Design of ordinal CRM

Wojciech Wójciak based on John Kirkpatrick work

Last run on 2022-11-02 by kirkpatj on rbaub000162.bas.roche.com

The statistical model and theory about it.

The patient's toxicity response is not binary (no DLT, DLT) but categorical (for example: no DLT, sub-DLT, DLT). Furthermore, multicategory toxicity responses are treated as ordinal due to their ordinal characteristics such as a monotone trend.

The model: Cumulative Link Models

These type of responses can be modeled using Generalized Linear Model (GLM), where the model is expressed in terms of the cumulative response probabilities. Hence, the name: Cumulative Link Model, or CLM. The CLM assumes the form:

$$h[p_{ij}] = h[P(Y_i \leq j)] = \alpha_j + \mathbf{x}_i^T \boldsymbol{\beta}, \qquad j = 1, \ldots, c-1,$$

where $c \in \{2, 3, \ldots\}$ denotes a number of toxicity categories on ordinal scale, and the components of the model are as follows:

1. h - link function. The most common link functions are "logit" and "probit", which result in Cumulative Logit Model and Cumulative Probit Model respectively.
2. Y_i - random variable that represents the response outcome category for subject $i \geq 1$, i.e. $supp(Y_i) = \{1, 2, \ldots, c\}$, $i \geq 1$. Probabilities $P(Y_i = j)$, $j = 1, \ldots, c$, for $i \geq 1$, are called category probabilities, and the cumulative response probabilities for subject $i \geq 1$ are defined as

   $$p_{ij} = P(Y_i \leq j) = P(Y_i = 1) + \cdots + P(Y_i = j), \qquad j = 1, \ldots, c;\ i \geq 1,$$

3. $\mathbf{x}_i$ - a vector of covariates corresponding to main effects and interactions for subject $i \geq 1$.
4. α_j, $j = 1, \ldots, c-1$ are "cutpoints" and $\boldsymbol{\beta}$ is a vector of effects (without intercept term). Although each response category has a corresponding cutpoint, the regression coefficients $\boldsymbol{\beta}$ are constant across categories. Each cumulative logs has its own intercept ("cutpoint") and

   $$-\infty < \alpha_1 < \alpha_2 < \cdots < \alpha_{c-1} < \infty,$$

   because $P(Y_i \leq j)$ increases in j at any fixed $\mathbf{x}_i$, and the logit is an increasing function of $P(Y_i \leq j)$, $i \geq 1$.

Note that a model for $logit[p_{ij}]$ alone is an ordinary logistic model for a binary response in which categories 1 to j represent "success" and categories $j+1$ to c represent "failure".

Why cumulative probabilities rather than category probabilities?

As explained by McCullagh and Nelder (1989), the choice and definition of response categories is often arbitrary. It is essential, therefore, if we are to arrive at valid conclusions, that the nature of those conclusions should not be affected by the number of choice of response categories. Such considerations lead to models that based on the cumulative probabilities rather than category probabilities. These two sets of probabilities are equivalent, but simple models for the cumulative probabilities are likely to have better properties for ordinal response scales than equally simple models based on the category probabilities.

The prior

Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{c-1})$ be a random vector, where $c \in \{2, 3, \ldots\}$ denotes a number of toxicity categories. Specifying priors for $\boldsymbol{\alpha}$ is challenging because of the ordering restriction. The following approaches have been suggested.

Truncated prior distribution

$P(\boldsymbol{\alpha}) = P(\alpha_1)P(\alpha_2|\alpha_1)\cdots P(\alpha_j|\alpha_{j-1})$, where $\alpha_1 \in \mathbb{R}$ and $supp(\alpha_j) = (\alpha_{j-1}, \infty)$, $j = 2, \ldots, c-1 \geq 2$. For example:

$$\alpha_1 \sim \mathcal{N}(0, \sigma_\alpha^2), \alpha_j|\alpha_{j-1} \sim \mathcal{N}(0, \sigma_\alpha^2)\, T(\alpha_{j-1}, \infty),\ j = 2, \ldots, c-1 \geq 2$$

See McKinley et al. (2013) (equation (11)) and the references given therein, i.e. Albert and Chib (1993), Johnson and Albert (1999), Congdon (2005). Eventually see James et al. (2022)

I see however the problem with this approach. Namely, if α_k for some $k \in \{1, \ldots c-2\}$ is large by a chance, then it forces that α_{k+1} will be even higher, regardless of whether it is practically justified. This maybe become an issue, if the number of categories c is high. So the alternative could be doubly-truncated distributions, which is somehow given in Congdon (2005), but it is unclear to me how this should be used exactly.

[JK] I agree this is a potential concern. However, my experience is that it has not ben an issue in the models I have fitted to date.

[JK] Codifying the potential priors like this is excellent. Thank you. Can you add an explicit discussion on the effect on the covariance structure specified by the user that each might induce?

[JK] As a minimal viable product, is it worth considering the requirement that the user-defined covariance matrix for the model parameters is diagonal? This has worked for me in the past, and avoids all issues of correlation!

Transformation of the intercepts to an unconstrained space

$\delta_1 := log(\alpha_1)$, $\delta_j := log(\alpha_j - \alpha_{j-1})$, $j = 2, \ldots, c-1 \geq 2$. Then, e.g. $\boldsymbol{\delta} \sim \mathcal{N}_{c-1}(\mu_\delta, \Sigma_\delta)$.

Note that it is required that $\alpha_1 > 0$.

See Albert and Chib (1997) for more details.

An ordered uniform distribution (Ishwaran, 2000)

This approach was mentioned in few sources, but it is still a bit unclear to me. For instance, in Congdon (2005), on page 239, the author writes: "An alternative suggested by Ishwaran (2000) is a uniform density

$$0 = \alpha_1 < \alpha_2 < \cdots < \alpha_{c-1} = U,$$

where U is equal to or less than c."

Latent response model

With this approach, the observed response Y is often taken to reflect an underlying continuous random variable Y^* with $c-1$ thresholds or cut points. Albert and Chib (1993) presented a Bayesian analysis that utilizes the latent (response) variable model:

$$Y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \qquad i \geq 1,$$

where ϵ_i are iid and $\epsilon_i \sim \mathcal{N}(0, 1), i \geq 1$. Here we have that

$$Y_i = j \quad \text{iff} \quad \alpha_{j-1} < Y_i^* \leq \alpha_j, \quad j = 2, \ldots, c-1,$$

where $\alpha_0 := -\infty$ and $\alpha_c := \infty$.

Following Congdon (2005) (see Chapter 7.2) that is based on Johnson and Albert (1999) (Chapter 4.3), the cut points are sampled in a way that takes account of the sampled Y^*. Thus, at iteration t:

$$\alpha_j^{(t)} \sim \mathcal{N}(0, V_\alpha)\, T(L_j, U_j), \quad j = 1, \ldots, c-1,$$

where V_α is preset and

$$L_j = max(Y^{*(t)},\, Y_i = j) U_j = min(Y^{*(t)},\, Y_i = j+1)$$

and

$$Y_i^* \sim \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, \gamma_i)\, T(\alpha_{y_{i-1}}, \alpha_{y_i}),$$

where $\gamma_i \sim Ga(0.5v, 0.5v)$, with v preset.

LogisticLogNormOrd

For simplicity of notation, unless I need to refer to particular subjects or to particular values of the explanatory variables, we replace p_{ij} (i.e. $P(X_i < j)$) in such equations by p_j (i.e. $P(Y < j)$), keeping in mind that in the model this is actually a conditional probability at each fixed value for the explanatory variables.

The following basic CLM can be considered as a single-agent toxicity model, in which the covariate is the natural logarithm of the dose x divided by the reference dose $x*$, i.e.:

$$logit[p_j] = \alpha_j + \beta\, log(\tfrac{x}{x*}).$$

The prior is:

$$log(\beta) \sim N(\mu_\beta, \sigma_\beta^2)$$

and for $\{\alpha_j\}_{j\in\{1,\ldots,c-2\}}$ I would propose: "truncated prior distribution" or "transformation of the intercepts to an unconstrained space".

References

[1] Agresti, A. (2010). Analysis of Ordinal Categorical Data. Hoboken, NJ: John Wiley & Sons.

[2] Agresti, A. (2013). Categorical Data Analysis. Hoboken, NJ: John Wiley & Sons.

[3] Agresti, A. (2015). Foundations of Linear and Generalized Linear Models. Hoboken, NJ: John Wiley & Sons.

[4] Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association, 88(422): 669–679.

[5] Albert, J. and Chib, S. (1997). Bayesian methods for cumulative, sequential and twostep ordinal data regression models. Technical report.

[6] Congdon, P. (2005). Bayesian Model for Categorical Data, Willey.

[7] Ishwaran, H. (2000). Univariate and Multirater Ordinal Cumulative Link Regression with Covariate SpecificCutpoints. The Canadian Journal of Statistics / La Revue Canadienne de Statistique, Vol. 28, No. 4 (Dec., 2000), pp. 715-730

[8] James, N. T., Harrell Jr, F. E., Shepherd, B. E. (2022) Bayesian Cumulative Probability Models for Continuous and Mixed Outcomes. arXiv:2102.00330v2 [stat.ME]

[9] Johnson, V. E. and Albert, J. H. (1999). Ordinal Data Modeling. Springer-Verlag, New York.

[10] McKinley, T. J., Morters, M., Wood, J. L. N. (2015) Bayesian Model Choice in Cumulative Link Ordinal Regression Models.

[11] McCullagh, P., and Nelder, J. A. (1989). Generalized Linear Models. 2nd ed. London: Chapman & Hall https://www.utstat.toronto.edu/~brunner/oldclass/2201s11/readings/glmbook.pdf

[12] Van Meter EM, Garrett-Mayer E, Bandyopadhyay D. Dose-finding clinical trial design for ordinal toxicity grades using the continuation ratio model: an extension of the continual reassessment method. Clin Trials. 2012 Jun;9(3):303-13. doi: 10.1177/1740774512443593. Epub 2012 Apr 30. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5531273/

Prototype code

Fictitious observed data

```
dose_grid <- c(5, 15, 45, 70, 100, 220, 300, 600, 1000, 1800, 4000, 10000, 16000)

x <- c(
  rep(5, 1),
  rep(15,  4),
  rep(45,  5),
  rep(70,  5),
  rep(100,  5),
  rep(220,  8),
  rep(300,  6),
  rep(600, 15),
  rep(1000,  8),
  rep(1800,  9),
  rep(4000, 10),
  rep(10000, 14),
  rep(16000,  2)
)

y <- c(
  rep(0, 1),
  rep(0, 4),
  rep(0, 5),
  rep(0, 5),
  rep(0, 5),
  rep(0, 7), 1,
  rep(0, 5), 1,
  rep(0, 10), 1, 1, 1, 1, 1,
  rep(0, 5), 1, 1, 1,
  rep(0, 8), 1,
  rep(0, 8), 1, 1,
  rep(0, 10), 1, 1, 1, 1,
  rep(1, 1),
  rep(2, 1)
)
```

Data-class

After many considerations, I think the best would be to slightly modify existing Data class by: - allowing y to take on values from 0, 1, 2, 3, ... - adding new slot y_categories, which keeps all the possible values for y, with default to y_categories = c(0L, 1L). This is mostly for validation purposes, but maybe it would also be useful for some other purposes, which I am not aware of at the moment. This proposed approach has a lot of advantages, e.g. it allows child classes of Data to benefit from having toxicities on an ordinal scale.

[JK] If we extend the existing `Data` class rather than subclassing it, how will the behaviours of models (such as `LogisiticLogNormal` etc) be defined when they are passed a `Data` object with `length(y_categories) > 2`?

[JK] It would be helpful to allow users to label the toxicity categories as they wish. This could be achieved either by allowing the `y_categories` slot to be a named vector or by providing sister slot, `y_labels`, say. In either case the default labels could be `c("No DLT", "DLT")` for the current (default) case and `paste0("Cat ", y_categories)` when the model is ordinal rather than binary.

[JK] There is an inconsistency in the naming of the new slot: it uses snake_case, whereas existing slots use camelCase. Is this pre-empting a planned refactoring?

Lets call it Data0 for a moment

```
data <- Data(
  x = x,
  y = y, # here is a problem because y slot accepts only 0 and 1 values here
  doseGrid = dose_grid,
)

.Data0 <- setClass(
  Class = "Data0",
  contains = "GeneralData",
  slots = c(
    x = "numeric",
    y = "integer",
    doseGrid = "numeric",
    nGrid = "integer",
    y_categories = "integer",
    xLevel = "integer",

    placebo = "logical"
  ),
  prototype = prototype(
    x = numeric(),
    y = integer(),
    doseGrid = numeric(),
    nGrid = 0L,
    xLevel = integer(),
    y_categories = c(0L, 1L),
    placebo = FALSE
  ),
  # validity = v_data # same as Data but more validation for yGrid.
)
```

Model-class

IT IS NOT READY YET. WILL BE PROPOSED AFTER THE THEORY IS ACCEPTED.

```
LogisticLogNormOrd <- function(mean, cov, ref_dose = 0) {
  params <- ModelParamsNormal(mean, cov)
  .LogisticLogNormOrd(
    params = params,
    ref_dose = ref_dose,
    datamodel = function() {
      for (i in 1:nObs) {
```

```
        x_rel[i] <- log(x[i] / ref_dose)
        for (j in 1:(y_categories-1)) {
          logit(p[i, j]) <- alpha[j] + beta * x_rel[i]
          y[i, j] ~ dbern(p[i, j])
        }
      }
    },
    priormodel = function() {
      theta ~ dmnorm(mean, prec)
      alpha <- theta[1:(y_categories-1)]
      beta <- exp(theta[y_categories])
    },
    modelspecs = function(from_prior) {
      ms <- list(mean = params@mean, prec = params@prec)
      if (!from_prior) {
        ms$ref_dose <- ref_dose
      }
      ms
    },
    init = function() {
      list(theta = c(0, -20)) # TODO
    },
    datanames = c("nObs", "y_categories", "y", "x"),
    sample = c("alpha", "beta")
  )
}
```