have:

$$\Pr(M_J) = \frac{1}{k+1} \binom{k}{|M_J|}^{-1} \cdot I(|M_J| \leq k/2) \tag{8}$$

where the search spans all models including at most $k/2$ characteristics. This choice favours model selection according to the Occam razor principle: the fewer characteristics employed, the more probable is the model. This rationale is reasonably objective. Combining (6) and (7) and (8), we get the required posterior.

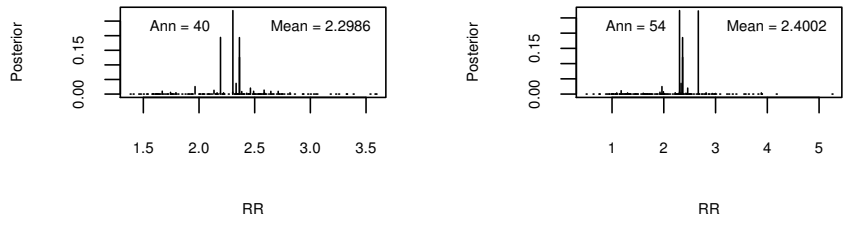

Figura 1 Risk ratios for two individuals who succeded and required the hint

5 An experiment

We carried on an experiment at the University of Florence, School of Engineering, Fall 2017. We asked to 160 students to solve a simple probabilistic question and we provided randomly an hint (the treatment T). Before the test we asked to the students if they wish to be helped or not ($E = 0$ and $E = 1$).
We are interested to investigate if there is a causal relationship between the hint and the ability to solve the question, for the students who wished to take the hint. We had 8 of these cases and for all of them we found a $RR > 2$ (obtained by averaging the results of different models according to their posterior probability, see figure (1)). This implies a lower bound for the probability of causation greater than 0.5 which suffices to indicate a causal relation between the hint and the positive result obtained by the 8 students.

As a result of our experiment we have also that, among the students who asked and had the hint, 24 did not succeed. We can image that one of them claimed that it was the hint which caused the failure. If we look now to the corresponding RR for such students, obtained as before by model averaging, we note that all of them have a values smaller then 1 (see figure 2). This is not conclusive that there is a causal relation between the hint and the failure.

In a civil trial this would not suggest to a judge to provide a compensation.

6 Conclusions

We introduced a typical Cause of Effect problem by means of an archetyppical example considering Ann and the effect of an aspirin on her headache. We have proposed a possible solution to make operational the choice of variables to include, so as to validate the fundamental assumptions underlying the assessment of Ann's probability of causation. We assume to have the possibility to perform a randomi-

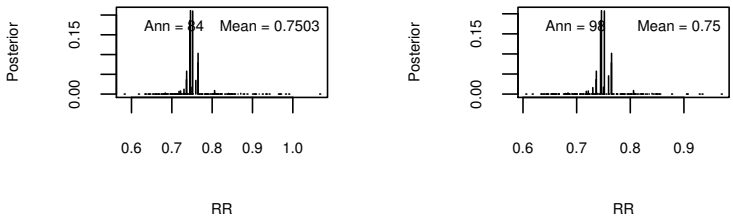

Figura 2 Risk ratios for two individuals who didn't succed and required the hint

zed sample from the Ann's population where, as usual, T is assigned following a randomized protocol and E, this is a novelty, is a question asking to the members of the sample about their preference to be treated or not.

Next step will be to extend the methodology to observational studies, to make possible in a wider range of cases the evaluation of the PC_A for Causes of Effects problems.

Acknowledgements The second author was supported by the project GESTA of the Fondazione di Sardegna and Regione Autonoma di Sardegna

Riferimenti bibliografici

1. Chen, J. and Chen, Z., Extended Bayesian Information Criteria for Model Selection with large Model Spaces. *Biometrika*, **95**, 3, 759-771, 2008.
2. Dawid, A. P., The role of scientific and statistical evidence in assessing causality. In *Perspectives on Causation*, (ed. R. Goldberg), 133-147. Hart Publishing, Oxford, 2011.
3. Dawid, A. P., Faigman, D. L., and Fienberg, S. E., Fitting science into legal contexts: Assessing effects of causes or causes of effects? (with Discussion and authors' rejoinder). *Sociological Methods and Research*, **43**, 359–421, 2014.
4. Dawid, A. P., Musio, M. and Fienberg, S. E., From Statistical Evidence to Evidence of Causality. *Bayesian Analysis*, **11**, 725-752, 2016.
5. Dawid, A. P., Musio, M. and Murtas, R., The Probability of Causation, *Law, Probability and Risk*, **16**, 4, 163-179, 2017.
6. Holland, P. W., Statistics and Causal Inference *Journal of the American Statistical Association*, **81**, 396, 945-960, 1986.
7. Rubin, D. B., Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, **66**, 688–701, 1974.