Towards Argumentation for Statistical Model Selection

Isabel SASSOON [a,1], Jeroen KEPPENS [a] Peter MCBURNEY [a]
[a] *King's College London, UK*

Abstract. The increase in routine clinical data collection coupled with an expectation to exploit this in support of evidence based decision making creates the requirement for a system to support clinicians in this analysis. This paper looks at applying argumentation to this problem, by collating all the relevant statistical approaches and their assumptions into a statistical knowledge base and then representing the model selection process through argumentation. This will form the foundation for the development of a prototype that will enable clinicians to answer their research questions with no statistics, informatics or administrative support.

Keywords. application of argumentation, automated statistical analysis, statistical model selection

1. Introduction

Routine clinical data collection is resulting in the proliferation of databases and the expectation to exploit it in support of evidence based decision making is a growing trend. In order to leverage clinical data in support of decision making, it is essential to analyse it with rigour. This rigour is based on using the most appropriate statistical assumptions and methods given the analysis objective and available data. The appropriate statistical method supports evidence based decision making by validating the clinician's research question or hypothesis by providing confidence in the conclusion. It is difficult for a clinician to choose an appropriate statistical method and the choice is not always straight forward, even for a statistician. The considerations as to what model to use depend not just on the clinician's research question and data but may also depend on background information from the clinician, and may vary from model to model. Easy to use statistical packages make the analysis easy to use but not the choice of model.

The aim of this work is to develop an intelligent model selection system to solve this problem by suggesting appropriate model(s) to the clinician based on the research question, the data and any external relevant input and to support these with arguments.This paper and work leverages a head and neck cancer database and analysis requirements. The output is a method to support the model selection process and facilitate analysis.

[1]Corresponding Author: Isabel Sassoon, Dept. of Informatics, King's College London, Strand, London WC2R 2LS, UK; E-mail: isabel.sassoon@kcl.ac.uk.

2. Background

Although the work presented herein can be applied more broadly to statistical analysis in various domains, it is motivated by the challenges that arise when clinicians wish to answer a research question using case data collected in their day-to-day practice.

Clinicians interact with statistical concepts at the design stage of a study, when selecting the models to use to analyse the data and when performing and interpreting the analysis. This paper focuses on the selection of the model as we are concentrating on analysing existing data rather than designing how the data is collected. Clinicians may not always be qualified in performing the analysis required in support of their research question and as such would involve a statistician. The statistician's role is to understand the data in the context of the research question and recommend the statistical analysis best suited to provide the results required.

Often, in empirical analyses of clinical data, models are chosen poorly or cannot be justified. A recent systematic review by [1] highlighted that reporting of survival analysis results, one important type of empirical analysis of clinical data, had increased within journal publications, however the quality of the reporting of the statistical analysis was improving slowly. The authors highlighted examples of this lack of quality such as failure to describe the study's follow up period, no clear definition of the event of interest and lack of significance testing. More pertinent to the aim of this paper is that the authors found a low proportion of articles that mention validation of model assumptions prior to use (proportional hazards assumptions for Cox modelling as a specific example). This paper will illustrate the proposed approach by means of survival analysis [7].

In some situations more than one analysis technique is appropriate to answer the specific research question, and opinions may vary as to what is the 'best' method to use. This depends on a range of preferences including the purpose of the model, desired outcomes (or example individual predictions vs. comparison between groups) and assumptions that are external to the data. Furthermore once a decision is made as to what model is used, this needs to be documented and justified. Previous work does not address this.

Some fundamental requirements from any method employed in statistical model selection are ability to deal with conflicting conclusions, ability to handle incomplete information and the facility to provide justification for the resulting recommendation. Argumentation has been shown through its use in decision support to accommodate these requirements. An additional desired feature is the ability to be driven by a separate knowledge base, to make the system flexible and expandable.

3. Method

The proposed approach is split into two parts (i) a knowledge base that contains all the statistical model definitions, objectives and assumptions; (ii) argumentation schemes to guide the model selection process. The knowledge base specifies how statistical models can achieve research objectives given certain assumptions. The knowledge base is used to instantiate the argumentation schemes. This split ensures that the knowledge base can be expanded and modified independently of the argumentation schemes.

In this paper we use the concept of a research objective to differentiate between different 'families' of analysis. One analysis objective (survival analysis) will contain all the models that are relevant to analyse time to event data. Another possible analysis objective would contain the models required for categorical outcome variable analysis.

Some similarities can be found between an argument for practical reasoning and our model selection problem. Practical reasoning is about what is sensible for someone to do given a situation, in our case the aim is deciding what model to use given the data and the circumstances at hand. In [2] the authors suggest that practical arguments similar to this one are best handled using argumentation schemes and associated critical questions. In this paper we have taken a similar approach.

3.1. Statistical knowledge base

The statistical knowledge base (SKB) consists of a set of objectives $O = \{o_1,\ldots,o_u\}$, a set of models $M = \{m_1,\ldots,m_v\}$ and a set of assumptions $A = \{a_1,\ldots,a_w\}$. Each objective corresponds to a type of research question that can be answered by means of a statistical analysis technique. Each model corresponds to a statistical analysis technique that can be employed to achieve an objective (i.e. answer a research question). Each assumption corresponds to a condition that ought to be met to some extent when employing a model. The objective of the system presented herein is to identify what model or models can be employed to achieve an objective by judging to how well a model's assumptions apply to the circumstances at hand.

Each model can be employed to achieve one or more objectives and each objective can be achieved by one or more models. Let $R_{OM} : O \times M$ be a relationship such that $(o_i,m_j) \in R_{OM}$ implies that objective o_i can be achieved by means of model m_j. Each model is associated with a set of assumptions that must be met adequately if a model is applied, irrespective of what objective that model is applied to. This is represented in the SKB by a relationship $R_{MA} : M \times A$ such that $(m_i,a_j) \in R_{MA}$ implies that assumption a_j must hold in a problem if model m_i is to be applied. Let $A(m_i) = \{a_j|(m_i,a_j) \in R_{MA}\}$ denote the set of assumptions of m_i.

It is not always the case that there exists a model that enables the achievement of an objective, such that all that model's assumptions hold in the circumstances at hand. In that case, it will be necessary to apply a model with some assumption violations. An assumption a_j is said to be critical to a model m_i if m_i must not be applied under any circumstances if a_j does not hold. Let $C \subset R_{MA}$ be the set of all model assumption pairs (m_i,a_j) such that a_j is critical to m_i, and let $A_c(m_i) = \{a_j|(m_i,a_j) \in C\}$ denote the set of critical assumptions of m_i.

Each assumption is either a specific property of the data set or a feature of the broader population of interest or the way in which the data set is collected from that population. The former type of assumption is normally assessed by applying a test on the data set that returns an assessment of the extent to which the data set satisfies that assumption. For example, the proportional hazards assumption is tested by running a model on the data with a time dependent covariate and the assumption is met if the time dependent covariate's coefficient is not significant.

Assessing the latter type of assumption normally relies on the judgement of a domain expert. For example, when the analysis objective is survival analysis there is a critical assumption of non informative censoring which in practice means that censoring occurs for reasons unrelated to the study. This may not be captured in the data and therefore relies on the clinician's judgment with regards to this potential correlation between censoring and survival time.

Therefore, the set of assumptions A is partitioned into a set of tests A_t and a set of queries A_q. Tests are assumptions that are assessed by applying a test on the available data set and queries are assumptions that are assessed by asking the clinician for an opinion. In what follows, let $A_t(m_i) = \{a_j | (m_i, a_j) \in A_t\}$ and $A_q(m_i) = \{a_j | (m_i, a_j) \in A_q\}$.

Table.1 shows the contents of the SKB for the analysis objective of survival analysis. The most popular methods are included, more complex cases at this stage, the system would recommend to consult a statistician to explore more sophisticated methods.

3.2. Schemes to generate arguments from the knowledge base

The SKB specifies how research objectives can be achieved by means of statistical models and under what assumptions. This statistical knowledge is applied to form argumentations by means of generic argumentation schemes. The process starts with the clinician selecting their analysis objective o_c. A number models m_i can achieve this objective provided their critical assumptions are met. Such arguments can be instantiated through AS1.

AS1: *Argument for a Possible Model*

- Model m_i achieves objective o_c
- The data set meets the set of assumptions $A'_t = A_t(m_i)$
- The research project meets the set of assumptions $A'_q = A_q(m_i)$
- $A_c(m_i) \subseteq A'_t \cup A'_q$

$\therefore$ m_i is a possible model

Assumptions that are validated through a query to the clinician, as they cannot be validated through data can be seen as arguments from expert opinion and human judgement will apply to this. Instantiating AS1 may identify one or more possible models. If there is more than one possible model, it is necessary to choose which one to apply. Depending on the aim of the research the characteristics of the preferred model may vary, in [9] the author discusses the different purposes of models. Statisticians and clinicians employ various valid arguments to justify that they prefer one model over another. Often, these arguments attack one another. A number of argumentation schemes are identified that define the reasons for preferring one model over another, so that these reasons and attack relationships can be modelled and reasoned with. All argumentation schemes describing a preference for one model over another are of the form of AS2.

AS2: *Argument for a Model Preference*

- m_i is a possible model
- m_j is a possible model
- there is a reason to prefer m_i over m_j $(*)$

$\therefore$ m_i is preferred over m_j

A number of versions of AS2 can be defined, each with a different reason to prefer one model over another by using the appropriate condition in $(*)$ in the generic definition provided in AS2. There may be a preference for models that violate fewer assumptions. Specifically, if the assumptions that are violated by one model constitute a strict subset of those violated by another, then the former model is preferred to the latter, then $(*)$ in

AS2 can be substituted by the following set of conditions:

- $A_i^{\perp} = [A_t(m_i) \backslash A_t'] \cup [A_q(m_i) \backslash A_q']$
- $A_j^{\perp} = [A_t(m_j) \backslash A_t'] \cup [A_q(m_j) \backslash A_q']$ *AS2(a)*
- $A_i^{\perp} \subset A_j^{\perp}$

A modification of the scheme above is possible where the assumptions are not nested. Therefore the criteria will be based on the number of assumptions violated without requiring them to be a subset. This would substitute $(*)$ with:

- $A_i^{\perp} = [A_t(m_i) \backslash A_t'] \cup [A_q(m_i) \backslash A_q']$
- $A_j^{\perp} = [A_t(m_j) \backslash A_t'] \cup [A_q(m_j) \backslash A_q']$ *AS2(b)*
- All violated assumptions in $A_i^{\perp} \cup A_j^{\perp}$ are of equal importance
- $n(A_i^{\perp}) < n(A_j^{\perp})$

There may be a preference for a more complex model that satisfies more assumptions. For example a parametric model, which has more validated assumptions will be preferred to a non parametric version in cases where accuracy in model estimates is key. In such cases $(*)$ will be substituted with:

- m_i satisfies the assumptions $A_i^{\top}$
- m_j satisfies the assumptions $A_j^{\top}$ *AS2(c)*
- $n(A_i^{\top}) > n(A_j^{\top})$

There may be scenarios where a possible model that requires fewer assumptions to be satisfied is preferred, this could be in cases where the desire is to robustly explore the data with a simpler model. In this case $(*)$ will be replaced by:

- $n[A_t(m_i) \cup A_q(m_i)] < n[A_t(m_j) \cup A_q(m_j)]$ *AS2(d)*

The clinician may have a belief that model m_j is the model to be applied to their research question. Typically, this is due to m_j being used in the vast majority of relevant past analysis work or publications. If the resulting list of possible models includes m_j and other models then there needs to be a good enough reason for the clinician not to run m_j and to employ an alternative model from those that are possible. In this case $(*)$ will be replaced by:

- Most peer reviewed research concerning objective o_c employs model m_i
- There are no strong reasons not to use m_i
- If most peer reviewed research concerning objective o_c employs model m_i and there are no strong reasons not to use m_i, m_i ought to be preferred over other suitable models *AS2(e)*

Further details for AS2 are available on [2]. An alternative approach to model preference could be through the use of a measure of value, and there are examples of this approach [3]. As this paper focuses on cases where there is no numeric value to be attributed to the differing outcomes that result from different model choices, a preference based approach will be tested in future work.

4. Case Study

This case study is based on analysis work undertaken on a head and neck cancer database aimed at ascertaining the safety of a new diagnostic procedure. This has been submitted for publication in the Journal of Clinical Oncology. We will be using a concise example

[2] http://www.dcs.kcl.ac.uk/pg/isassoon/COMMA2014

of the type of analysis objective assessed in this study in order to illustrate the application of the proposed method.

The clinician choses the following research question from the drop down, this sets the objective of the analysis to o_c. The research question in this context could be "Is there a survival difference between patients with different levels of tumour differentiation?"

Assuming there are N patients, and the data available is (t_i, c_i, z_i) for $i = 1, ..., n$ where t_i is the follow up time or survival time for patient i, c_i is the event indicator and z_i is the covariate.

The next step is to retrieve all the model options from the SKB where $(o_c, m_j) \in R_{OM}$. In this case there are three potential models, m_1, m_2, m_3 can all achieve objective o_c. These are: m_1 Kaplan Meier, m_2 Cox Proportional Hazards and m_3 Weibull model [7]. Note that this is not an exhaustive list but this represents the more frequently used models in these situations, in order to demonstrate the proposed method.

The next step involves instantiating AS1 to ascertain which models are possible. In this example the details of the required queries and tests are in Table.1. The critical assumptions for the models are: $A_c(m_1) = a_1$, $A_c(m_2) = a_1, a_3$ and $A_c(m_3) = a_1, a_4$. The assumptions validation returned: a_1 is true, a_3 is true but a_4 is false. Instantiating AS1 for m_1 based on the assumption validation will result in the following:

- Model m_1 achieves objective o_c
- The data set meets the set of assumptions $A'_t = \emptyset$
- The research project meets the set of assumptions $A'_q = a_1$
- $A_c(m_1) A'_t \cup A'_q = A_c(m_1)$

$\therefore$ m_1 is possible

Applying AS1 to m_2 results in m_2 being a possible model. As a_4 is not TRUE then no argument for the claim that m_3 is a possible model can be produced. This leaves two possible models: m_1 and m_2. The instantiation of AS2 will output the preferred model. As discussed in Section 3 there are multiple ways of deciding the preferred model. Detailed instantiations of these are available on www.dcs.kcl.ac.uk/pg/isassoon/COMMA2014.

Instantiating AS2(a) and AS2(b) results in no preference between m_1 and m_2, AS2(c) results in preferring m_2, AS2(d) and AS2(e) both result in preferring m_1. In this case study the actual analysis was executed using m_1, this in agreement with the preferred model that resulted from both AS2(d) and AS2(e).

5. Discussion and related work

Our proposed method supports the clinician through the analysis process by recommending the most appropriate model and ruling out models that do not support the objective or have their assumptions violated. The use of argumentation schemes and the SKB in conjunction with the argumentation process facilitates the documentation of the model recommendation. The use of the SKB to store all the model definitions and assumptions as a separated module will enable the incorporation of additional models, assumptions and analysis objectives without affecting the argumentation mechanism for model selection. Our method is able to handle different considerations when selecting a preferred model from the possible models and this differentiates it from other approaches.

Model	Assumption	Critical	Test or Query
m_1 : Kaplan Meier	a_1 : non informative censoring	Yes	Query: Clinician to confirm
m_1 : Kaplan Meier	a_2 : heavy censoring	No	Test: Proportion censored = $\frac{N-\sum c_i}{N} > 70\%$
m_2 : Cox Proportional Hazards	a_1 : non informative censoring	Yes	Query: Clinician to confirm
m_2: Cox Proportional Hazards	a_3 : proportional hazards	Yes	Test: Run R function that calculates tests of the proportional-hazards assumption for each z_i, $\forall z_i : p-value > 0.05$
m_2: Cox Proportional Hazards	a_2 : heavy censoring	No	Test" Proportion censored = $\frac{N-\sum c_i}{N} > 70\%$
m_3: Weibull	a_4 : distributional assumption	Yes	Test: R script to test linear relation $log(-log(\hat{S(t)}))$ vs. $log(t)$ where $\hat{S(t)}$ is the survival estimate from the KM model (m_1)
m_3 : Weibull	a_1 : non informative censoring	Yes	Clinician to confirm
m_3: Weibull	a_2 : heavy censoring	No	Proportion censored = $\frac{N-\sum c_i}{N} > 70\%$

Table 1. Detailed SKB for o_c Survival Analysis including tests on the data and queries to the clinician

There are examples of applications of argumentation that are similar to the one in this paper. This paper looks at 'diagnosing' the appropriate model to apply to a research question, there are examples of implementations where argumentation looks at deciding on the best treatment for a patient [3]. In this paper the authors demonstrate how argumentation can assist in cases where there are multiple treatment options as well as value considerations. They also make use of a knowledge base. In [5], [6] the authors implement a system to flag abnormal reactions to medications using argumentation, this highlights the advantage of using argumentation in documenting the reasons for flagging a patient reaction. Another pertinent implementation [10] uses argumentation as part of the analysis process to ascertain whether external data can replace missing data, this also places argumentation as part of the analysis process.

The market for statistical analysis tools includes specialist tools for the clinician and the statistician however these offer little guidance on the overall model selection process. Some will recommend the best analysis based on the distributional assumptions of the data in isolation, whilst others will flag a break in the assumptions within the results outputs if it occurs. The application of expert systems to automate statistical analysis has not had major developments in the past years. A review paper [4] describes the range of tasks and the desired features of such a system, some are very relevant to our problem. These include the need for any such system to be able to explain itself, cater for user error, recommend the most powerful technique, adapt for data quality issues, incorporate new techniques and self document.

A recent paper shows some renewed interest in automation of statistical analysis, in [8] the authors present initial work on a project they term "Automatic Statistician". The approach and type of analysis tackled is different from the one our paper focuses on. The authors [8] focus on automating time series data analysis by exploring all possible modelling options before selecting the model that best explains the data. An additional emphasis of [8] is the natural language text of the resulting analysis report, this is an aspect that we are aiming to address as part of future phases.

Future work includes the expansion of the SKB to include weights on the different assumptions and the inclusion of model purpose and preferences when selecting the recommended model. The role of the critical questions in the context of the proposed argumentation schemes will be explored. We are planning to increase the scope of the SKB to include further methods within the survival analysis domain and additional objectives such as comparison of continuous/measurements as an outcome or classification. Additional areas for future work include the introduction of an ontology to support a more flexible research question input. This would enable the clinician to formulate their research question in more flexible and familiar terminology, as the ontology would relate it back to the key concepts required by the proposed method. A prototype implementation is also planned.

6. Acknowledgements

The authors would like to thank Mark McGurk for his generous support on this project.

References

[1] V. Abraira, A. Muriel, J. I. Emparanza, J. I. Pijoan, A. Royuela, M. N. Plana, A. Cano, I. Urreta, and J. Zamora. Reporting quality of survival analyses in medical journals still needs improvement. a minimal requirements proposal. *J Clin Epidemiol*, 66(12):1340–6.e5, Dec 2013.

[2] K. Atkinson, T. Bench-capon, and P. Mcburney. Justifying practical reasoning. In *Proceedings of the Fourth International Workshop on Computational Models of Natural Argument (CMNA 2004*, pages 87–90, 2004.

[3] K. Atkinson, T. J. M. Bench-Capon, and S. Modgil. Argumentation for decision support. In S. Bressan, J. Küng, and R. Wagner, editors, *DEXA*, volume 4080 of *Lecture Notes in Computer Science*, pages 822–831. Springer, 2006.

[4] M. Dami and A. Vella. Expert systems in statistics. *Univesity of Piraeus: Journal of Economics and Business*, 38(3):390–402, 1988.

[5] M. A. Grando, L. Moss, D. Glasspool, D. H. Sleeman, M. Sim, C. J. Gilhooly, and J. Kinsella. Argumentation-logic for explaining anomalous patient responses to treatments. In M. Peleg, N. Lavrac, and C. Combi, editors, *AIME*, volume 6747 of *Lecture Notes in Computer Science*, pages 35–44. Springer, 2011.

[6] M. A. Grando, L. Moss, D. H. Sleeman, and J. Kinsella. Argumentation-logic for creating and explaining medical hypotheses. *Artificial Intelligence in Medicine*, 58(1):1–13, 2013.

[7] J. Klein and M. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer, 2003.

[8] J. R. Lloyd, D. K. Duvenaud, R. B. Grosse, J. B. Tenenbaum, and Z. Ghahramani. Automatic construction and natural-language description of nonparametric regression models. *CoRR*, abs/1402.4304, 2014.

[9] P. McBurney. What are models for? In *Proceedings of the 9th European Conference on Multi-Agent Systems*, EUMAS'11, pages 175–188, Berlin, Heidelberg, 2012. Springer-Verlag.

[10] P. R. Quinlan, A. Thompson, and C. Reed. An analysis and hypothesis generation platform for heterogeneous cancer databases. In B. Verheij, S. Szeider, and S. Woltran, editors, *COMMA*, volume 245 of *Frontiers in Artificial Intelligence and Applications*, pages 59–70. IOS Press, 2012.