Assessing Uniformity in Sampling of Sedentary Times

Xiang Meng
xmeng@g.harvard.edu

JSM 2021 Novel Methodology for the Analysis of Physical Activity Data Measured by Accelerometers

9 August 2021

HARVARD UNIVERSITY

Collaborators

- Harvard University: Peng Liao, Susan Murphy
- University of Michigan: Walter Dempsey, Predrag (Pedja) Klasnja
- University of Washington: Nick Reid

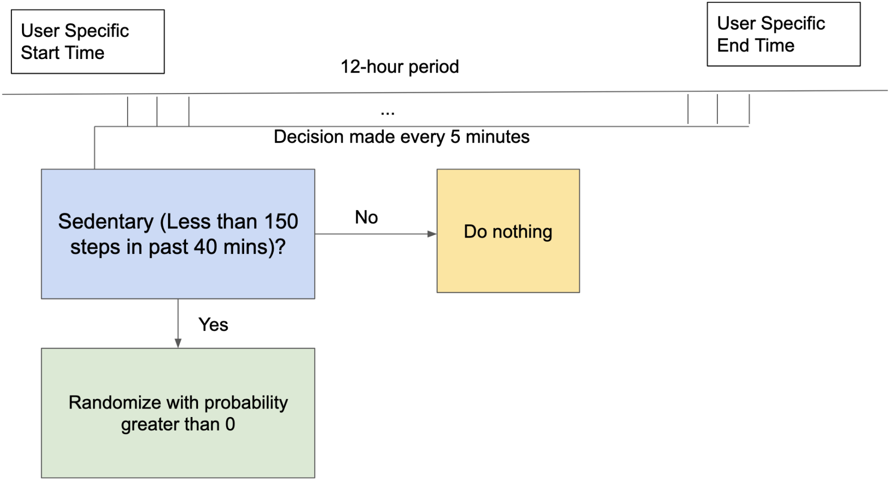
Contents

A Key Question to Mobile Intervention System

- Mobile intervention: Change one's behaviour
- Trade-off in number of interventions

- Context also matters.
- Key Question: How often, and in what context should a mobile intervention be provided?
- One way to address the key question is Micro-Randomized Trial (MRT).
- Our task: To design a MRT to address the key question.

Example: MRT to Reduce Sedentary Behavior



A Sampling Algorithm of Just-in-time Intervention is Critical

- Want to design a MRT to learn when to intervene, specifically at sedentary times.
- A sampling algorithm should spread intervention uniformly across all sedentary times.
- Uniformity
  - Uniform number of interventions per day
  - Uniform distribution on sedentary times
- To achieve this
  - Set a hard constraint on number of intervention per day
  - At each sedentary time, predict the remaining sedentary times

HeartSteps Study

- Aim: Develop a mobile activity assistant for individuals who are newly diagnosed with Stage 1 Hypertension
- Intervention: A push notification to encourage exercise.
- Decision time: Every 5 mins
- Notation:
  - $t = 1, 2, 3...$, decision time. $\mathcal{T}_d$ set of decision times for day d;
  - Each day is separated into three 4-hour blocks. Denote $\mathcal{B}_{d,k}$ as kth block on day d, $\bigcup_{k=1}^{3} \mathcal{B}_{d,k} = \mathcal{T}_d$
  - A_t: intervention; H_t: history; I_t: sedentary indicator

Description of the Sampling Algorithm, i.e. Randomisation Probabilities

- Goal:
  1. An average of 0.5 Intervention per Block $\mathcal{B}_{d,k}$
  2. Uniform distribution in sedentary times
- For Goal 1: Substract all previous probabilities from the hard constraint. $N_1^* - \sum_{s \in \mathcal{B}_{d,k}, s \leq t-1} I(I_s = 1)\pi_s$
- For Goal 2: If remaining sedentary times g_t is known, by setting $N_1^* = 0.5$ we can sample uniformly with probability

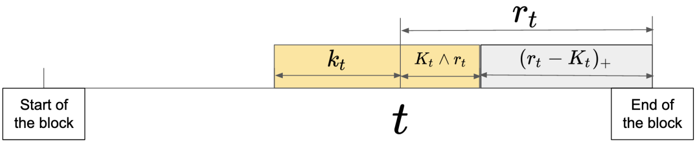
$$\frac{N_1^* - \sum_{s \in \mathcal{B}_{d,k}, s \leq t-1} I(I_s = 1)\pi_s}{1 + g_t}$$

- Truncate to $[0.005, 0.995]$

Tuning parameters:

- g_t: forecast of the remaining sedentary times in the block.
- N_1^*: desired average number of interventions per block.

Model for Prediction of the Remaining Sedentary Times g_t



- Notation:
  - k_t: run length.
  - K_t: remaining run length of the current run
  - r_t: remaining time in the current block
  - F_t: the fraction of sedentary times outside the run.
- Estimate for g_t:

$$\hat{g}_t = \hat{E}[K_t \wedge r_t] + \hat{E}[F_t]\hat{E}[(r_t - K_t)_+]$$

- $\hat{E}[K_t \wedge r_t], \hat{E}[F_t], \hat{E}[(r_t - K_t)_+]$ are non-parametric averages from prior data.

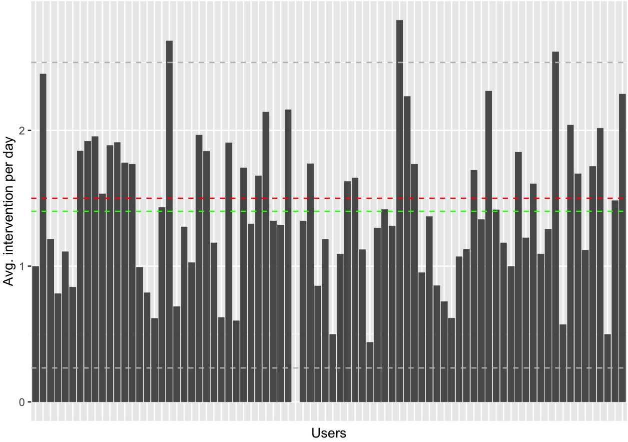
How many interventions were delivered on each block and each day?

- We first calculate average interventions per user per block over all days.
- Then for each block, we do a summary for all users. Total 81 users are considered.
- Goal: 0.5 per block and 1.5 per day.

	Median	Mean	Std. Dev.
Block 1	0.476	0.471	0.219
Block 2	0.451	0.456	0.241
Block 3	0.448	0.476	0.259
Daily	1.333	1.404	0.568

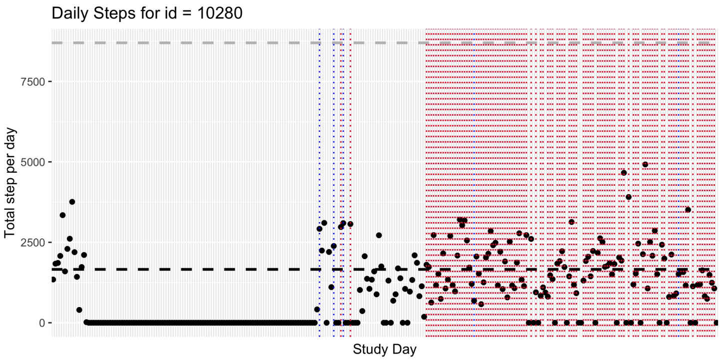
- Algorithm is good at targeting the goal.
- Std. Dev. increases with block because number of sedentary times decrease with block.

Average interventions sent per day across users



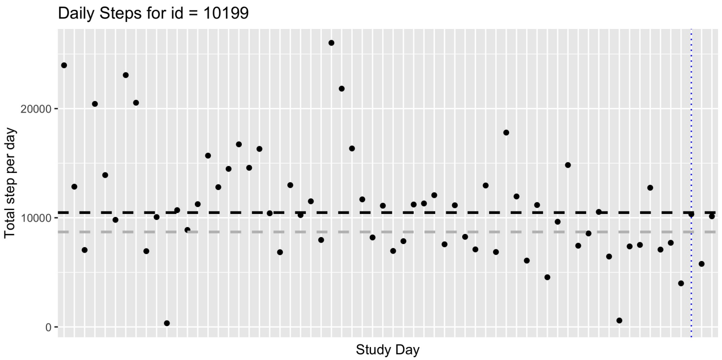
- Red line is the target (1.5 per day)
- Green line is average over users (1.404 per day).
- Grey lines: high (2.5 per day) and low (0.25 per day).

Three Users with daily avg intervention greater than 2.5



- Each point is daily sum of Steps for the User.
- Black horizontal line: average daily steps for **this user**.
- Gray horizontal line: average daily steps across **all (user,day)**.
- Blue vertical line: at least one sedentary time on the day.
- Red vertical line: at least one intervention on the day
- Other two users have the same pattern.

One user with daily avg intervention lower than 0.25



- Each point is daily sum of Steps for the User.
- Black horizontal line: average daily steps for **this user**.
- Gray horizontal line: average daily steps across **all (user,day)**.
- Blue vertical line: at least one sedentary time on the day.
- Red vertical line: at least one intervention on the day

Is the Algorithm Uniform Across all Sedentary Times?

- Notation: $N_{u,d,k}$ as the number of sedentary time per (user, day, 4-hour block).
- Measure: KL divergence **between the actual** treatment probability p_a and the **uniform** distribution defined by $p_u = 0.5/N_{u,d,k}$.
- We calculate mean KL divergence on each (user, day, 4-hour block) by

$$\frac{1}{N_{u,d,k}} \sum_{t=1}^{N_{u,d,k}} KL(p_{a,t}, p_{u,t})$$

where $KL(p,q) = p\log(p/q) + (1-p)\log((1-p)/(1-q))$

A Comparison Using KL Divergence

A baseline algorithm—Block sampling

- First, from prior data, calculate average number of sedentary times across all (user, day) in block k: $\hat{M}_k, k = 1, 2, 3$.
- Then set the randomisation probability to be $0.5/\hat{M}_k$ for sedentary time in block k. M1: 0.0280, M2: 0.0255, M3: 0.0256

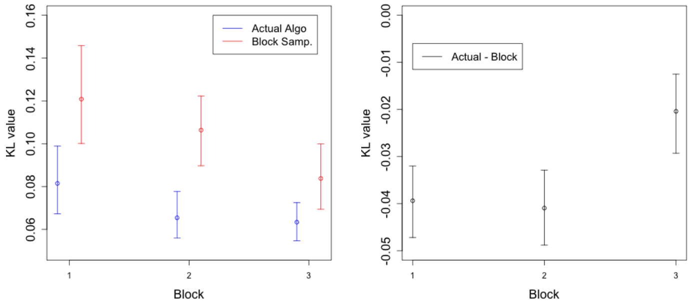


Figure: Left: KL between the actual probability(blue)/block sampling(red) and the uniform probability. Right: CI of their difference

Discussion: How generalizable is our analysis?

- Missing sedentary times could lead to bias.
- Type A: Missing sedentary time due to bad data propagation.
- Type B: Missing sedentary time due to user behaviour.
- User population will differ in real life
  - Our analysis still gives strong indication if underlying dynamics of missingness are similar.

Open Problems

- Other ways to measure whether the algorithm is uniform?
- How to model sedentary time such that it can form a good test bed to test algorithms?
- How would one design an algorithm with better performance?

Thanks for your attention!

Any questions?