Inferring the Outcome-Oriented Sleep States

Haoqi, Robert, Brandon

March 22, 2023

Abstract

An abstract.

We have n participants. Every participant has an outcome $Y^{(i)}$ and covariates $L^{(i)}$. The i-th participant has timeseries $\{S^{(i)}_{1:T^{(i)}}, X^{(i)}_{1:T^{(i)}}\}$. We simplify it to $\{\bar{S}, \bar{X}\}$ by dropping i and using a bar to represent the timeseries, where S_t and X_t are the AASM sleep stage and (representation of) signals of the t-th epoch respectively.

We assume there is a hidden timeseries $\bar{Z}$ that generates $\bar{X}$ and $\bar{S}$ (Figure 1). Our goal is to

- first, infer $\bar{Z}$ from the observed data $\{\bar{S}, \bar{X}, Y, L\}$;
- second, estimate causal estimand if $\bar{Z}$ is "intervened", what's the effect on Y.

In the sections below, we derive detailed steps to achieve the two goals in two conditions where the outcome happens in the future vs already exists. We finally provide a case study.

1 When the outcome happens in the future

The outcome is a timeseries $\bar{Y} = Y_{0:K} \in \mathbb{R}^{K+1}$ where $Y_k = 1$ means the outcome happens at time k, and $Y_k = 0$ means the outcome does not happen at time k. k is on a longer timescale, usually years or months, compared to t which is epochs in one night's sleep. We have $p(Y_0 = 0) = 1$ and $p(Y_{k+1} = 1 | Y_k = 1) = 1$ by definition.

The diagram is shown in Figure 1. Note that although we show Z_{t+1} is dependent on Z_t and not $Z_{1:t-1}$ (Markov property), this need not be true. More complicated techniques for sequences, such as a transformer or recurrent network, can be used.

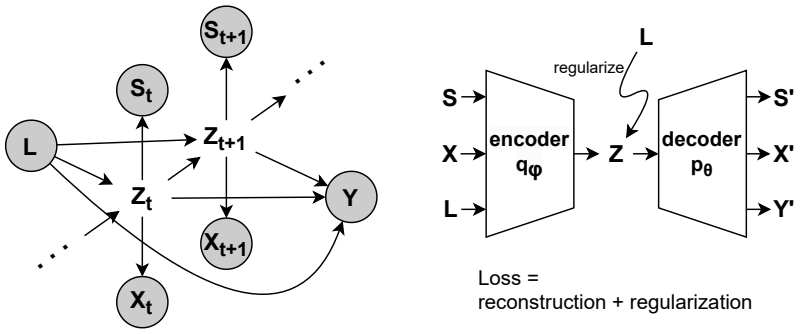

Figure 1: (left) Diagram for future outcome (observations are shaded). (right) The equivalent variational autoencoder [1].

1.1 Infer $\bar{Z}$

In amortized variational inference, we need to find the variational distribution as a function of observed data (encoder, q_ϕ) that maximizes the evidence lower bound (ELBO). We have

$$D_{KL}\left[q(Z|S, X, L)||p(Z|S, X, L, Y)\right]$$

$$= \sum_Z q(Z|S,X,L) \left[\log q(Z|S,X,L) - \log p(Z|S,X,L,Y)\right]$$
$$= \sum_Z q(Z|S,X,L) \left[\log q(Z|S,X,L) - \log p(Z,S,X,L,Y) + \log p(S,X,L,Y)\right] \geq 0 \ . \tag{1}$$

Therefore,

$$\log p(S,X,L,Y) \geq \text{ELBO}(p_\theta, q_\phi) = \sum_Z q(Z|S,X,L) \left[\log p(Z,S,X,L,Y) - \log q(Z|S,X,L)\right]$$
$$= \sum_Z q(Z|S,X,L) \Big[\underbrace{\log p(S|Z) + \log p(X|Z) + \log p(Y|L,Z)}_{\text{reconstruction error}} + \underbrace{\log p(Z|L) - \log q(Z|S,X,L)}_{\text{regularization}} \Big] + \log p(L)$$
$$= \mathbb{E}_{q_\phi(Z|S,X,L)} \left[f_{\theta,\phi}(Z)\right] + \text{constant} \ . \tag{2}$$

1.1.1 Loss function when Z is categorical

Differentiating ELBO w.r.t. ϕ is

$$\begin{aligned}\nabla_\phi \text{ELBO}(p_\theta, q_\phi) &= \nabla_\phi \mathbb{E}_{q_\phi(Z|S,X,L)} \left[f_{\theta,\phi}(Z)\right] \\ &= \mathbb{E}_{q_\phi(Z|S,X,L)} \left[(\nabla_\phi \log q_\phi(Z|S,X,L)) f_{\theta,\phi}(Z) + \nabla_\phi f_{\theta,\phi}(Z)\right] \\ &= \mathbb{E}_{q_\phi(Z|S,X,L)} \left[\nabla_\phi (\log q_\phi(Z|S,X,L) \widetilde{f_{\theta,\phi}(Z)} + f_{\theta,\phi}(Z))\right] \ . \end{aligned} \tag{3}$$

Differentiating ELBO w.r.t. θ is

$$\begin{aligned}\nabla_\theta \text{ELBO}(p_\theta, q_\phi) &= \nabla_\theta \mathbb{E}_{q_\phi(Z|S,X,L)} \left[f_{\theta,\phi}(Z)\right] \\ &= \mathbb{E}_{q_\phi(Z|S,X,L)} \left[\nabla_\theta f_{\theta,\phi}(Z)\right] \\ &= \mathbb{E}_{q_\phi(Z|S,X,L)} \left[\nabla_\theta (\log q_\phi(Z|S,X,L) \widetilde{f_{\theta,\phi}(Z)} + f_{\theta,\phi}(Z))\right] \ . \end{aligned} \tag{4}$$

Therefore, the surrogate loss function is

$$\frac{1}{M} \sum_{z_m \sim q_\phi(Z|S,X,L)} \log q_\phi(Z|S,X,L) \widetilde{f_{\theta,\phi}(Z)} + f_{\theta,\phi}(Z) \ , \tag{5}$$

where $\widetilde{f_{\theta,\phi}(Z)}$ means it's held constant.

If we assume the Markov temporal structure, the surrogate loss function can be further written as

$$\frac{1}{M} \sum_{z_m \sim \prod_t q_\phi(Z_t|Z_{t-1},S_t,X_t,L)} \sum_t \log q_\phi(Z_t|Z_{t-1},S_t,X_t,L) \widetilde{f_{\theta,\phi}(\bar{Z})} + f_{\theta,\phi}(\bar{Z}) \ , \tag{6}$$

where

$$\begin{aligned} f_{\theta,\phi}(\bar{Z}) &= \log p(\bar{S}|\bar{Z}) + \log p(\bar{X}|\bar{Z}) + \log p(\bar{Y}|L,\bar{Z}) + \log p(\bar{Z}|L) - \log q(\bar{Z}|\bar{S},\bar{X},L) \\ &= \sum_{t=1}^{T} \left[\log p(S_t|Z_t) + \log p(X_t|Z_t) + \log p(Z_t|Z_{t-1},L) - q(Z_t|Z_{t-1},S_t,X_t,L)\right] \\ &\qquad + \log p(Y_K|L,\bar{Z},Y_{1:K-1}=0) + \sum_{k=1}^{K-1} \log p(Y_k=0|L,\bar{Z},Y_{1:k-1}=0) \ . \end{aligned} \tag{7}$$

However, this estimator tends to have high variance. There are two approaches that can be used in combination.

First, only keep downstream of Z_t when estimating $\widetilde{f_{t,\theta,\phi}(\bar{Z})}$ in $\sum_t \log q_\phi(Z_t|Z_{t-1}, S_t, X_t, L)\widetilde{f_{t,\theta,\phi}(\bar{Z})}$

$$\widetilde{f_{t,\theta,\phi}(\bar{Z})} = \left(\sum_{s=t+2}^{T} \cdots\right) + \cdots . \tag{8}$$

Second, use baseline b, i.e., a running average of recent samples of $\widetilde{f_{t,\theta,\phi}(\bar{Z})}$

$$\log q_\phi(Z_t|Z_{t-1}, S_t, X_t, L)\left(\widetilde{f_{t,\theta,\phi}(\bar{Z})} - b\right) . \tag{9}$$

1.1.2 Loss function when Z is continuous

We use the reparameterization trick,

$$\begin{aligned}\text{ELBO}(p_\theta, q_\phi) &= \mathbb{E}_{q_\phi(\bar{Z}|\bar{S},\bar{X},L)}\left[f_{\theta,\phi}(\bar{Z})\right] \\ &= \mathbb{E}_{r(\bar{\epsilon})}\left[f_{\theta,\phi}(g(\bar{\epsilon}))\right] ,\end{aligned} \tag{10}$$

where

$$r(\bar{\epsilon}) = \{\mathcal{N}(0,1)\}_{t=1}^{T} \tag{11}$$

$$g(\bar{\epsilon}) = \{Z_t \sim g_\mu(Z_{t-1}, S_t, X_t, L) + \epsilon_t \cdot g_\sigma(Z_{t-1}, S_t, X_t, L)\}_{t=1}^{T} \tag{12}$$

$$\nabla_\theta \text{ELBO}(p_\theta, q_\phi) = \mathbb{E}_{r(\bar{\epsilon})}\left[\nabla_\theta f_{\theta,\phi}(g(\bar{\epsilon}))\right] \tag{13}$$

$$\nabla_\phi \text{ELBO}(p_\theta, q_\phi) = \mathbb{E}_{r(\bar{\epsilon})}\left[\nabla_\phi f_{\theta,\phi}(g(\bar{\epsilon}))\right] . \tag{14}$$

Therefore, the surrogate loss function is

$$\frac{1}{M}\sum_{\bar{\epsilon}\sim r_(\bar{\epsilon})} f_{\theta,\phi}(g(\bar{\epsilon})) . \tag{15}$$

1.2 Estimate causal estimand

We will use survival analysis for the time-to-event type of outcome and censoring. Here, the censoring occurs as (1) administrative stop of the follow-up or loss to follow-up, denoted as $\bar{C}$; and (3) death as the competing risk, denoted as $\bar{D}$.

The estimand is

$$\mathbb{E}[Y_{K+1}^{\bar{Z}=\bar{z},\bar{C}=0,\bar{D}=0}] , \tag{16}$$

which is the expectation of not developing the outcome until time $K+1$ since the baseline, if the hidden sleep states are set to $\bar{z}$ and there is no censoring and no competing risk.

Based on Young et al. [2], the estimand can be identified using g-formula,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=0}^{K} p(Y_{k+1}=1|L^{(i)}, \bar{Z}=\bar{z})\prod_{j=0}^{k} p(Y_{k+1}=0|L^{(i)}, \bar{Z}=\bar{z}) , \tag{17}$$

where the building block $p(Y_{k+1}|L, \bar{Z}=\bar{z})$ is estimated in Equation (**??**).

The identification assumptions are ??.

2 When the outcome is an existing condition

The outcome is a binary variable $Y \in \{0,1\}$ where $Y=1$ means the outcome is currently present, and $Y=0$ means the outcome is currently not present. The diagram is shown in Figure 2.

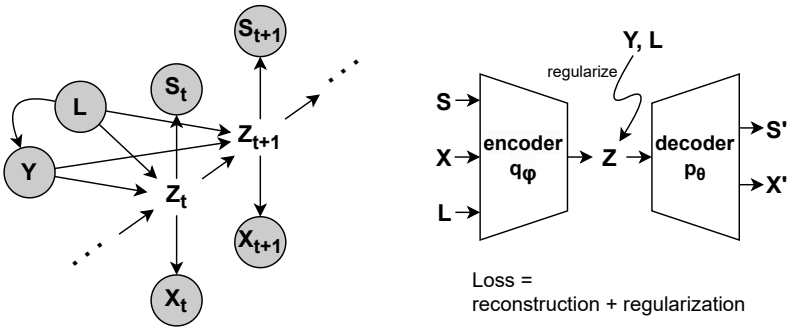

Figure 2: (left) Diagram for current outcome (observations are shaded). (right) The equivalent variational autoencoder [1].

2.1 Infer $\bar{Z}$

Similarly, we have

$$\begin{aligned}
&\log p(S,X,L,Y) \geq \text{ELBO}(p_\theta, q_\phi) = \sum_Z q(Z|S,X,L)\left[\log p(Z,S,X,L,Y) - \log q(Z|S,X,L)\right] \\
&= \sum_Z q(Z|S,X,L)\Big[\underbrace{\log p(S|Z) + \log p(X|Z)}_{\text{reconstruction error}} + \underbrace{\log p(Z|Y,L) - \log q(Z|S,X,L)}_{\text{regularization}}\Big] + \log p(Y,L) \\
&= \mathbb{E}_{q_\phi(Z|S,X,L)}\left[f_{\theta,\phi}(Z)\right] + \text{constant} \,. \qquad (18)
\end{aligned}$$

If we assume the Markov temporal structure, the surrogate loss function can be written as

$$\frac{1}{M}\sum_{z_m \sim \prod_t q_\phi(Z_t|Z_{t-1},S_t,X_t,L)} \sum_t \log q_\phi(Z_t|Z_{t-1},S_t,X_t,L)\widetilde{f_{\theta,\phi}(\bar{Z})} + f_{\theta,\phi}(\bar{Z}) \,, \qquad (19)$$

where

$$\begin{aligned}
f_{\theta,\phi}(\bar{Z}) &= \log p(\bar{S}|\bar{Z}) + \log p(\bar{X}|\bar{Z}) + \log p(\bar{Z}|Y,L) - \log q(\bar{Z}|\bar{S},\bar{X},L) \\
&= \sum_{t=1}^{T}\left[\log p(S_t|Z_t) + \log p(X_t|Z_t) - \log q(Z_t|Z_{t-1},S_t,X_t,L)\right] + \log p(\text{summary}(Z)|Y,L) \,. \qquad (20)
\end{aligned}$$

2.2 Estimate causal estimand

?

3 A case study

We study the case for determining the optimal thresholds for slow wave activity amplitude and percent in 30-second epoch for new N2 and N3, which maximizes its association between the percent of new N3 with the current condition of dementia. The current condition of dementia is defined according to criteria as in Ye et al. [3].

The specification is

- $S_t \in \{\text{W, R, N1, N2+N3}\}$
- $Z_t \in \{\text{W, R, N1, newN2, newN3}\}$
- $X_t \in \mathbb{R}^2$: SWA amplitude, SWA percent
- L is null for simplicity
- $Y \in \{0,1\}$: yes or no to current dementia (before or within 1 year after sleep study)

- $Z_t \sim q_\phi(Z_t|Z_{t-1}, S_t, X_t, L)$: if $S_t \in \{\text{W,R,N1}\}$, then $Z_t = S_t$; else if X_t(SWA amplitude) >amplitude threshold and X_t(SWA percent) >percent threshold, then newN3; else newN2
- $p_\theta(\text{summary}(Z)|Y, L)$ is $p_\theta(\text{new N3 \%}|Y, L) \sim \text{Beta}(\alpha = \exp(b_1^\top [Y\ L]), \beta = \exp(b_2^\top [Y\ L]))$
- $S_t \sim p_\theta(S_t|Z_t)$: if $Z_t \in \{\text{W,R,N1}\}$, then $S_t = Z_t$; else N2+N3
- $p_\theta(X_t|Z_t) \sim \text{LogNormal}(\mu(Z_t), \sigma^2(Z_t))$ for SWA amplitude and $\text{Beta}(\alpha(Z_t), \beta(Z_t))$ for SWA percent (with 0 and 1 squeezed by $(x(N-1)+0.5)/N$ [4])

The retrospective dataset has ? participants who underwent overnight diagnostic PSG sleep study at the MGH sleep clinic from ? to ?. The average age is ? years. ?% are females. The average BMI is ? kg/m^2. Over the ? years of follow-up, ? (?%) had been diagnosed with dementia until ?. There is IRB approval.

We fit the model and obtained the following results.

The optimal SWA amplitude threshold is ?μV. The optimal SWA percent threshold is ?%. The C-index for predicting dementia using the optimal thresholds is ? (?–?). The C-index using the conventional thresholds is ? (?–?). The C-index using the covariates L only is ? (?–?).

References

[1] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[2] Jessica G Young, Mats J Stensrud, Eric J Tchetgen Tchetgen, and Miguel A Hernán. A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in medicine*, 39(8):1199–1236, 2020.

[3] Elissa M Ye, Haoqi Sun, Parimala V Krishnamurthy, Noor Adra, Wolfgang Ganglberger, Robert J Thomas, Alice D Lam, and M Brandon Westover. Dementia detection from brain activity during sleep. *Sleep*, 46(3):zsac286, 2023.

[4] Michael Smithson and Jay Verkuilen. A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods*, 11(1):54, 2006.