A multinomial spatio-temporal model for sexual risk behaviour in AGYW

HIV Inference Lab Group Meeting

Adam Howes

Imperial College London

August 2021

Background

- Want to estimate the proportion p_k of AGYW in $K = 4$ risk categories
  - $k = 1$: Not sexually active `nosex12m`
  - $k = 2$: Cohabiting, single partner `sexcohab`
  - $k = 3$: Non-regular partner(s) `sexnonreg`
  - $k = 4$: Key population, female sex worker `sexpaid12m`
- Used as a step towards estimating total incidence by group
  - By district, age, sex, time
- Different policy responses depending on where new infections are

Goal

- Estimate proportions indexed by
  - District $i = 1, \dots, n$
    - 13 AGYW Global Fund priority countries (but fitting individually): Botswana, Cameroon, Eswatini, Kenya, Lesotho, Malawi, Mozambique, Namibia, South Africa, Tanzania, Uganda, Zambia and Zimbabwe
  - Survey $t = 1, \dots, T$
    - The available DHSs
  - Age $a \in \{15\text{–}19, 20\text{–}24, 25\text{–}29\}$
  - Category k
- e.g. in Lilongwe, in 2015, how many AGYW aged 20-24 are cohabiting with a single partner?

Why not just use the raw data

1. Usual high variance from low counts problem, improved by spatio-temporal smoothing
2. The raw proportions don't add up to one

Multinomial likelihood

- We have proportions $p_1, \ldots, p_4$ which are all $\in [0,1]$ and $\sum_k p_k = 1$
- $\implies$ suggests using a multinomial likelihood

  $\mathbf{y}_{ita} \sim \text{Multinomial}(m_{ita}; p_{ita1}, \ldots, p_{itaK})$

$$p(\mathbf{y}_{ita} \mid m_{ita}, p_{ita1}, \ldots, p_{ita4}) = \frac{m_{ita}!}{y_{ita1}! \times \cdots \times y_{ita4}!} p_{ita1}^{y_{ita1}} \times \cdots p_{ita4}^{y_{ita4}}$$

Multivariate model

- Requires a multivariate model in that each observation depends on K probabilities
  - Really $K - 1$ numbers, since they are constrained to sum to one
- Usually this is done by modelling the “contrasts” between categories
  - How much more likely to be in this category (or group of categories) than that category (or group of categories)
  - Different ways to do this
- These types of models are popular in economics to model choice
  - Treat the latent field like utility
- Some keywords: discrete choice model, multinomial logit model

Baseline category

- Pick one of the $k \in \{1, \ldots, K\}$ as your baseline category, say $k = 2$
- Define log-odds for being in category $k \neq 2$ versus $k = 2$

$$\log\left(\frac{p_{itak}}{p_{ita2}}\right)$$

- Three of these: $\{2\}$ versus $\{1\}$, $\{2\}$ versus $\{3\}$ and $\{2\}$ versus $\{4\}$

Nested

- Attractive when it is appropriate to model individuals as making choices sequentially
- $\{1\}$ versus $\{2, 3, 4\}$, $\{2\}$ versus $\{3, 4\}$ and $\{3\}$ versus $\{4\}$
- Choosing 1. whether or not to have sex, 2. conditional on having sex, whether to have one cohabiting partner or irregular partners seems reasonable
- Choosing 3. whether or not to be a FSW based upon choosing to have irregular partners is less reasonable
  - i.e. not like you first choose to have many partners then choose to be FSW, more like choose to be FSW based on economic factors say

Workaround requried

- Want to fit this model quickly and easily using `R-INLA`
- Sadly `R-INLA` doesn't work for likelihoods which depend upon multiple elements of the latent field (like this multinomial model)
- There is a workaround, called the Poisson trick
  - A multinomial likelihood can be rewritten as a Poisson likelihood (with some additional nuisance parameters)
  - So this respecifies the model in terms of a Poisson likelihood for each category of the observation

Sidenote: ordinal

- $\pi_k = \sum_{l \leq k} p_l$ be the cumulative probability of category k where $\pi_1 \leq \pi_2 \leq \cdots \leq \pi_K$.
- Ordinal logistic regression is based upon $K - 1$ cumulative logits $\text{logit}(\pi_k) = \eta_k$ where the linear predictors are ordered
- One way to do this is to assume η_k only differ in their intercept (also called cut-point)
- I wonder if you can fit this with a single linear predictor and a category random effect?
  - Possibly wouldn't need Poisson trick

Poisson trick

- Based on the following fact
- Let $y_k \sim \text{Pois}(\lambda_k)$, $k = 1, \ldots, K$ then $\mathbf{y} \mid n \sim \text{Multinomial}(\mathbf{y}; n, p_1, \ldots, p_K)$
  - $m = \sum_k y_k$
  - $p_j = \lambda_j / \sum_{k=1}^{K} \lambda_k$
- In words, given their sum, Poisson counts are jointly multinomially distributed (McCullagh and Nelder 1989)
  - In some some way you can obtain the multinomial likelihood (what we want) from Poisson likelihoods (what we have to work with)

Poisson trick

- In the multinomial likelihood, the sample size is treated as fixed
- Instead treat $m = \sum_k y_k$ as random. Let $\Lambda = \sum_{k=1}^{K} \lambda_k$ and suppose that $m \sim \text{Poisson}(\delta\Lambda)$ (Lee, Green, and Ryan 2017)

$$
\begin{aligned}
P(\mathbf{y}, m) = P(m)P(\mathbf{y} \mid m) &= \exp(-\delta\Lambda)\frac{(\delta\Lambda)^m}{m!} \times \frac{m!}{\prod_k y_k!} \prod_k \left(\frac{\lambda_k}{\Lambda}\right)^{y_k} \\
&= \prod_k \left(\frac{\exp(-\delta\lambda_k)\,(\delta\lambda_k)^{y_k}}{y_k!}\right)
\end{aligned}
$$

Poisson trick

- To find the marginal $P(\mathbf{y})$ sum over the possible values of m (0 to ∞)

$$P(\mathbf{y}) = \prod_k \left(\frac{\exp(-\delta\lambda_k)\,(\delta\lambda_k)^{y_k}}{y_k!} \right)$$

- $\implies y_k \sim \text{Poisson}(\delta\lambda_k)$ independently

Poisson trick

- "The Poisson surrogate model eliminates m from the denominator of the multinomial probabilities. This makes sense intuitively, as we do not expect the multinomial sums to provide any useful information in estimating the fixed effects." (Lee, Green, and Ryan 2017)
- Example of a broader class of Poisson transformations where the theme is to consider the normalisation constant as another parameter (Barthelmé and Chopin 2015)
- I guess that this must mean the normalisation constant is "estimated" exactly always

Poisson log-linear model

- $y_{itak} \sim \text{Poisson}(\lambda_{itak})$
- $\log(\lambda_{itak}) = \eta_{itak}$ the linear predictor for district i, survey t, age-group a and category k
- Including observation-specific random effects $\theta_{ita} \sim \mathcal{N}(0, \tau_\theta^{-1})$ in the model $\eta_{itak} = \theta_{ita} + \cdots$ assures exact reproduction of the multinomial denominators
  - By "observation" I mean observation of the multinomial e.g. a vector `c(2, 5, 3, 0)`
  - These variables correspond to $\log(\delta)$ from before. So we're just having a free parameter which can
- Then you put whatever mixed model things you'd like in

No intercept

- Don't include an intercept, i.e. `-1` in the formula
- This is because we are interested in recovering probabilities from the model

$$p_j = \frac{\lambda_j}{\sum_{k=1}^{K} \lambda_k} = \frac{\exp \eta_j}{\sum_{k=1}^{K} \exp \eta_j}$$

- So adding a constant β_0 on to each of the η_j does nothing (so I'd hope that if you did include a constant it would turn out to be zero if you fitted the model)

$$\frac{\exp\left(\eta_j + \beta_0\right)}{\sum_{k=1}^{K} \exp\left(\eta_j + \beta_0\right)} = \frac{\exp \eta_j}{\sum_{k=1}^{K} \exp \eta_j}$$

- Note that this `exp(x) / sum(exp(x))` function is called "softmax"

Nine models

- Always use age-category random effects $\alpha_{ak} \sim \mathcal{N}(0, \tau_\alpha^{-1})$
  - I haven't been including category random effects, but based on my research for this presentation I think I should have been. I think because I have the age-category it's already mopping up the intercept
- Three choices for spatial-category random effects ϕ_{ik}:
  - None
  - IID $\phi_{ik} \sim \mathcal{N}(0, \tau_\phi^{-1})$
  - BYM2
- Three choices of time-category random effects γ_{tk}
  - None
  - IID $\gamma_{tk} \sim \mathcal{N}(0, \tau_\gamma^{-1})$
  - AR1
- Some amount of nuance / choice to defining the structured {spatial (BYM2), temporal (AR1)}-category interactions

First way

- Let `area_idx` be an indicator the the area
- `area_idx.k = ifelse(category == k, area_idx, NA)` is that same indicator but just for the category `k`
- Then (say you want spatial) you can use `f(area_idx.k, model = "bym2", graph = adjM)` in your `R-INLA` formula
- This results in a set of K independent spatial random effects

$$\phi_{ik} = \sqrt{\tau_{\phi_k}^{-1}} \left(\sqrt{\pi_k} \cdot u_{ik} + \sqrt{1 - \pi_k} \cdot v_{ik} \right), \quad k = 1, \ldots, K$$

- Key point is that you have different hyperparameters for each category $\{\tau_{\phi_k}, \pi_k\}$
  - Do we want this?
- Also have to define `area_idx.k` and write `f(...)` more times (annoying)

Second way

- Alternative (in `R-INLA`) is to use the `group` option which lets you define "Gaussian Kronecker product Markov random fields"
  - Sounds fancy, isn't that fancy
- "Precison matrix" (probably it doesn't have full rank so it's not really a precision) defined by $\mathbf{Q} = \mathbf{Q}_1 \otimes \mathbf{Q}_2$
- In `R-INLA` it's more like the "structure matrix" is defined by $\mathbf{R} = \mathbf{R}_1 \otimes \mathbf{R}_2$
  - So that the precision is $\mathbf{Q} = \tau \mathbf{R}$

Example of group option

- Let's say we're trying to define the time-category interaction random effects, using an AR1 on time
- If there are three surveys, and we suppose $\rho = 1$ then the precision matrix for this AR1 is

$$\begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

- To get the structure matrix $\mathbf{R_1}$ we'd divide by `exp(mean(log(diag(Q))))`

Example of group option

- For the "category random effects" grouping variable we want IID structure, so that the precision (and structure matrix $\mathbf{R_2}$) are

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- The R-INLA call for this is f(sur_idx, model = "ar1", group = cat_idx, control.group = list(model = "iid"), ...)
- The new structure matrix looks like

$$\mathbf{R} = \begin{pmatrix} \mathbf{R_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R_1} \end{pmatrix}$$

- Important difference: just have one set of hyperparameters $\{\tau_\gamma, \rho\}$
  - i.e. there is pooling across categories doing it this way

Survey weights

- Use survey weights to calculate effective observation `x_eff`
- May not be integers i.e. `c(1.87, 5.32, 2.20, 0.15)`
- Deal with this by using `xPoisson` (generalisation of the Poisson to non-integer data) rather than `Poisson`
- Might be equivalent to weighted log-likelihood approach

Data issues for `sexpaid12m` category

- Two different questions asked in the DHS, and answered very differently
- Before 2015, asked about most recent three partners. One of the categories for these is paid
  - Very rare to list one of these as paid, using this question often think there are essentially no FSW
- After 2015, asked if have given or received gifts or money in exchange for sex
  - Better answered but still has issues
    - Missing some FSW from survey sampling frame
    - Not all "gifts of money in exchange for sex" is sex work

Current approach

- Can't fit a (meaningful) spatio-temporal model when the question asked / response is so different
- Have been fitting single-survey spatial models to the most recent DHS
- Otherwise have created a three-category model, moving all `sexpaid12m` into `sexnonreg` to create a new `sexnonregplus` category
  - Haven't checked that some of the `sexpaid12m` shouldn't go into `sexcohab` rather than `sexnonreg` yet (sorry Katie!)
- Challenge is what to do about the `sexpaid12m` category

What to do about the `sexpaid12m` category

1. Ignore it
2. Try to use other data sources to help

- Previous estimates from workbook are based upon national estimates of FSW population size (perhaps the UNAIDS Key Population Atlas)

Johnston et al. (2021, preprint)

Deriving and interpreting population size estimates for adolescent and young key populations at higher risk of HIV transmission: men who have sex with men, female sex workers and transgender persons

- Disaggregates the UNAIDS published population size estimates by age (nice, but not by district, which we'd also want) using proportion of sexually active adults
- Kinh is a coauthor and warns that
  - The estimates should (often) be seen as expert opinion rather than based on data
  - Several countries had no data
  - Rounding up when the number is too small

Laga et al. (2021, preprint)

Mapping female sex worker prevalence (aged 15-49 years) in sub-Saharan Africa

- Collates available FSW population size estimates
- Small-area model with covariates used to extrapolate to district level
- Has code available
- Probably the best single source to rely on, done a lot of the work already

Hodgkins et al. (2021, preprint)

HIV prevalence, population sizes, and HIV prevention among men who paid for sex in sub-Saharan Africa: a meta-analysis of 82 population-based surveys (2000-2020)

- Proportion of men who pay for sex (CFSW) estimated from survey data
- Could be linked to proportion FSW by some model e.g.

$$p_{\text{CFSW}} = B \times p_{\text{FSW}}$$

  with a strongly informative prior $B \sim p(B)$ (around 10 say)
  - Don't know if this is a good model, anyone have any ideas?
  - Extract estimates from Hodgkins and plot versus estimates of FSW to see if it looks linear
  - Of course the ratio is going to vary, but perhaps it's quite stable?
- Fit a model on the good DHS `sexpaid12m` question, then use it where there isn't the question

$$\log(p_{\mathrm{CFSW}}) - \log(p_{\mathrm{FSW}}) = \log(B) \tag{1}$$

$$\log\left(\frac{p_{\mathrm{CFSW}}}{p_{\mathrm{FSW}}}\right) = \log(B) \tag{2}$$

References I

Barthelmé, Simon, and Nicolas Chopin. 2015. "The Poisson Transform for Unnormalised Statistical Models." *Statistics and Computing* 25 (4): 767–80.

Lee, Jarod YL, Peter J Green, and Louise M Ryan. 2017. "On the" Poisson Trick" and Its Extensions for Fitting Multinomial Regression Models." *arXiv Preprint arXiv:1707.08538*.

McCullagh, Peter, and John A Nelder. 1989. *Generalized Linear Models*. Routledge.