A Markov Chain Monte Carlo Procedure with Simulated Annealing for Short Form Creation

Klint Kanopka, Victoria Zhang, Daphna Harel



Collaborators



Victoria Zhang Master's Student xz2661@nyu.edu



Daphna Harel Associate Professor of Applied Statistics ASH daphna.harel@nyu.edu



Onions Puppy

Research Question:

Can we develop a method to create easy to administer and score short forms that reproduce clinical diagnosis?

Why shorten forms?

- Long forms can be a burden on clinicians and respondents
- Longer forms may require specialized administration
- Sometimes forms get reused for slightly different purposes than what they were originally intended for
 - Constructs may no longer align perfectly
 - Some items may contribute little unique information for this new purpose



How are short forms used?





How are short forms used?



- Weights are often **ignored** in clinical practice!
- Instead, users just use unweighted sum scores



Why not cut out the middle man?









Where's the catch?

- As it turns out, we have great tools to drop items and weight the rest
 - LASSO regression
 - Bayesian methods with spike-and-slab priors
- With the unweighted sum score rule, things get trickier!
- One approach: We can check every possible short form and choose the best one based on some criteria
 - If you have 20 items, you have $2^{20}-2 = 1,048,574$ possible short forms to check
 - Each form has multiple different scoring thresholds
 - If you have 30 items, you have over a billion forms to check!
- Another approach: Randomly check a bunch of forms until we get tired of doing it and pick the best one you find
 - It's certainly *an* approach
 - Not particularly principled, however

📍 NYU STEINHARDT



Markov Chain Monte Carlo (MCMC)

- Exists halfway between checking everything and checking randomly
- Instead of manually testing all possible short forms, MCMC leverages random sampling to search for a well-performing subset
- The trick is that we weight all of the changes we make by how good they perform, and use those guide future choices and (hopefully) explore better solutions!



MCMC

- Start with a random short form (say, ten items out of our 20)
- Propose a small change
 - Normally we select one item from the possible pool and either add or remove it
 - If you have a specific form length you're trying to hit, you can swap one item out at random for another
- Evaluate if the change improves performance
 - If yes, keep the change! It's better!
 - If not, maybe keep it anyway with a probability proportional to the quality of the solution
 - This lets you theoretically explore every possible solution and get yourself "unstuck" if you find a local maxima that isn't fully optimal!
- Repeat for a long time, and pick the best performing solution



MCMC with Simulated Annealing

- Start with a random short form (say, ten items out of our 20)
- Propose a small change
 - Normally we select one item from the possible pool and either add or remove it
 - If you have a specific form length you're trying to hit, you can swap one item out at random for another
- Evaluate if the change improves performance
 - If yes, keep the change! It's better!
 - If not, maybe keep it anyway with a probability proportional to the quality of the solution
 - This lets you theoretically explore every possible solution and get yourself "unstuck" if you find a local maxima that isn't fully optimal!
- Repeat for a long time, and pick the best performing solution
- Gradually reduce the probability of accepting worse solutions
 - This lets you explore more at first and then narrow down to find the best solution in a region



Form Evaluation

- Form evaluation requires **two stages:**
 - Within a given shortening, **what cut score** produces the best performance?
 - Across shortenings, **what combination of items and cut score** produces the best performance?
- Each proposed form must be evaluated at each possible cut score to determine the best version of that shortening
- The best version of each possible shortening are compared across proposed forms to arrive at a final choice
- But what does "best" even mean?



Form Evaluation

- We need a metric to compare different short forms and it needs to be flexible to proposed uses:
 - WBA = q*Sensitivity + (1-q)*Specificity
 - Where q is a weight from 0 to 1
- Sensitivity is the **true positive rate**
 - Are you correctly classifying people with AUD as having AUD?
- Specificity is the **true negative rate**
 - Are you correctly classifying people without AUD as not having AUD?
- Different values of q matter in different circumstances!



Simulation Studies

Simulation Study

- Simulation datasets
 - N = 1000 individuals responding to J = 20 dichotomous items
 - Introduced a simulation condition J*: the proportion of "informative" items
 - Tested J* ∈ {1/3, 1/2, 2/3} (7, 10, 14 informative items)
- Two simulations
 - Fixed short form length ($k \in \{3, ..., 18\}$)
 - Floating short form length k
- Evaluated short forms using weighted balance accuracy (WBA)
 - Evaluate for $q \in \{\frac{1}{3}, \frac{1}{2}, \frac{2}{3}\}$
- Parallelized for efficiency
 - Each condition was estimated using 5 chains to check for convergence



Proportion of best solutions that contain each of the 20 items

Key Observation: Our method automatically retains informative items and automatically drops uninformative items!



Empirical Application

Data Source

- 233 young Swiss men answering 20 dichotomous questions designed to screen for alcohol use disorder (AUD)
 - Data from Baggio et al. (2020) and Iglesias et al. (2018)
 - Available from the Item Response Warehouse (Domingue et al., 2024).
- Items ask the presence or absence of the eleven DSM-5 criteria and nine other alcohol-related consequences
- Alongside self-report data, the dataset contains a dichotomous gold standard judgement for the presence or absence of AUD obtained from a clinician-administered Diagnostic Interview for Genetic Studies (Berney et al., 2002)



20-item Alcohol Use Disorder (AUD) Screener

In the past 12 months,...

1	has your drinking alcohol caused you more than once to miss a class,
	work or to fail to look after your family at home?
2	did you more than once drive a car or another vehicle (such as a bicycle, motorcycle
	or moped) shortly after you had had several drinks with alcohol?
3	did you resume your drinking habits even though your drinking had caused
	problems with your partner, friend or acquaintances?
4	did you find you needed a lot more alcohol to become high or drunk than you used to?
5	did you start feeling nervous or shaky for a full day or more after you had cut
	down on your drinking?
6	did you often find yourself drinking more and for longer periods of time than you intended?
7	did you try to cut down on your drinking, but couldn't?
8	did you find yourself spending a great deal of time obtaining, using, or recovering
	from the effects of alcohol?
9	did you give up activities you care about (e.g. school, work or being with friends
	and family) because of your drinking?
10	did you continue drinking even though you were aware that alcohol had repeatedly
	caused you anxiety, depression or health problems?
11	have you had such a strong desire or urge to drink that you could not help drinking?
12	I drank alcohol or took drugs or medicine in order to get over any of the bad
	secondary effects of drinking alcohol
13	I had a mental blackout after drinking alcohol
14	While drinking alcohol, I did something that I badly regretted later
15	I had unplanned sex because I was drunk
16	I had sex without a condom because I was drunk
17	I had an accident or I got injured because I was drunk
18	I came into a conflict with the police or with the authorities more than once
	because of my alcohol use
19	I came into an argument or into a fight while drinking alcohol or straight after
20	I damaged property, because I was drunk

20-item Alcohol Use Disorder (AUD) Screener



9	did you give up activities you care about (e.g. school, work or being with friends
	and family) because of your drinking?
10	did you continue drinking even though you were aware that alcohol had repeatedly

caused you anxiety, depression or health problems?

14	While drinking alcohol, I did something that I badly regretted later
15	I had unplanned sex because I was drunk
16	I had sex without a condom because I was drunk
17	I had an accident or I got injured because I was drunk
18	I came into a conflict with the police or with the authorities more than once
	because of my alcohol use
19	I came into an argument or into a fight while drinking alcohol or straight after
20	I damaged property, because I was drunk

WBA for best performing short form of a given length

Key Observation: Under the unweighted sum scoring constraint, shorter forms can outperform longer forms when reproducing clinical diagnosis!





Discussion

- Sometimes our goal is not measurement, but accurate reproduction of expert judgement
- Our method efficiently creates shortened forms that can outperform longer forms
- Shortening the 20 item AUD form took under five minutes on a 2017 MacBook Pro with a 2.3 GHz Dual-Core i5 processor and 16GB of RAM
- We want people to start using this!
 - Reach out if you'd like code or help implementing it
 - An R package is on the way!





klint.kanopka@nyu.edu

klint@bsky.social

