Graphical Models for Surrogates

Steffen L. Lauritzen
Aalborg University, Department of Mathematical Sciences
Fredrik Bajers Vej 7G
DK-9220 Aalborg, Denmark
steffen@math.auc.dk

Recently, it has been demonstrated that graphical models promise some potential for expressing causal concepts, see for example Pearl (2000), Lauritzen (2001), or Dawid (2002). The causal interpretation is most direct in models based on directed acyclic graphs, whereas causal interpretation for chain graph models generally is more subtle and complex (Lauritzen and Richardson 2002).

In the articles cited, such concepts as confounding, partial compliance, causal sufficiency of covariates, and prediction of treatment effects were discussed and illuminated.

In this article we will use graphical models to illustrate and analyse the notion of a *surrogate outcome*, such as also discussed e.g. in Frangakis and Rubin (2002).

Causal Markov models

The basic causal interpretation of a directed graphical model is as follows. We consider a directed acyclic graph (DAG) where nodes V of $\mathcal{D}$ represent random variables. The DAG is said to be *causal* for a probability distribution P with respect to a subset of variables $B \subseteq V$, if it holds for all $A \subseteq B$ that

$$p(x \,||\, x_A^*) = \prod_{\alpha \in V \setminus A} p(x_\alpha \,|\, x_{\mathrm{pa}(\alpha)}) \Big|_{x_A = x_A^*} . \tag{1}$$

Here $p(x \,||\, y) = P(X = x \,|\, Y \leftarrow y)$ refers to the distribution of x after an *intervention* which gives Y the specific value y. Generally $p(x \,||\, y) \neq p(x \,|\, y)$, where the latter is the more conventional conditional distribution, obtained as $p(x \,|\, y) = p(x, y)/p(y)$. Generally we distinguish between *intervention conditioning* and *observation conditioning*.

Note that the causal Markov property specifies a relation between *different* probability measures, each representing the law associated with a specific intervention. Thus, in leftmost of the causal DAGs below

we have that $p(y \,||\, x) = p(y \,|\, x)$ and $p(x \,||\, y) = p(x)$, whereas these relations are reversed in the rightmost graph, i.e. there it holds that $p(y \,||\, x) = p(y)$ and $p(x \,||\, y) = p(x \,|\, y)$.

The formula (1) has appeared in various forms in Pearl (1993) and Spirtes, Glymour and Scheines (1993). It is implicit in Robins (1986) and in other literature.

The causal Markov property of the DAG gives an easy way of reading off important conditional independence properties directly from the DAG itself. We refer e.g. to Dawid (2002) or Lauritzen (2001) for further details.

Surrogate outcomes

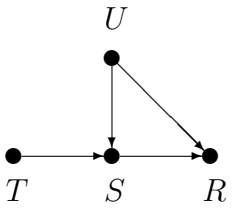
In certain therapeutic trials it may be difficult or practically infeasible to measure a response R of primary interest, for example if this is survival time for an individual. It may however be possible at an early stage to measure a post-treatment variable S which is known to predict the final outcome R well. Such an intermediate response S is said to be a *surrogate outcome* for the effect of T on R.

However, there is some ambiguity about the precise properties that S should have to justify it to be considered a surrogate. Prentice (1989) suggests a relevant property to be the independence of response and treatment, conditionally on surrogate outcome, i.e. S is a *statistical surrogate* if the relation $R \perp\!\!\!\perp T \mid S$ holds.

However, as argued by Frangakis and Rubin (2002), using a statistical surrogate for treatment evaluation can lead to serious and undesirable biases and misinterpretation of experimental results in the presence of potential confounding variables. In particular a statistical surrogate does not obey what the latter authors denote *causal necessity*. It may happen that the treatment can have no effect on S and still have a proper effect on the response R. Essentially this is a consequence of differences between conditioning by intervention and conditioning by observation, as mentioned earlier in this article.

Clearly, for a surrogate S to be useful it must also predict the response R well. This aspect is not discussed in this article.

The causal graphical model formulation leads naturally to the notion of a surrogate in a simple, randomised clinical trial through the diagram below:



where T is treatment, R is response, S is intermediate response, and U represents possible unmeasured confounding variables. The missing edge between U and T reflects the independence $T \perp\!\!\!\perp U$ ensured by allocating treatments T at random to individuals.

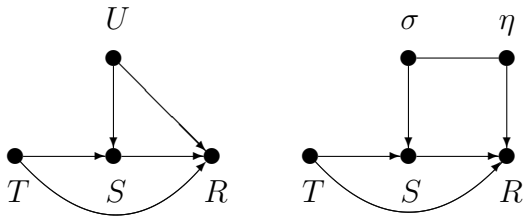
The missing edge between T and R reflects the property that the entire effect of T on R is mediated through the surrogate S, represented by the conditional independence $R \perp\!\!\!\perp T \mid (S, U)$. The latter allows the relationship between S and R to be modified by unknown confounders U.

We say that such a variable S is a *strong surrogate* for the effect of T on R. Note that a statistical surrogate is typically not a strong surrogate and vice versa. The diagram above does *not* imply $R \perp\!\!\!\perp T \mid S$.

Note also that a strong surrogate S also satisfies the notion of causal necessity. No effect of T on S would correspond to a missing arrow between T and S and this will automatically yield that T has no effect on R either.

Principal surrogates

Frangakis and Rubin (2002) define the notion of a *principal surrogate* using the method of potential responses (Neyman 1923; Rubin 1974). To describe this notion in terms of graphical models we need to consider *mapping variables* (Heckerman and Shachter 1995) — a variant of potential responses. The mapping variables describe deterministically how the treatment affects the responses in the coarsest possible way, as illustrated in the rightmost of the diagrams below, where the diagram to the left represents an unspecified and unobserved confounder U. Dawid (2002) uses the term 'canonical functional model' for the diagram involving mapping variables.



Here σ is a map which identifies how T affects S: $S(T, \sigma) = \sigma(T)$ and similarly η identifies the effect of the response as $R(T, S, \eta) = \eta(T, S)$. The mapping variables σ, η are describe completely how U moderates the treatment effects, so that further knowledge about U is irrelevant.

To illustrate the idea of a mapping variable, consider the situation in which S and T each have two levels, $T \in \mathcal{T} = \{1, 2\}$ and $S \in \mathcal{S} = \{L, H\}$. The table below gives the values of S for different values of T, where we have in mind that S represents low or high level of a substance which can be easily measured. There are four possible maps from $\mathcal{S}$ to $\mathcal{T}$:

T	$\sigma =$ sicker	$\sigma =$ normal	$\sigma =$ healthier	$\sigma =$ special
1	L	L	H	H
2	L	H	H	L

The mapping variable $\sigma =$ normal is representing the case where the second treatment raises the level of S, etc.

In a similar way there are eight possible maps from $\mathcal{T} \times \mathcal{S}$ to $\mathcal{R}$ if also the response is binary.

Frangakis and Rubin (2002) define the notion of a *principal surrogate* in a way which in our formulation would be implied by the condition $R \perp\!\!\!\perp T \,|\, (S, \sigma)$, since the mapping variable σ defines what they term *principal strata*. The authors in fact only demand this conditional independence to hold for such σ which are constant maps, i.e. for $\sigma =$ sicker and $\sigma =$ healthier in the example above.

This is thus considerably weaker than the graphical condition on a strong surrogate, which demands absence of the directed link from T to S. Thus, *a strong surrogate is a principal surrogate,* but not vice versa.

Identifying strong surrogates

In general it can be a difficult task to very that a post-treatment variable indeed satisfies the conditions needed for it to be a strong or principal surrogate, as both notions involve unobserved or even unobservable variables.

However, the diagram defining a variable to have the property of a strong surrogate is identical to a similar diagram describing the notion of *instrumental* variables. More precisely, it holds that *S is a strong surrogate for the effect of T on R if and only if T is an instrumental variable for the effect of S on R.*

It thus follows that the property of S being a strong surrogate for the effect of T on R can be *falsified* in principle from observation of (T, S, R) only, if S is a discrete variable. This is true because the instrumental inequality (Pearl 1995) implies that if S is a strong surrogate we must have

$$\max_s \sum_r \max_t p(r, s \,|\, t) \leq 1,$$

and this condition is in principle falsifiable from empirical observations.

REFERENCES

Dawid, A. P. (2002). Influence diagrams for causal modelling and inference. International Statistical Review, **70**, 161–89.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. Biometrics, **58**, 21–9.

Heckerman, D. and Shachter, R. (1995). Decision-theoretic foundations for causal reasoning. Journal of Artificial Intelligence Research, **3**, 405–30.

Lauritzen, S. L. (2001). Causal inference from graphical models. In Complex stochastic systems, (ed. O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg), pp. 63–107. Chapman and Hall/CRC Press, London/Boca Raton.

Lauritzen, S. L. and Richardson, T. S. (2002). Chain graph models and their causal interpretation (with discussion). Journal of the Royal Statistical Society, Series B, **64**, 321–61.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. In Polish. English translation of Section 9 by D. Dabrowska and T. P. Speed in *Statistical Science* **5** (1990), 465–480.

Pearl, J. (1993). Graphical models, causality and intervention. Statistical Science, **8**, 266–9. Comment to Spiegelhalter et al. (1993).

Pearl, J. (1995). Causal inference from indirect experiments. Artificial Intelligence in Medicine, **7**, 561–82.

Pearl, J. (2000). Causality: Models, reasoning, and inference. Cambridge University Press, Cambridge, UK.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials. Statistics in Medicine, **8**, 431–40.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods — application to control of the healthy worker survivor effect. Mathematical Modelling, **7**, 1393–512.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. Journal of Educational Psychology, **66**, 688–701.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., and Cowell, R. G. (1993). Bayesian analysis in expert systems (with discussion). Statistical Science, **8**, 219–83.

Spirtes, P., Glymour, C., and Scheines, R. (1993). Causation, prediction and search. Springer-Verlag, New York. Reprinted by MIT Press.

RÉSUMÉ

On considére la representation des notions causales par des modèles graphiques probabilistes. On discute la representation des variables qui sont capable de prévoir le comportement des resultats finals d'une experiment: les variables surrogats.