A New Approach to Regression Models for Binary Outcomes

Anders Huitfeldt[1], OpenMethodology Collaborators[2] and Other authors to be confirmed[3]

[1]Department of Epidemiology, Royal College of Surgeons in Ireland

April 13, 2020

Abstract

In this paper, we introduce a new class of nested regression models for binary outcomes. Under certain conditions, the parameters of this model can be interpreted as causal counterfactual transition parameters associated with the effect of the covariate, but the model may be useful even if the parameters do not have this causal interpretation. We describe the properties of the model and discuss its motivation and attractiveness relative to traditional models for binary outcomes such as logistic, log-binomial, identity link and probit regression. We then derive inference results for the parameters. Finally, we propose a general form of the model, of which both logistic regression and the probabilistic transition model are special cases. This general form allows some variables to be associated with a risk-multiplicative parameter, and other variables to be associated with an odds-multiplicative parameter.

1. BACKGROUND

Several different regression models exist for modelling the conditional distribution of a binary outcome variable. The most commonly used models include logistic regression models, log-binomial models, linear probability models and probit models. These models can all be understood in the generalized linear model (GLM) framework, where they are associated with different link functions. The link function is closely related to the specific effect measure used to the model the treatment effects; for example: the logistic model uses a log odds link and is therefore based on the odds ratio; the log-binomial model uses a log probability link and is therefore based on risk ratios, whereas the linear probability model uses an identity link and is therefore based on risk differences. Models for binary outcomes may be associated with different shortcomings. Among the most widely recognized shortcomings are non-collapsibility (meaning that the model is associated with an effect measure whose marginal value is not equal to a weighted average of the conditional effect measures) and predictions outside of the range $[0, 1]$.

In this paper, we propose a new class of non-GLM models for binary outcomes. In section two, we propose the Switchrisk Model, which is based on an effect parameter called the associational switch risk ratio. Switchrisk Models have a number of advantages over existing models: In contrast with logistic regression, the effect parameters are readily collapsible, do not rely on unintuitive odds ratios, and may have an attractive causal interpretation. In contrast with Log-Binomial Models and Linear Probability Models, the Switchrisk Model never predicts risks outside the range $[0, 1]$. In section three, we provide inference results for the Switchrisk Model. In section four, we generalize the principles behind the Switchrisk Model and propose Generalized Probability Transition Models, a new framework for modelling binary outcomes based on iterated application of associational probability transition operators. We show that Switchrisk Models, Logistic Models, Log-Binomial Models and Linear Probability Models are all special cases of the Generalized Probability Transition Model, which also allows mixed models where different covariates are associated with different effect parameters. We argue that such mixed models may be useful in settings where only the parameter associatated with the primary intervention variable will be given a causal interpretation.

2. SWITCHRISK MODELS

2.1 Switch risk ratios

The Switchrisk Model is a multivariable regression model for binary outcomes, based on an effect parameter called the *switch risk ratio*. While the switch risk ratio is a simple and readily interpretable effect parameter, it is not currently widely known

or used. We therefore begin by reviewing the history and interpretation of this measure.

The risk ratio is asymmetric, such that the inferences from a risk ratio model for the probability of death (Y=1) are fundamentally different from similar inferences from a risk ratio model for the probability of survival (Y=0). For this reason, the risk ratio is often conceptualized as two separate parameters, which we will here call the risk ratio and the survival ratio. Each parameter can be defined either as a causal parameter (in terms of unobservable counterfactual quantities) or as an associational parameter (in terms of observed variables):

$$\text{Associational RR} \stackrel{def}{=} \frac{\Pr(Y=1|A=1)}{\Pr(Y=1|A=0)}$$

$$\text{Associational SR} \stackrel{def}{=} \frac{1-\Pr(Y=1|A=1)}{1-\Pr(Y=1|A=0)} = \frac{\Pr(Y=0|A=1)}{\Pr(Y=0|A=0)}$$

$$\text{Causal RR} \stackrel{def}{=} \frac{\Pr(Y^{a=1}=1)}{\Pr(Y^{a=0}=1)}$$

$$\text{Causal SR} \stackrel{def}{=} \frac{1-\Pr(Y^{a=1}=1)}{1-\Pr(Y^{a=0}=1)} = \frac{\Pr(Y^{a=1}=0)}{\Pr(Y^{a=0}=0)}$$

Van der Laan et al [1] proposed the causal switch risk ratio, which selects the causal risk ratio if $\Pr(Y^{a=1}=1) < \Pr(Y^{a=0}=1)$, and the causal survival ratio if $\Pr(Y^{a=1}=1) > \Pr(Y^{a=0}=1)$, thereby ensuring that one always selects the version of the risk ratio that is in the range $[0,1]$. This setup has the desirable property that it never leads to predictions outside the range of valid probabilities. While the parameter in Van der Laan et al was explicitly a causal switch risk ratio, it is also possible to define an analogous parameter in associational terms. Baker and Jackson [3] proposed the generalized relative risk reduction (GRRR) as a measure of meta-analytic effect; GRRR is best understood as a notationally convenient representation of the associational switch risk ratio; it is a number in the range $[-1,1]$ which is equal to $1-\frac{1-\Pr(Y=1|A=1)}{1-\Pr(Y=1|A=0)}$ if the risk is higher in the exposed, and equal to $-1+\frac{\Pr(Y=1|A=1)}{\Pr(Y=1|A=0)}$ if the risk is lower in the exposed.

We note that the motivation for using models based on the switch risk ratio depends in part on being able to give the effect parameter a causal interpretation. A full data generating mechanism which motivates this model was described by Huitfeldt et al, in terms of Counterfactual Outcome State Transition probabilities[2]. In the presence of exchangeability and monotonicity, associational switch risk ratio corresponds directly to counterfactual outcome state transition parameters (GRRR = G_A-H_A if exchangeability + monotonicity). However, a causal interpretation is not necessary to derive utility from models based on switch risk ratios. In the remainder of this manuscript, we will focus exclusively on observational quantities, in order to discuss the associational model in terms of its statistical properties.

2.2 Probabilistic transition operators

In section 2.4, we propose multivariable models based on associational risk ratios, partially adopting the notational conventions introduced in Baker and Jackson. We first define the probabilistic transition operator $\odot$ and the effect coefficients γ. Together, these objects are used to parametrize the model.

The probabilistic transition operator $\odot$ acts on a probability p and a real number R as follows:

$$p \odot R \stackrel{def}{=} \begin{cases} 1 - (1-p) \times (1-R) & \text{if } R > 0 \\ p & \text{if } R = 0 \\ p \times (1+R) & \text{if } R < 0 \end{cases}$$

We refer to the output of this operation as q. Note that if R is in the range [-1,1], then q is in the range [0,1], i.e. is a valid probability. It will be useful to also define the inverse operation of $\odot$, $\odot^*$, :

$$p \odot^* q \stackrel{def}{=} \begin{cases} 1 - \frac{1-q}{1-p} & \text{if } q > p \\ 0 & \text{if } q = p \\ -1 + \frac{q}{p} & \text{if } q < p \end{cases}$$

It is worth pointing out here that this inverse operation is not anticommutative: $p \odot^* q \neq -q \odot^* p$, this relates to the asymmetry of the risk ratio. Therefore, models based on $\odot$ will not be invariant to temporal order. This will be discussed in greater detail later in the manuscript.

2.3 Gamma parameters

In our model, each predictor will have a γ coefficient associated with it, defined as follows:

$$\gamma_A \stackrel{def}{=} \Pr(Y = 1|A = 0) \odot^* \Pr(Y = 1|A = 1)$$

Gamma coefficients have the following interpretation: If $\gamma_A = 0$, A and Y are independent. If $\gamma_A > 0$, A positively predicts Y and $1-\gamma_A$ is the multiplicative parameter that describes the relationship between $\Pr(Y = 0|A = 1)$ and $\Pr(Y = 0|A = 0)$. If $\gamma_A < 0$, A negatively predicts Y, and $1+\gamma_A$ is the multiplicative parameter that describes the relationship between $\Pr(Y = 1|A = 1)$ and $\Pr(Y = 1|A = 0)$.

2.4 Multivariable models with binary predictors

Suppose we have a temporally ordered data generating structure, $A \to B \to C \to Y$ and we are interested in modelling the distribution of binary outcome variable Y, conditional on A, B and C. We next describe a nested model for the distribution of Y conditional on A,B, C, with an intercept parameter for the baseline risk β (in the range $[0, 1]$)[1] and effect coefficients γ (in the range $[-1, 1]$) associated with each independent variable in the model. Initially, we will assume that all predictors are binary; this assumption will be relaxed later.

All effect parameters will be defined conditional on future covariates taking the value 0. The preliminary notation which is used in this draft proposal, is that the subscript on the gamma parameter first shows the past covariates, then a semicolon, then the variable which it describes the effect of, then a semicolon, then all future covariates (with lower case), which are set to 0. For example, the parameter for the effect of B is $\gamma_{A;B;c=0}$.

If all predictors are binary, the fully saturated model is then the following nested set of equation, which together define the conditional distribution of Y in terms of β and γ coefficients, for all values of a, b and c:

$$\Pr(Y = 1|A = 0, B = 0, C = 0) = \beta$$

$$\Pr(Y = 1|A = 1, B = 0, C = 0) = \Pr(Y = 1|A = 0, B = 0, C = 0) \odot \gamma_{A;b=0,c=0}$$

$$\Pr(Y = 1|A = a, B = 1, C = 0) = \Pr(Y = 1|A = a, B = 0, C = 0) \odot \gamma_{A=a;B;c=0}$$

$$\Pr(Y = 1|A = a, B = b, C = 1) = \Pr(Y = 1|A = a, B = b, C = 0) \odot \gamma_{A=a,B=b;C}$$

In order reduce the number of parameters, one may assume effect homogeneity by forcing the effects in subgroups defined by earlier covariates to be equal. For example, if the parameter for C is equal between subgroups defined by A and B, $\gamma_{A=a,B=b;C}$ is constant for all values of a and b, and is therefore a single parameter. Specific forms of interaction can be allowed for example allowing $\gamma_{A=a,B=b;C}$ to differ depending on a or b, thereby increasing the number of model parameters.

2.5 Models with continuous predictors

The idea for continuous predictors is less set in stone; what follows is a rough draft:

If there are continuous predictors (if e.g. B in continuous), we consider the "baseline" value of B to be the lowest observed value of the predictor. $B = \overset{min}{B}$ plays the same role in the model with continuous B as B=0 does for binary B. The "full effect" of exposure to B occurs for those observations with highest observed

[1]Baseline risk is here defined as the distribution of Y conditional on A=0, B=0, C=0

value $\overset{max}{B}$ of the predictor. In these models, "linearity" will mean that the gamma coefficient associated with each incremental increase in the predictor is constant.

To formalize this, let the possible values of B increase in increments of ϵ, and define K = $\frac{\overset{max}{B} - \overset{min}{B}}{\epsilon}$.

The incremental effect of covariate B, when increasing by the k'th unit of ϵ, $\gamma_{A=a;B;c=0;k}$ is then defined as

$$\Pr(Y = 1|A = a, B = \overset{min}{B}_{+\epsilon}, C = 0) = \Pr(Y = 1|A = a, B = \overset{min}{B}, C = 0) \odot \gamma_{A=a;B;c=0,k=1}$$

$$\Pr(Y = 1|A = a, B = \overset{min}{B}_{+2\times\epsilon}, C = 0) = \Pr(Y = 1|A = a, B = \overset{min}{B}_{+\epsilon}, C = 0) \odot \gamma_{A=a;B;c=0;k=2}$$

If linearity is assumed, $\gamma_{A=a;B;c=0;k}$ is constant over k

2.6 Collapsibility

Gamma coefficients are collapsible, at least in the weak sense that collapsibility weights exist. Whether there exists a simple identifying expression for collapsibility weights is less immediately clear to me (but seems likely). It is clear that if the effect is either positive in every stratum, or negative in every stratum, then such weights exist and are trivial extensions of those described in Huitfeldt et al (2019)[4]. If some strata have positive effects and other strata have negative effects, then collapsilibity weights may be more challenging.

3. MAXIMUM LIKELIHOOD ESTIMATION

This is a very rough draft section:

Let

$$f_\gamma(p, V) = \begin{cases} p \odot \gamma & \text{if } V = 1 \\ p & \text{if } V = 0 \end{cases}$$

For any individual i, their probability of $Y = 1$ given model parameters β and γ and their covariates A_i , B_i and C_i is then given by:

$\Pr(Y = 1|A = A_i, B = B_i, C = C_i, \beta, \gamma) = f_{\gamma_{A_i,B_i;C}}(f_{\gamma_{A_i;B;c=0}}(f_{\gamma_{A;b=0,c=0}}(\beta, A_i), B_i), C_i)$

The overall likelihood function is then:

$\prod_i \Pr(Y = 1|A = A_i, B = B_i, C = C_i, \beta, \gamma)^{Y_i} \times (1 - \Pr(Y = 1|A = A_i, B = B_i, C = C_i, \beta, \gamma))^{1-Y_i}$

Note that the likelihood is non-differentiable when a γ parameter is equal to zero. One may therefore need approximate or numerical methods to maximize it. This is outside of my expertise, and the success of this project depends on finding a mathematically oriented statistical collaborator who is willing to take the lead on this section of the manuscript.

4. GENERALIZED PROBABILITY TRANSITION MODEL

4.1 Similar operations with odds ratios

Next, we define the operation $\diamond$ and its inverse operation $\diamond*$. These are similar to $\odot$ and $\odot*$, but but based on odds ratios instead of risk ratios:

$$p \diamond R \stackrel{def}{=} \begin{cases} 1 - \frac{\frac{1-p}{p} \times (1-R)}{1+\frac{1-p}{p} \times (1-R)} & \text{if } R > 0 \\ p & \text{if } R = 0 \\ -1 + \frac{\frac{p}{1-p} \times (1+R)}{1+\frac{p}{1-p} \times (1+R)} & \text{if } R < 0 \end{cases}$$

$$p \diamond^* q \stackrel{def}{=} \begin{cases} 1 - \frac{\frac{1-q}{q}}{\frac{1-p}{p}} & \text{if } q > p \\ 0 & \text{if } q = p \\ -1 + \frac{\frac{q}{1-q}}{\frac{p}{1-p}} & \text{if } q < p \end{cases}$$

We can construct a model similar to the one in section 2, by replacing $\odot$ with $\diamond$. In this model, the parameters are δ in the range $[-1, 1]$, but can be trivially rewritten as odds ratios:

$$\Pr(Y = 1|A = 0, B = 0, C = 0) = \beta$$

$$\Pr(Y = 1|A = 1, B = 0, C = 0) = \Pr(Y = 1|A = 0, B = 0, C = 0) \diamond \delta_{A;b=0,c=0}$$

$$\Pr(Y = 1|A = a, B = 1, C = 0) = \Pr(Y = 1|A = a, B = 0, C = 0) \diamond \delta_{A=a;B;c=0}$$

$$\Pr(Y = 1|A = a, B = b, C = 1) = \Pr(Y = 1|A = a, B = b, C = 0) \diamond \delta_{A=a,B=b;C}$$

This model is equivalent to logistic regression. Note that in contrast to $\odot^*$, $\diamond^*$ is anticommutative. Therefore, models based on $\diamond$ are invariant to temporal order of the covariates.

4.2 Other possible operations

Similar operations can be described based on essentially any effect measure, including the risk difference, the risk ratio, the survival ratio and the arcsin differences. Models based on these operations will have the following properties:

- Invariant to temporal order if inverse operation is anti-commutative
- Will not give invalid predictions if operation is closed on [0,1] in the range of valid parameter values

- Will be collapsible if operation is (?can we state a result about this in terms of mathematical properties of the operator?)

Of particular interest will be the operation for the effect of removing treatment. This is equivalent to $\odot$ but with the coding of A reversed ($A = 0 \Rightarrow A* = 1$ and $A = 1 \Rightarrow A* = 0$).

4.3 Mixed Probability Transition Models

Next, we consider general model where some variables are associated with the $\diamond$ operations, and other variables with the $\odot$ operation (or other operations). The mixed model may be useful if investigators do not wish to make assumptions about temporal order for baseline covariates. The baseline covariates A and B are then associated with a coefficient based on $\diamond$ and are used simply to estimate the risk among those untreated with C; and the intervention variable C has a coefficient based on $\odot$, which one may give a causal interpretation:

$$\Pr(Y = 1|A = 0, B = 0, C = 0) = \beta$$

$$\Pr(Y = 1|A = 1, B = 0, C = 0) = \Pr(Y = 1|A = 0, B = 0, C = 0) \diamond \delta_{A;b=0,c=0}$$

$$\Pr(Y = 1|A = a, B = 1, C = 0) = \Pr(Y = 1|A = a, B = 0, C = 0) \diamond \delta_{A=a;B;c=0}$$

$$\Pr(Y = 1|A = a, B = b, C = 1) = \Pr(Y = 1|A = a, B = b, C = 0) \odot \gamma_{A=a,B=b;C}$$

REFERENCES

[1] Mark J. Van Der Laan, Alan Hubbard, and Nicholas P. Jewell. Estimation of treatment effects in randomized trials with non-compliance and a dichotomous outcome. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3):463–482, 2007.

[2] Anders Huitfeldt, Andrew Goldstein, and Sonja A. Swanson. The choice of effect measure for binary outcomes: Introducing counterfactual outcome state transition parameters. *Epidemiologic Methods*, 7(1), 2018.

[3] Rose Baker and Dan Jackson. A new measure of treatment effect for random-effects meta-analysis of comparative binary outcome data. *ArXiv:1806.03471*, 2018.

[4] Anders Huitfeldt, Mats J. Stensrud and Etsuji Suzuki. On the collapsibility of measures of effect in the counterfactual causal framework. *Emerging Themes in Epidemiology*, 16:1742-7622, 2019.

ACKNOWLEDGEMENTS