A generative, predictive model for menstrual cycle lengths that accounts for potential self-tracking artifacts in mobile health data

Kathy Li
Applied Physics and Applied Mathematics
Columbia University
New York, NY 10027
kathy.li@columbia.edu

Iñigo Urteaga
Applied Physics and Applied Mathematics
Columbia University
New York, NY 10027
inigo.urteaga@columbia.edu

Amanda Shea
Clue by BioWink
Adalbertstraße 7-8, 10999
Berlin, Germany
amanda.shea@biowink.com

Virginia J. Vitzthum
Kinsey Institute & Department of Anthropology
Indiana University
Bloomington, IN 47405
vitzthum@indiana.edu

Chris H. Wiggins
Applied Physics and Applied Mathematics
Columbia University
New York, NY 10027
chris.wiggins@columbia.edu

Noémie Elhadad
Biomedical Informatics
Columbia University
New York, NY 10032
noemie.elhadad@columbia.edu

Abstract

Mobile health (mHealth) apps such as menstrual trackers provide a rich source of self-tracked health observations that can be leveraged for statistical modeling. However, such data streams are notoriously unreliable since they hinge on user adherence to the app. Thus, it is crucial for machine learning models to account for self-tracking artifacts like skipped self-tracking. In this abstract, we propose and evaluate a hierarchical, generative model for predicting next cycle length based on previously tracked cycle lengths that accounts explicitly for the possibility of users forgetting to track their period. Our model offers several advantages: 1) accounting explicitly for self-tracking artifacts yields better prediction accuracy as likelihood of skipping increases; 2) as a generative model, predictions can be updated online as a given cycle evolves; and 3) its hierarchical nature enables modeling of an individual's cycle length history while incorporating population-level information. Our experiments using real mHealth cycle length data from 5,000 menstruators show that our method yields state-of-the-art performance against neural network-based and summary statistic-based baselines.

1 Introduction

The menstrual cycle is an important indicator of health, influencing illnesses, fertility, and sexual well-being [1, 2, 3, 4, 5, 6, 7]. Two open questions in menstrual health research are how to accurately predict the occurrence of the next menstrual cycle and how to understand the underlying mechanisms of menstruation. The rise of popularity in menstrual trackers has enabled access to large-scale, longitudinal cycle length data for rigorous inquiry [8, 9, 10, 11]. However, such self-tracked mHealth data poses the risk of users skipping tracking. By leveraging a generative model, researchers may be able to better distinguish true menstrual cycle behavior from these self-tracking artifacts. We propose a generative model that encodes measures of "typical" cycle length (i.e., cycle length regularity) and propensity to skip tracking (i.e., self-tracking adherence) separately.
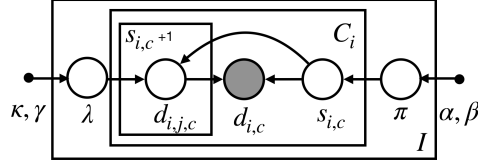
Figure 1: Hierarchical graphical model for proposed generative process. Individual-level parameters λ_i (average cycle length without skipping) and π_i (probability of skipping a cycle) are drawn from population-level distributions characterized by hyperparameters $u = [\kappa, \gamma, \alpha, \beta]$. Shaded circles represent observed data, open circles represent latent variables, and dots represent hyperparameters.

Our generative approach accounts for the possibility of the observed cycle containing more than one skipped cycle and updates its prediction as the cycle proceeds. We assess our model's predictive ability by comparing it against alternative baselines. We also distinguish "regular" and "irregular" users based on a previously proposed definition of variability [11] to observe how performance differs by variability group.

2   Methods

Proposed generative model.   Our hierarchical, generative model for self-tracked menstrual cycle lengths incorporates population-wide knowledge (via informative priors for hyperparameters) and learns individualized cycle length patterns (via per-user parameters and predictions). The proposed model outlined in Figure 1 extends Poisson regression (that represents expected user cycle lengths via parameter λ_i) with individual latent variables π_i that accommodate each user's propensity for skipping tracking. As such, we can decouple true from observed cycle lengths $d_{i,c}$: the variable $s_{i,c}$ indicates the number of not reported (skipped) cycles for user i and observed cycle c. The proposed generative process is detailed in the Supplementary Information.

Hyperparameter inference.   Population-level hyperparameter inference is achieved via type-II maximum likelihood estimation. Expert knowledge is incorporated via hyperparameter priors. The data log-likelihood under the proposed model is computed by an importance sampling (IS) approach that offers reduced execution runtimes (see Supplementary Information for details).

Individualized predictions.   We compute the expectation of the conditional predictive posterior distribution to make predictions for each user's next cycle length, d^*. Due to the flexibility of the proposed generative model, we can update predictions online by conditioning on the day of the current cycle, i.e., $E[d^*|d^* > d_{current}, d_i, \hat{u}]$, where $d_{current}$ refers to the current day of the cycle and d^* denotes the predicted cycle length (i.e., if $d_{current} = 1$, then d^* must be at least 2). To compute the expected cycle length, our model offers two possibilities: 1) set $s = 0$ to assume the next reported cycle will be truth (i.e., that the next observed cycle will not be skipped); or 2) set $s \geq 0$ and integrate it out, assuming the next reported cycle may not be truth (i.e., accounting for the user possibly skipping their next cycle tracking). We evaluate both options, with details on the prediction procedure provided in the Supplementary Information.

Dataset.   We leverage a de-identified self-tracked dataset from Clue [12], a popular and accurate menstrual tracker, comprised of $378,694$ users, with a median age of 25 years, a median of 11 cycles tracked, and a median cycle length of 29 days (cohort details in Supplementary Information). We focus on period data (i.e., the users' self-reports on which days they have their period), since we only use cycle lengths as model input. We define a menstrual cycle as the span of days from the first day of a period through to and including the day before the first day of the next period [13]. We randomly select 5,000 menstruators with more than 11 cycles tracked from the full cohort and use the first 11 cycles as input. We preserve the proportion of regular to irregular users (i.e., 90% to 10%), based on previous work on menstrual variability measured via the median absolute cycle length difference [11], and we utilize this variability definition as a lens through which to interpret our results.

Alternative baselines.   To evaluate the predictive performance of our proposed model, we consider summary statistic-based and neural network-based baselines (details in Supplementary Information):

- `Mean and median baselines`: the predicted next cycle for each user is the average (or median) of their previously observed cycle lengths.
- `CNN`: a 1-layer convolutional neural network with a 3-dimensional kernel.
- `RNN`: a 1-layer bidirectional recurrent neural network with a 3-dimensional hidden state.
- `LSTM`: a 1-layer Long Short-Term Memory neural network with a 3-dimensional hidden state.

3 Results

We first compare our cycle length prediction accuracy at day 0 of the next cycle (i.e., when the user reports their period). Table 1 showcases that our model outperforms neural network-based models and is competitive with summary statistic-based baselines. Our model prediction improves when integrating out s, demonstrating that accommodating for skipped cycles is beneficial for prediction accuracy. We note that prediction performance for the regular group is clearly superior to that of the irregular group, which may be due to increased difficulty of predicting highly variable cycle lengths.

Table 1: Prediction RMSE results by model on day 0 (C=10)

Model	Full	Regular	Irregular
Mean	7.810	6.810	13.877
Median	7.786	6.729	14.099
CNN	8.341	7.227	15.022
LSTM	8.137	7.038	14.704
RNN	7.781	6.781	13.841
Proposed model	7.682	6.674	13.757
Proposed model ($s = 0$)	7.861	6.697	14.636

A benefit of the proposed model is that we can generate day-by-day predictions by explicitly conditioning on the current (non-period) cycle day. We observe that our model outperforms baselines as the cycle proceeds. Figure 2 showcases prediction results at cycle day $21, 29$, and 60. On cycle day 29, our model begins to outperform the alternative baselines across number of training cycles C. This performance gap is magnified beyond day 29, indicating the clear advantage of evolving predictions by day. The performance fluctuations over C across all evaluated models are due to randomness in the data – the sampled users may experience unusual cycles at varying times within their cycle tracking history
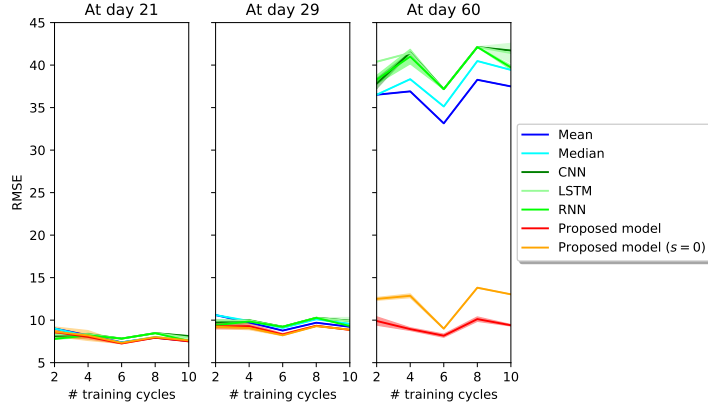

Figure 2: Prediction RMSE over number of training cycles C, for different cycle days. $C = 2$ means 2 input cycles were used to predict the third and so on. Note that all baseline model performances deteriorate past day 29 due to their static nature – they lack prediction updates with current cycle day.

In addition to our model's ability to compute online predictions, another distinguishing feature is that it explicitly accounts for the possibility of skipped cycles. As the cycle day proceeds, the probability of skipping increases in a multimodal fashion (see Figure 3). For instance, if the current cycle day has already passed day 29 and no period has been tracked yet, then the likelihood of a skipped cycle is high, which results in an increased probability of the cycle length being around 60 days. This can be seen in the conditional predictive posterior distribution of the next cycle length $p\left(d^{*}|\hat{u}, d_{i}, d^{*} > d_{current}\right)$ as the current day in the cycle ($d_{current}$) evolves in Figure 3. When we assume $s = 0$ as in Figure 3a, the predictive posterior is unimodal; however, when we account for potential cycle skips as in Figure 3b, the distribution becomes multimodal – as the current day within a cycle proceeds past 29, there is an increased likelihood of a cycle skip, resulting in probability peaks in the cycle length distribution for d^{*} around 60 and 90 (i.e., one or two skipped cycles). The combined effect of accounting for skipping and evolving prediction by day enables our model to dynamically update its predictions, providing a clear advantage over static baselines. The benefit of accounting for skipped cycles as the cycle proceeds is especially noticeable for those with more variable cycles, as seen in Figure 4.

(a) Conditional predictive posterior assuming next cycle is not skipped ($s = 0$).

(b) Conditional predictive posterior accommodating that next cycle may be skipped ($s \geq 0$).
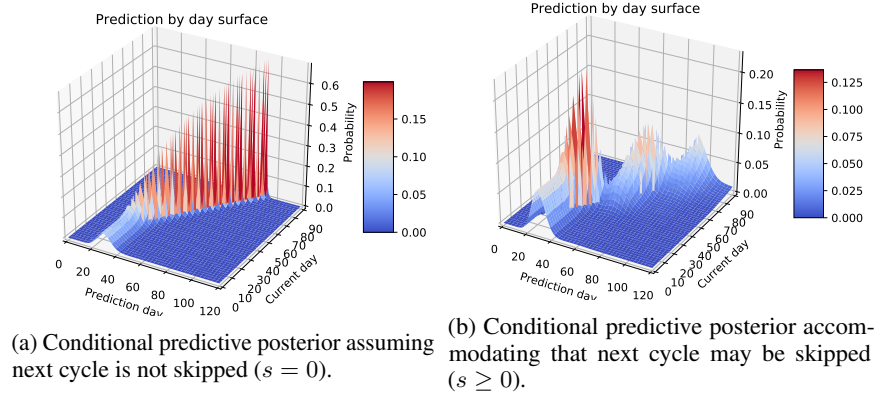
Figure 3: Comparison of predictive posterior over prediction day d^* (i.e., what the next cycle is predicted to be) and current day $d_{current}$ (i.e., day in next cycle) for the same user, assuming either that next observed cycle is truth or that next observed cycle may contain skipped cycles. Note that multimodality occurs when $s \geq 0$; chance of skipping peaks as typical cycle length is surpassed.
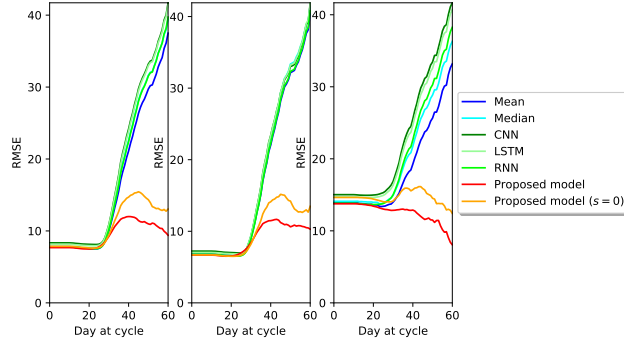

Figure 4: Comparison of prediction RMSE over current day for the whole population (left), regular (middle) and irregular (right) users. Note that for irregular users the benefit of accommodating for skipped cycles is especially apparent – the RMSE decreases steadily past day 29 for the red line (right), whereas for the regular group (middle), prediction is updated as expected cycle length is passed, as indicated by a bump in RMSE.

For the irregular users (right panel), model error (the red line) decreases steadily past cycle day 29, while for regular users (middle panel), there is a bump past cycle day 29 that decreases as day 60 approaches. This showcases how the proposed model characterizes the higher variability of irregular users by anticipating that the cycle may extend past day 29. For regular users, it's not as likely that the cycle will extend past what is expected, yet the model adapts to skipped cycles when they occur.

4 Conclusion

The ability to leverage mHealth data to improve predictions of menstrual cycle length is crucial to better understanding women's health. We have developed a model that takes into consideration the specific nature of mHealth data, namely the presence of self-tracking artifacts and the online nature of tracking, that allows for dynamically-updated predictions. By doing so, we are able to outperform several baselines. These insights have practical applications in mHealth – for instance, users may be alerted when they are suspected to have skipped tracking a cycle. Generative, predictive models for cycle lengths hold many promising avenues for further exploration. For instance, moving to per-regularity group instead of per-population inference may improve prediction. In addition, the model we've proposed here is the necessary first step to a proven hybrid deep learning approach [14] that models experimental details generatively and complex biological details in a deep way. While our model is able to outperform baselines by using only previous cycle lengths as predictors for next cycle length, a hybrid model where cycles are generated from the proposed generative model with hyperparameters learned by a deep model that uses both symptom (e.g., headache, abdominal pain) and cycle length information as input may improve individualized predictions further.

5 Acknowledgements

References

[1] Yi-Xin Wang, Mariel Arvizu, Janet W Rich-Edwards, Jennifer J Stuart, JoAnn E Manson, Stacey A Missmer, An Pan, and Jorge E Chavarro. Menstrual cycle regularity and length across the reproductive lifespan and risk of premature mortality: prospective cohort study. *bmj*, 371, 2020.

[2] Enrico Carmina and Rogerio A. Lobo. Polycystic Ovary Syndrome (PCOS): Arguably the Most Common Endocrinopathy Is Associated with Significant Morbidity in Women. *The Journal of Clinical Endocrinology & Metabolism*, 84(6):1897–1899, 1999.

[3] A. Zittermann, I. Schwarz, K. Scheld, T. Sudhop, H. K. Berthold, K. von Bergmann, H. van der Ven, and P. Stehle. Physiologic fluctuations of serum estradiol levels influence biochemical markers of bone resorption in young women. *The Journal of Clinical Endocrinology and Metabolism*, 85(1):95–101, January 2000.

[4] Caren G. Solomon, Frank B. Hu, Andrea Dunaif, Janet E. Rich-Edwards, Meir J. Stampfer, Walter C. Willett, Frank E. Speizer, and Joann E. Manson. Menstrual cycle irregularity and risk for future cardiovascular disease. *The Journal of Clinical Endocrinology and Metabolism*, 87(5):2013–2017, May 2002.

[5] V. B. Popat, T. Prodanov, K. A. Calis, and L. M. Nelson. The menstrual cycle: a biological marker of general health in adolescents. *Annals of the New York Academy of Sciences*, 1135:43–51, 2008.

[6] C Parente Barbosa, AM Bentes De Souza, B Bianco, and DM Christofolini. The effect of hormones on endometriosis development. *Minerva Ginecologica: A Journal on Obstetrics and Gynecology*, 63(4):375 – 386, Aug 2011.

[7] Lynne T. Shuster, Deborah J. Rhodes, Bobbie S. Gostout, Brandon R. Grossardt, and Walter A. Rocca. Premature menopause or early menopause: long-term health consequences. *Maturitas*, 65(2):161–166, February 2010.

[8] Emma Pierson, Tim Althoff, and Jure Leskovec. Modeling Individual Cyclic Variation in Human Behavior. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 107–116, Republic and Canton of Geneva, Switzerland, 2018. International World Wide Web Conferences Steering Committee.

[9] Laura Symul, Katarzyna Wac, Paula Hillard, and Marcel Salathé. Assessment of menstrual health status and evolution through mobile apps for fertility awareness. *Nature Digital Medicine*, 2(64), 2019.

[10] Jonathan R. Bull, Simon P. Rowland, Elina Berglund Scherwitzl, Raoul Scherwitzl, Kristina Gemzell Danielsson, and Joyce Harper. Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. *Nature Digital Medicine*, 2(83), 2019.

[11] Kathy Li, Iñigo Urteaga, Chris H. Wiggins, Anna Druet, Amanda Shea, Virginia J. Vitzthum, and Noémie Elhadad. Characterizing physiological and symptomatic variation in menstrual cycles using self-tracked mobile health data. *Nature Digital Medicine*, 3(79), 2020.

[12] Clue by BioWink GmbH, Adalbertstraße 7-8, 10999 Berlin, Germany. https://helloclue.com/, 2019.

[13] Virginia J. Vitzthum. The ecology and evolutionary endocrinology of reproduction in the human female. *American Journal of Physical Anthropology*, 140(S49):95–136, 2009.

[14] Wesley Tansey, Kathy Li, Haoran Zhang, Scott W. Linderman, Raul Rabadan, David M. Blei, and Chris H. Wiggins. Dose-response modeling in high-throughput cancer drug screenings: An end-to-end approach, 2020 (to appear in Biostatistics).

[15] Alan E. Treloar, Ruth E. Boynton, Borghild G. Behn, and Byron W. Brown. Variation of the human menstrual cycle through reproductive life. *International journal of fertility*, 12(1 Pt 2):77–126, 1967.

[16] Leonard Chiazze, Franklin T. Brayer, Jr. John J. Macisco, Margaret P. Parker, and Benefict J. Duffy. The Length and Variability of the Human Menstrual Cycle. *The Journal of the American Medical Association*, 203(6):377–380, 1968.

[17] Rebecca J Ferrell, Kathleen A O'Connor, German Rodriguez, Tristan Gorrindo, Darryl J Holman, Eleanor Brindle, Rebecca C Miller, Deborah E Schechter, Lauren Korshalla, James A Simon, Phyllis K Mansfield, James W Wood, and Maxine Weinstein. Monitoring reproductive aging in a 5-year prospective study: aggregate and individual changes in steroid hormones and menstrual cycle lengths with age. *Menopause*, 12:567–757, 2005.

[18] Siobán D. Harlow, Margery Gass, Janet E. Hall, Roger Lobo, Pauline Maki, Robert W. Rebar, Sherry Sherman, Patrick M. Sluss, Tobie J.de Villiers, and for the STRAW+10 Collaborative Group. Executive Summary of the Stages of Reproductive Aging Workshop + 10: Addressing the Unfinished Agenda of Staging Reproductive Aging. *The Journal of Clinical Endocrinology & Metabolism*, 97(4):1159–1168, 04 2012.

[19] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

Supplementary Information

Dataset cohort details

Our cohort consists of users aged 21-33 (because cycles are more likely to be ovulatory and less variable in their lengths during this age interval [15, 16, 17, 13, 18]) with natural menstrual cycles (i.e., no hormonal birth control or intrauterine device). To rule out cases that indicate insufficient engagement with the app, we remove users from our cohort who have only tracked two cycles and cycles for which the user has not provided period data within 90 days.

Generative process details

We provide details on the generative process for self-tracked mHealth cycle lengths, which draws per-user specific parameters from population level shared priors:

- **Observed variables:** Observed cycle length $d_{i,c}$, with $c = \{1, \cdots, C_i\}$ cycle lengths for each individual $i = \{1, \cdots, I\}$. Each true cycle length (for user i, cycle c, out of the number of skipped cycles j) is drawn from a Poisson distribution, $d_{i,j,c} \sim p(d_{i,j,c}|\lambda_i) = Pois(d_{i,j,c}|\lambda_i)$. The sum of independent Poissons is a different Poisson distribution, so the observed cycle length ($d_{i,c} = \sum_{j=0}^{s_{i,c}+1} d_{i,j,c}$) is also drawn from a Poisson, conditioned on the number of skipped cycles,

$$d_{i,c} \sim Pois(\lambda_i(s_{i,c} + 1)) . \tag{1}$$

- **Latent variables:** $s_{i,c}$ denotes the number of skipped (not reported) cycles, with $c = \{1, \cdots, C_i\}$ cycle lengths for each individual $i = \{1, \cdots, I\}$. The number of skipped cycles is drawn from a truncated Geometric distribution with a maximum number of skipped cycles S,

$$s_{i,c} \sim p(s|\pi_i) = \frac{\pi_i^s(1-\pi_i)}{\sum_{s=0}^{S} \pi_i^s(1-\pi_i)} = \frac{\pi_i^s}{\sum_{s=0}^{S} \pi_i^s} = \frac{\pi_i^s(1-\pi_i)}{(1-\pi_i^{(S+1)})} \quad \text{for } s \in \mathbb{N} . \tag{2}$$

- **Parameters λ_i:** the Poisson rate parameters for each individual $i = \{1, \cdots, I\}$. Per-user Poisson rate parameters λ_i are drawn from a population-level Gamma distribution

$$\lambda_i \sim p(\lambda|\kappa, \gamma) = \frac{\gamma^\kappa}{\Gamma(\kappa)} \lambda^{\kappa-1} e^{-\gamma\lambda} \quad \text{for } \lambda > 0 \text{ and } \kappa, \gamma > 0. \tag{3}$$

- **Hyperparameters of the Poisson rate parameter:** κ, γ of a Gamma distribution prior at the population level.

- **Parameters π_i:** the probability of skipping a cycle for each individual $i = \{1, \cdots, I\}$. The probability of an individual skipping a cycle is drawn from a population-level Beta distribution

$$\pi_i \sim p(\pi|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1}(1-\pi)^{\beta-1} , \quad \text{for } \pi \in [0, 1] \text{ and } \alpha, \beta > 0. \tag{4}$$

- **Hyperparameters of the geometric distribution parameters:** α, β of the population level Beta distribution prior on skipping probabilities.

Inference details

Given a dataset of C_i cycle lengths for I users, we perform hyperparameter inference via type-II maximum likelihood estimation. We compute a Monte Carlo importance sampling (IS) approximation to the negative log-likelihood $-\ln(p(d|u)) = -\ln(\sum_i p(d_i|u))$, due to the impossibility of integrating out the number of skipped cycles $s_{i,c}$ analytically. In particular, we compute an approximation to $p(d_i|u)$ with M samples,

$$p(d_i|u) = \frac{1}{M} \sum_m p(\theta_m|u) \frac{p(d_i|\theta_m)}{p(\theta_m|u_0)} = \sum_m w_m p(\theta_m|u) , \tag{5}$$

where $w_m = \frac{1}{M}\frac{p(d_i|\theta_m)}{p(\theta_m|u_0)}$, u represents the hyperparameters $[\alpha, \beta, \kappa, \gamma]$, and θ represents the parameters $[\lambda, \pi]$. Note that the IS weights w_m do not depend on the hyperaparameters u, and thus can be precomputed before inference (i.e., optimization). This method is less computationally demanding than a classical Monte Carlo approximation.

We compute the probability $p(d_i|\theta_m)$ by integrating out the probability of skipping $s_{i,c}$, which is drawn from a truncated geometric distribution as in Eqn. (2):

$$p(d_i|\theta_m) = \prod_{c=1}^{C_i} p(d_c|\theta_m) = \prod_{c=1}^{C_i}\sum_{s=0}^{S} p(d_c|\lambda_m, s)p(s|\pi_m) \tag{6a}$$

$$= \prod_{c=1}^{C_i}\sum_{s=0}^{S}((\lambda_m(s+1))^{d_c}e^{-\lambda_m(s+1)}/d_c!)\left(\frac{\pi_m^s(1-\pi_m)}{\sum_{s=0}^{S}\pi_m^s(1-\pi_m)}\right) \tag{6b}$$

$$= \prod_{c=1}^{C_i}\frac{\lambda_m^{d_c}e^{-\lambda_m}}{d_c!}\sum_{s=0}^{S}((s+1)^{d_c}e^{-\lambda_m s})\left(\frac{\pi_m^s}{\sum_{s=0}^{S}\pi_m^s}\right) \tag{6c}$$

$$= \prod_{c=1}^{C_i}\phi(\lambda_m)\frac{\sum_{s=0}^{S}(s+1)^{d_c}(\pi_m e^{-\lambda_m})^s}{\sum_{s=0}^{S}\pi_m^s} \tag{6d}$$

$$= \prod_{c=1}^{C_i}\phi(\lambda_m)\frac{\sum_{s=0}^{S}(s+1)^{d_c}(\pi_m e^{-\lambda_m})^s}{\frac{1-\pi_m^{S+1}}{1-\pi_m}} \tag{6e}$$

$$= \prod_{c=1}^{C_i}\frac{1-\pi_m}{1-\pi_m^{S+1}}\phi(\lambda_m)\sum_{s=0}^{S}(s+1)^{d_c}(\pi_m e^{-\lambda_m})^s \tag{6f}$$

where d_c represents one cycle length for a given user, C_i is the number of cycles for user i, S is the maximum value of s, and ϕ is the Poisson density.

Prediction details

In order to update our predictions of per-user cycle length as each subsequent day passes, we are interested in the posterior of the next reported cycle length d^*, conditioned on previous cycle lengths d_i for a user i and the day of the current cycle $d_{current}$,

$$p(d^*|d^* > d_{current}, d_i, \hat{u}) = \frac{p(d^*, d^* > d_{current}|d_i, \hat{u})}{p(d^* > d_{current}|d_i, \hat{u})} = \frac{p(d^*|d_i, \hat{u})I(d^* > d_{current})}{p(d^* > d_{current}|d_i, \hat{u})} \tag{7a}$$

where we explicitly indicate that $p(d^*, d^* > d_{current}|d_i, \hat{u}) = 0$ if $d^* \le d_{current}$.

We are interested in computing the expectation of the conditional predictive posterior,

$$E[p(d^*|d^* > d_{current}, d_i, \hat{u})] = \sum_{d^*} d^* p(d^*|d^* > d_{current}, d_i, \hat{u}) \tag{8a}$$

$$= \sum_{d^*} d^*\frac{p(d^*|d_i, \hat{u})I(d^* > d_{current})}{p(d^* > d_{current}|d_i, \hat{u})} \tag{8b}$$

$$= \frac{\sum_{d^*} d^* p(d^*|d_i, \hat{u})I(d^* > d_{current})}{p(d^* > d_{current}|d_i, \hat{u})} \tag{8c}$$

$$= \frac{\sum_{d^*>d_{current}} d^* p(d^*|d_i, \hat{u})}{p(d^* > d_{current}|d_i, \hat{u})} \tag{8d}$$

$$= \frac{\sum_{d^*=d_{current}+1}^{D} d^* p(d^*|d_i, \hat{u})}{\sum_{d^*=d_{current}+1}^{D} p(d^*|d_i, \hat{u})} \tag{8e}$$

The key term above is $p(d^*|d_i, \hat{u})$:

$$p(d^*|d_i, \hat{u}) = \frac{\int d\lambda d\pi q(\lambda)b(\pi)\sum_{s*} p(s^*|\pi)p(d^*|s^*, \lambda)p(d_i|\lambda, \pi)}{\int d\lambda d\pi q(\lambda)b(\pi)p(d_i|\lambda, \pi)} \ , \tag{9}$$

where d_i are the cycle lengths for a user i and s_i are the number of skipped cycles for a user, and d^*, s^* are the next reported cycle length and next number of skipped cycles, respectively. For the truncated geometric distribution on skipping probabilities, we compute the above as

$$p(d^*|d_i, \hat{u}) = \frac{\sum_{m=1}^{M} \frac{1-\pi_m}{1-\pi_m^{S+1}} \sum_{s^*=0}^{S} \pi_m^{s^*} p(d^*|s^*, \lambda_m)p(d_i|\lambda_m, \pi_m)}{\sum_{m=1}^{M} p(d_i|\lambda_m, \pi_m)} \ . \tag{10a}$$

We compute $p(d^*|d_i, \hat{u})$ from $d^* = 0$ to D, normalizing appropriately over d^* for each value of $d_{current}$, using $p(d_i|\lambda_m, \pi_m)$ (as specified in Eqn. (6) of the description of inference) and $p(d^*|s^*, \lambda_m) = Pois(\lambda(s^*+1))$ (i.e., the Poisson PMF). Note that we must normalize $p(d^*|s^*, \lambda)$ over $d^* = 0$ to D.

Implementation details

We optimize the negative loglikelihood $-\ln(p(d|u)) = -\ln(\sum_i p(d_i|u))$ with $p(d_i|u)$ as in Eqn. (6) with respect to hyperparameters u via stochastic gradient descent. Specifically, we utilize Adam [19], an adaptive gradient method. All models have been implemented using PyTorch with minibatches of size 100.

All neural network-based models are trained (with dropout) on the observed cycle lengths for the whole cohort. Predictions are based on each per-user available cycle lengths.

Since we sequentially predict next cycle length, our train-test split is over the number of cycle lengths available, i.e., we train the models on C cycles and predict the $C + 1$th cycle, where $C = \{2, 4, 6, 8\}$.

For reproducibility, we provide the settings for priors, learning rate, and other details for each of the models below:

- CNN: number of layers = 1, kernel size = 3, stride = 1, padding = 0, dilation = 1, nonlinearity = $tanh$, dropout = 0.9, training criterion = MSE, epoch convergence criteria as maximum number of epochs = 1000, loss convergence criteria $\epsilon_{loss} = 1e - 3$, optimizer = Adam, learning rate = 0.01.

- RNN: number of layers = 1, hidden size = 3, nonlinearity = $tanh$, dropout = 0.9, epoch convergence criteria as maximum number of epochs = 1000, loss convergence criteria $\epsilon_{loss} = 1e - 3$, optimizer = Adam, learning rate = 0.01.

- LSTM: same as above.

- Proposed model: $u_0 = [\kappa_0 = 180, \gamma_0 = 6, \alpha_0 = 2, \beta_0 = 20]$, $S = 100$ (for both inference and prediction), $M = 1000$ (for both inference and prediction), epoch convergence criteria as maximum number of epochs = 1000, loss convergence criteria $\epsilon_{loss} = 1e - 3$, optimizer = Adam, learning rate = 0.01.

- Proposed model (s=0): same as above, with $S = 100$ in inference but $S = 0$ for next-cycle length prediction.