Statistics for HIV surveillance

StatML Kick-off Camp

Adam Howes

Imperial College London

September 2021

Welcome!

- Welcome to the CDT!
- Congratulations for many of you on completing your previous studies under some non-ideal circumstances
- PhDs can be challenging, learning process
  - I hope I'm a better researcher than I was two years ago
- Important to find the right work environment, draw on each other

About me / my project

- Surveillance of HIV in sub-Saharan Africa in collaboration with UNAIDS
- Main interest is in methods and applications of Bayesian statistics towards policy-relevant questions

Spatio-temporal

- Observations are indexed by space s and time t
- For example, often we have a sequence of national surveys
- In social science it's natural that we observe people in space and time
- Key property of spatio-temporal data is correlation structure

Bayesian statistics

- Bayesian statistics makes sense to me
- Perhaps more importantly, there isn't a lot to remember
  1. Write down a model for your process
  2. Some kind of algorithm to obtain the posterior
  3. Use the posterior for all decision theory

Data and theory

- The data we work with is incomplete, biased, noisy
- Mainly concerned with getting roughly the right answer
- Happy to use informative priors or scientific knowledge of domain experts to help

1. Size of the sample

- Running a (household) survey is expensive
- We only get small-sample sizes, and a low coverage of space-time
- The field of “small-area” estimation should really be called “small-data” estimation
- We use the fact that spatio-temporal data is correlated to improve things with smoothing

2. Who is in the sample?

- The sample probably often isn't representative
- Some people aren't on the sampling frame at all
- Example: for HIV, there are some key population groups (men who have sex with men, transgender people, injecting drug users, sex workers) who are hard to reach with surveys

3. Are the answers they give accurate?

- Survey questions can be ambiguous
- People may not be able to or want to answer questions
- Example: age at first sex is an important epidemiological parameter, but men and women (and old and young) give answers which are difficult to make line up

Sexual risk categories model

- Some work in progress of mine, to hopefully make the above concrete!
- Young women are an important population for HIV prevention
  - 10% of the population and 25% of all new infections (UNAIDS 2021)
- Not enough resources to provide programs for all young women in high prevalence settings $\implies$ standardised method for accessing population at risk required

Category	HIV related risk	HIV risk ratio
None	Not sexually active	0
Low	Sexually active, one cohabiting partner	1
High	Non-regular sexual partner(s)	1.72
Very high	Young women from key populations	13

Table 1: HIV risk categories and risk ratios.

Goal

- Estimate proportion of young women in each risk category
  - Districts $i = 1, \ldots, n$ in 13 AGYW Global Fund priority countries
  - Survey $t = 1, \ldots, T$
  - Age $a \in \{15\text{–}19, 20\text{–}24, 25\text{–}29\}$
- e.g. in Lilongwe (Malawi), in 2015, how many AGYW aged 20-24 are cohabiting with a single partner?

Spatial random effects

- Tried none, IID and the Besag model (most common structured spatial random effect model)

$$b_i \mid \mathbf{b}_{-i} \sim \mathcal{N}\left(\frac{\sum_{j:j\sim i} b_j}{n_{\delta i}}, \frac{1}{n_{\delta i}}\right),$$

Temporal random effects

- Tried none, IID and an AR1 model (most common structured spatial random effect model)

$$
\begin{aligned}
a_1 &\sim \left(0, \frac{1}{1-\rho^2}\right), \\
a_t &= \rho a_{t-1} + \epsilon_t, \quad t = 2, \ldots, T,
\end{aligned}
$$

Model ID	1	2	3	4	5	6	7	8	9
Spatial structure	S1	S2	S3	S1	S2	S3	S1	S2	S3
Temporal structure	T1	T1	T1	T2	T2	T2	T3	T3	T3

Table 2: All of the models considered

References I

UNAIDS. 2021. "2021 UNAIDS Global AIDS Update - Confronting inequalities - Lessons for pandemic responses from 40 years of AIDS." Geneva, Switzerland.