Supplementary Material for "A Flexible Framework for Incorporating Patient Preferences into Q-Learning"

A Incorporating Model Selection into our Framework

Note that we require specifying a parametric model $P_\theta(\mathbf{H}_3|\mathbf{E})$ for $\Pr(\mathbf{H}_3|\mathbf{E})$ with parameter vector θ. In practice, we do not know in advance what $\Pr(\mathbf{H}_3|\mathbf{E})$ is. In some cases, an approximately correct model for $\Pr(\mathbf{H}_3|\mathbf{E})$ can be chosen using a combination of previous literature, statistical expertise and domain knowledge. In other cases, however, assuming a correctly-specified model for $\Pr(\mathbf{H}_3|\mathbf{E})$ can be a rather strong assumption. Instead, our methodology would be more robust if there was a way to select among multiple proposed parametric models $M_1(\mathbf{H}_3|\mathbf{E},\theta_1),...,M_P(\mathbf{H}_3,|\mathbf{E},\theta_P)$ for $\Pr(\mathbf{H}_3|\mathbf{E})$ and make a weaker assumption that only one of our models was correctly specified. Here θ_p denotes the parameter vector associated with the pth proposed parametric model $M_p(\mathbf{H}_3|\mathbf{E},\theta_p)$ for $\Pr(\mathbf{H}_3|\mathbf{E})$. One way to select among models is to partition our data as $\mathcal{D} = \mathcal{D}_T \cup \mathcal{D}_V$, train each parametric model $M_p(\mathbf{H}_3|\mathbf{E},\theta_p), 1 \le p \le P$ as $M_p(\mathbf{H}_3|\mathbf{E},\hat{\theta}_p)$ on training set $\mathcal{D}_T$ and then evaluate the estimated models using the observed log-likelihood $\sum_{\mathbf{H}_3 \in \mathcal{D}_V} \log \int_{\mathcal{E}} M_p(\mathbf{H}_3|\mathbf{E},\hat{\theta}_p) \Pr(\mathbf{E}) d\mathbf{E}$ on the held-out validation set $\mathcal{D}_V$. Under some reasonable identifiability conditions, the generalization observed log-likelihood will be minimized uniquely at the true probability model $\Pr(\mathbf{H}_3|\mathbf{E})$ making it a valid loss to tune hyperparameters and select models.

As discussed in the Theoretical Results section, there are many cases where $M_1(\mathbf{H}_3|\mathbf{E},\theta_1),...,M_P(\mathbf{H}_3,|\mathbf{E},\theta_P)$ are technically unknown but the associated parameter vectors $\theta_1,...,\theta_P$ can still be estimated by maximizing a partial likelihood. Specifically, suppose that $\Pr(\mathbf{H}_3|\mathbf{E}) = f(\mathbf{H}_3|\mathbf{E})g(\mathbf{H}_3)$ where $g(\mathbf{H}_3)$ is unknown and $f_1(\mathbf{H}_3|\mathbf{E},\theta_1),...,f_P(\mathbf{H}_3|\mathbf{E},\theta_P)$

are all parametric models for $f(\mathbf{H}_3|\mathbf{E})$. Let $M_p(\mathbf{H}_3|\mathbf{E},\theta_p) = f_p(\mathbf{H}_3|\mathbf{E},\theta_p)g(\mathbf{H}_3)$. As before, we can train each parametric model $M_p(\mathbf{H}_3|\mathbf{E},\theta_p), 1 \le p \le P$ as $M_p(\mathbf{H}_3|\mathbf{E},\hat{\theta}_p)$ on training set $\mathcal{D}_T$ by solving $\text{argmax}_\theta \log \sum_{\mathbf{H}_3 \in \mathcal{D}_T} \int_{\mathcal{E}} f_p(\mathbf{H}_3|\mathbf{E},\theta) \Pr(\mathbf{E}) d\mathbf{E}$, as this solution is equivalent to $\text{argmax}_\theta \log \sum_{\mathbf{H}_3 \in \mathcal{D}_T} \int_{\mathcal{E}} M_p(\mathbf{H}_3|\mathbf{E},\theta) \Pr(\mathbf{E}) d\mathbf{E}$. We can then evaluate the estimated models using the equivalent partial log-likelihood on the held-out validation set $\mathcal{D}_V$. The next lemma provides some theoretical justification for such a procedure, the proof of which is a straightforward application of Lemma 5.35 of Van der Vaart (1998). Here $F_p(\mathbf{H}_3|\mathbf{E}) = f_p(\mathbf{H}_3|\mathbf{E},\hat{\theta}_p)$ and $F_p(\mathbf{H}_3) = \int_{\mathcal{E}} F_p(\mathbf{H}_3|\mathbf{E}) \Pr(\mathbf{E}) d\mathbf{E}$.

Lemma A.1. *Suppose $\mathcal{F} = \{f_1, ..., f_P\}$ are models for $f(\mathbf{H}_3|\mathbf{E})$ where $F_p(\mathbf{H}_3|\mathbf{E})g(\mathbf{H}_3), 1 \le p \le P$ define valid probability measures. Suppose too that there exists an $F_p \in \mathcal{F}$ such that $F_p(\mathbf{H}_3|\mathbf{E}) = f(\mathbf{H}_3|\mathbf{E})$ and $F_p(\mathbf{H}_3) \neq F(\mathbf{H}_3)$ for every other $F \in \mathcal{F}$. Then $\text{argmax}_{F \in \mathcal{F}} \mathbb{E}\left[\log F(\mathbf{H}_3)\right] = f(\mathbf{H}_3|\mathbf{E})$.*

Another possibility is to use estimated satisfaction value $\mathbb{E}_{\mathcal{D}_V}\left[\frac{\pi_{p,1}(A_1|\mathbf{H}_1)\pi_{p,2}(A_2|\mathbf{H}_2)}{\mu_1(A_1|\mathbf{H}_1)\mu_2(A_2|\mathbf{H}_2)} B_2\right]$ to evaluate model M_p, where $\pi_p = (\pi_{p,1}, \pi_{p,2})$ is the length-2 DTR sequence estimated from our procedure with M_p chosen as the parametric model for $\Pr(\mathbf{H}_3|\mathbf{E})$ and $\mu = (\mu_1, \mu_2)$ is the behavioral policy. This is an importance sampling or inverse probability weighted estimator for $\mathbb{E}_{\pi_p}[B_2]$ (Precup et al., 2000; Zhao et al., 2015). An advantage of this approach is that it guarantees our estimated DTR yields decent reported satisfaction even if none of the proposed models for $\Pr(\mathbf{H}_3|\mathbf{E})$ are correctly specified. The drawback of this approach, however, is that it does not choose models based on accuracy of estimating $\Pr(\mathbf{H}_3|\mathbf{E})$ and will thus be a poor choice if estimation and inference on the posterior distribution of $\mathbf{E}$ is of interest. Moreover, as reported satisfaction is only a noisy and imperfect measure of the true latent utilities, using satisfaction-based value may fail to choose the model that maximizes the true utility-based value $V(\pi_p)$, particularly in finite-sample settings.

B Extended Details of Generative Model

Below we fully specify the standard generative model for BEST. For each variable, the specified distribution is conditioned on the variables previously introduced in the lines above. Thus for example, the specified distribution X_{1j}, Binomial$(n = 10, p = 0.5)$, is conditional on $\mathbf{V}$ and $W_{1j}, 1 \leq j \leq 12$ (and thus X_{1j} is independent of these variables).

$$\mathbf{V} \sim \mathcal{N}_2(0, \mathbf{I}),$$

$$\mathbf{E} = \text{SoftMax}(\mathbf{V}) = \frac{(e^{\mathbf{V}}, 1)}{1 + \text{sum}(e^{\mathbf{V}})},$$

$$W_{1j} \overset{iid}{\sim} \text{Bernoulli}(p = \beta_{0,j,1} + \beta_{1,j,1}^T \mathbf{V}), 1 \leq j \leq 12,$$

$$\mathbf{W}_1^R \sim \text{Cat}(p), \quad \Pr(\mathbf{W}_1^R = \mathbf{w}) = \frac{\exp(-\lambda_1 T(\mathbf{w}, \mathbf{E}^R))}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_1 T(\mathbf{v}, \mathbf{E}^R))},$$

$$X_{1j} \overset{iid}{\sim} \text{Bin}(n = 10, p = 0.5), 1 \leq j \leq 3,$$

$$A_1 \sim \text{Uniform}(\mathcal{A}_1),$$

$$W_{2j} \overset{ind}{\sim} \text{Bernoulli}(p = \beta_{0,j,2} + \beta_{1,j,2}^T \mathbf{V}), 1 \leq j \leq 12,$$

$$\mathbf{W}_2^R \sim \text{Cat}(p), \quad \Pr(\mathbf{W}_2^R = \mathbf{w}) = \frac{\exp(-\lambda_2 T(\mathbf{w}, \mathbf{E}^R))}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_2 T(\mathbf{v}, \mathbf{E}^R))},$$

$$X_{2j} \overset{ind}{\sim} \text{Bin}\left(n = 10, p = \sigma\left[\sum_{k \in \{1,2,3,4\}} \gamma_{0,k,j,1} I(A_1 = k) + \frac{X_{1j} - \mathbb{E}[X_{1j}]}{\sqrt{\text{Var}[\mathbf{X}_{ij}]}} \sum_{k \in \{1,2,3,4\}} \gamma_{1,k,j,1} I(A_1 = k)\right]\right), 1 \leq j \leq 3,$$

$$B_1 \sim \text{Categorical}(p), \quad \Pr(B_1 \leq k) = \sigma(\alpha_{0,k,1} - \alpha_{1,1} \mathbf{E}^T \mathbf{X}_2),$$

$$C \sim \text{Uniform}(\mathcal{C}),$$

$$A_2 = \begin{cases} A_1 & C = 1 \\ A_1 + a, \quad a \sim \text{Uniform}(\mathcal{A}_1 \setminus A_1) & C = 2 \text{ or } (A_1 = 1 \text{ and } C \geq 2) \\ B(A_1 + a) + (1 - B)a, \quad B \sim \text{Bernoulli}(0.5), a \sim \mathcal{U}(\mathcal{A}_1 \setminus A_1) & C = 3 \text{ and } A_1 \neq 1 \\ a, \quad a \sim \text{Uniform}(\mathcal{A}_1 \setminus A_1) & C = 4 \text{ and } A_1 \neq 1, \end{cases}$$

$$Y_j \overset{ind}{\sim} \text{Bin}\left(n = 10, p = \sigma\left[\sum_{k \in \{1,2,3,4\}} \gamma_{0,k,j,2} I(k \in A_2) + \frac{X_{2j} - \mathbb{E}_\mu[X_{2j}]}{\sqrt{\text{Var}_\mu[X_{2j}]}} \sum_{k \in \{1,2,3,4\}} \gamma_{1,k,j,2} I(k \in A_2)\right]\right), 1 \leq j \leq 3,$$

$$B_2 \sim \text{Categorical}(p), \quad \Pr(B_2 \leq k) = \sigma(\alpha_{0,k,2} - \alpha_{1,2} \mathbf{E}^T \mathbf{Y}),$$

where the parameters of our generator are themselves simulated as follows:

$$\theta = (\beta, \alpha, \gamma),$$

$$\beta_{0,k,1} = 0,\ \beta_{1,k,1} \overset{iid}{\sim} \mathcal{N}_2(0,1) \quad (1 \leq k \leq 12),$$

$$\beta_{0,k,2} = 0,\ \beta_{1,k,2} = \sqrt{0.8}\beta_{1,k,1} + \sqrt{0.2}\epsilon_{\beta,k},\ \epsilon_{\beta,k} \overset{iid}{\sim} \mathcal{N}_2(0,1) \quad (1 \leq k \leq 12),$$

$$\alpha_{1,1} = 0.5,\ \alpha_{0,k,1} = 0.75k \quad (1 \leq k \leq 6),$$

$$\alpha_{1,2} = 0.6,\ \alpha_{0,k,2} = \alpha_{0,k,1} + 0.5 \quad (1 \leq k \leq 6),$$

$$\lambda_1 = 0.5, \lambda_2 = 2,$$

$$\gamma_{0,i,j,1} \overset{iid}{\sim} N(0, 0.5^2),\ \gamma_{1,i,j,1} \overset{iid}{\sim} N(0,1) \quad (1 \leq i \leq 4, 1 \leq j \leq 3),\ \text{and}$$

$$(\gamma_{0,i,j,2}, \gamma_{1,i,j,2}) = \sqrt{0.8}(\gamma_{0,i,j,1}, \gamma_{0,i,j,2}) + \sqrt{0.2}\epsilon_{\gamma,i,j},\ \epsilon_{\gamma,i,j} \overset{iid}{\sim} \mathcal{N}_2(0, (0.5^2, 1)) \quad (1 \leq i \leq 4, 1 \leq j \leq 3).$$

Recall that our experiments are run over 10 seeds. For each seed, we first simulate model parameters from their respective priors described above, and then we simulate the dataset given the simulated model parameters.

For the ablation generative model, we set $\gamma_{1,k,j,2} = 0$, $\gamma_{0,k,0,2} = 0$ and $\gamma_{0,k,3,2} = -\gamma_{0,k,2,2}$. These modifications change the distribution of the outcomes to $Y_1 \sim \text{Binomial}(n = 10, p = 0.5)$, $Y_2 \sim \text{Binomial}(n = 10, p = \sigma(f(A_2)))$ and $Y_3 \sim \text{Binomial}(n = 10, p = \sigma(-f(A_2)))$ where $f(A_2) = \text{expit}(\Pr(Y_2|A_2, \mathbf{X}_2))$. Now Y_1 can be thought of as a random intercept in our utilities $\mathbf{E}^T\mathbf{Y}$, while Y_2 and Y_3 can be thought of as competing outcomes. Moreover, for each seed, we simulated $\gamma_{0,k,2,2}$ from $N(0,1)$instead of $N(0, 0.5^2)$, so as to increase effect sizes and reduce signal-to-noise ratio. As $\mathbf{X}_2$ is now independent of Y, it may appear that covariates are no longer relevant to the DTR. However, this is not the case: the observed data is still useful for estimating the expected value of $\mathbf{E}$, which determines the best sequence of treatments to take.

C Mathematical Proofs

Theorem C.1. *$\hat{\theta}_n \to_p \theta_0$ provided with probability one and all $\theta \in \Theta$: (C1) $\Pr(\mathbf{H}_3|\mathbf{E}) = M_{\theta_0}(\mathbf{H}_3|\mathbf{E})$ for some $\theta_0 \in \Theta$; (C2) $M_\theta(\mathbf{H}_3|\mathbf{E})$ is continuous in θ; (C3) $|M_\theta(\mathbf{H}_3|\mathbf{E})| < F(\mathbf{H}_3|\mathbf{E})$ for some $\mathbb{E}_{\theta_0}[F(\mathbf{H}_3|\mathbf{E})] < \infty$; (C4) $|\log M_\theta(\mathbf{H}_3)| \leq F(\mathbf{H}_3)$ for some $\mathbb{E}_{\theta_0}[F(\mathbf{H}_3)] < \infty$; (C5) $M_{\theta_0}(\mathbf{H}_3) \neq M_\theta(\mathbf{H}_3)$ for all $\theta \neq \theta_0$. Moreover, $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d \mathcal{N}(0, I(\theta_0)^{-1})$, provided with probability one and all $\theta_1, \theta_2 \in N_\epsilon(\theta_0) = \{\theta : ||\theta - \theta_0||_2 < \epsilon\}$: (N1) $I(\theta_0)$ is non-singular; (N2) $|M_{\theta_1}(\mathbf{H}_3|\mathbf{E}) - M_{\theta_2}(\mathbf{H}_3|\mathbf{E})| \leq F(\mathbf{H}_3|\mathbf{E})||\theta_1 - \theta_2|$ for some $\mathbb{E}_{\theta_0,\mathbf{E}}[F^2(\mathbf{H}_3|\mathbf{E})] < \infty$; (N3) $M_\theta(\mathbf{H}_3) > c$ for some $c > 0$; (N4) $M_{\theta_1}(\mathbf{H}_3|\mathbf{E})$ is continuously differentiable; (N5) $||\nabla_\theta M_{\theta_1}(\mathbf{H}_3|\mathbf{E})||_\infty < G(\mathbf{H}_3|\mathbf{E})$ for some $\mathbb{E}_{\theta_0,\mathbf{E}}[G^2(\mathbf{H}_3|\mathbf{E})] < \infty$.*

Proof. By Theorem 5.7 of Van der Vaart (1998), $\hat{\theta}_n \to_p \theta_0$ provided (A1) $L_n(\hat{\theta}_n) \geq L_n(\theta_0) - o_p(1)$; (A2) $\sup_{\theta: d(\theta,\theta_0) \geq \epsilon} M(\theta) < M(\theta_0)$ for all $\epsilon > 0$; (A3) $\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \to_p 0$. Define $L_n(\theta) = \sum_{i=1}^n \log M_\theta(\mathbf{H}_3^i)$ and $L(\theta) = \mathbb{E}\left[\log M_\theta(\mathbf{H}_3^i)\right]$. Assumption (A1) is satisfied by definition of $\hat{\theta}_n$.

By (C1), $\Pr(\mathbf{H}_3) = \int_{\mathcal{E}} \Pr(\mathbf{H}_3|\mathbf{E}) \Pr(\mathbf{E}) d\mathbf{E} = \int_{\mathcal{E}} M_{\theta_0}(\mathbf{H}_3|\mathbf{E}) \Pr(\mathbf{E}) d\mathbf{E} = M_{\theta_0}(\mathbf{H}_3)$. Thus by (C1) and (C5), we have by Lemma 5.35 of Van der Vaart (1998) that $L(\theta)$ is uniquely maximized at θ_0. By (C3), it must be that $M_\theta(\mathbf{H}_3) = \int_{\mathcal{E}} M_\theta(\mathbf{H}_3|\mathbf{E}) \Pr(\mathbf{E}) d\mathbf{E} < \infty$ almost surely, and thus by (C2) and the dominated convergence theorem, $M_\theta(\mathbf{H}_3)$ is continuous in θ almost surely. By (C4), $\log M_\theta(\mathbf{H}_3) < \infty$ almost surely, which means $M_\theta(\mathbf{H}_3) > 0$ almost surely, which means $\log M_\theta(\mathbf{H}_3)$ is continuous in θ almost surely. Therefore, by compactness of Θ and (C5), assumption (A2) is satisfied by Problem 5.27 of Van der Vaart (1998). Finally, by (almost-sure) continuity of $\log M_\theta(\mathbf{H}_3)$, (C4) and compactness of Θ, we have by example 19.8 of Van der Vaart (1998) that $\{M_\theta(\mathbf{H}_3) : \theta \in \Theta\}$ defines a P_{θ_0}-Glivenko-Cantelli class. Therefore assumption (A3) is satisfied.

By Theorem 5.39, $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d \mathcal{N}(0, I(\theta_0)^{-1})$ provided (B1) $\hat{\theta}_n \to_p \theta_0$, (B2) $I(\theta_0)$

is non-singular, (B3) $\log M_\theta(\mathbf{H}_3)$ is Lipschitz continuous in an neighborhood of θ_0 with Lipschitz constant $F(\mathbf{H}_3)$ square-integrable and (B4) $M_\theta(\mathbf{H}_3)$ is Hellinger differentiable. (B2) is satisfied by assumption and (B1) is satisfied by conditions (C1)-(C5).

By (N2) and Jensen's inequality, $|M_{\theta_1}(\mathbf{H}_3) - M_{\theta_2}(\mathbf{H}_3)| \leq \mathbb{E}_E[F(\mathbf{H}_3|\mathbf{E})]||\theta_1 - \theta_2|$ and by (N3), $\log'(M_\theta(\mathbf{H}_3))$ is bounded by $1/c$. As the composition of Lipschitz continuous functions are also Lipschitz continuous with the Lipschitz constant being the product of those of the composing functions (Shalev-Shwartz and Ben-David, 2014), $|\log M_{\theta_1} - \log M_{\theta_2}(\mathbf{H}_3)| \leq \frac{1}{c}\mathbb{E}_E[F(\mathbf{H}_3|\mathbf{E})]||\theta_1 - \theta_2|$ with $\mathbb{E}_{\theta_0}\frac{1}{c}\mathbb{E}_E[F(\mathbf{H}_3|\mathbf{E})] < \infty$. Thus condition (B3) is satisfied.

By (N5), $\mathbb{E}_E[G(\mathbf{H}_3|\mathbf{E})] < \infty$ almost surely. Thus we have by the Leibniz integral theorem and (N4) that $\nabla_\theta M_\theta(\mathbf{H}_3) = \mathbb{E}_E[\nabla_\theta M_\theta(\mathbf{H}_3|\mathbf{E})]$, and we have by the dominated convergence theorem that $\nabla_\theta M_\theta(\mathbf{H}_3)$ is continuous. As $\nabla_\theta \sqrt{M_\theta(\mathbf{H}_3)} = \frac{1}{\sqrt{M_\theta(\mathbf{H}_3)}}\sqrt{M_\theta(\mathbf{H}_3)}$ and $M_\theta(\mathbf{H}_3) > c$ for some $c > 0$ by (N3), $\nabla_\theta\sqrt{M_\theta(\mathbf{H}_3)}$ is also continuous. Finally, under our assumptions $I(\theta) = \mathbb{E}_{\theta_0}\left[\frac{1}{M_\theta(\mathbf{H}_3)^2}\mathbb{E}_E[\nabla_\theta M_\theta(\mathbf{H}_3|\mathbf{E})]\mathbb{E}_E[\nabla_\theta M_\theta(\mathbf{H}_3|\mathbf{E})]^T\right]$ with each element of this matrix bounded by $\frac{1}{c}\mathbb{E}_E^2[G(\mathbf{H}_3|\mathbf{E})] \leq \frac{1}{c}\mathbb{E}_E[G^2(\mathbf{H}_3|\mathbf{E})]$ and $\frac{1}{c}\mathbb{E}_{\mathbf{E},\theta_0}[G^2(\mathbf{H}_3|\mathbf{E})] < \infty$, and thus using the dominated convergence theorem once more, we have that $I(\theta) = \mathbb{E}_{\theta_0}\left[\frac{1}{M_\theta(\mathbf{H}_3)^2}\mathbb{E}_E[\nabla_\theta M_\theta(\mathbf{H}_3|\mathbf{E})]\mathbb{E}_E[\nabla_\theta M_\theta(\mathbf{H}_3|\mathbf{E})]^T\right]$ is continuous. As $\sqrt{M_\theta(\mathbf{H}_3)}$ is continuously differentiable and $I(\theta)$ is continuous, we have by lemma 7.6 of Van der Vaart (1998) that $M_\theta(\mathbf{H}_3)$ is Hellinger differentiable, satisfying condition (B4). $\square$

Theorem C.2. *$V(\widehat{\pi}_n) - V(\pi^*) \to_p 0$ provided with probability one; (V1) $||\widehat{\mathbb{E}}_n[\mathbf{E}|\mathbf{H}_2] - \mathbb{E}[\mathbf{E}|\mathbf{H}_2]||_{P_{\theta_0}} \to_p 0$; (V2) $||\widehat{\mathbb{E}}_n[\mathbf{Y}|\mathbf{H}_2, A_2] - \mathbb{E}[\mathbf{Y}|\mathbf{H}_2, A_2]||_{P_{\theta_0}} \to_p 0$; (V3) $\widehat{\mathbb{E}}_n[\max_{A_2}\widehat{Q}_{n,2}(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1] - \mathbb{E}[\max_{A_2}\widehat{Q}_{n,2}(\mathbf{H}_1, A_1)|\mathbf{H}_1, A_1]|||_{P_{\theta_0}} \to_p 0$; (V4) $\Pr(A_2|\mathbf{H}_2), \Pr(A_1|\mathbf{H}_1) > \epsilon$ for some $\epsilon > 0$.*

Proof. Note that:

$$\begin{aligned}
||\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)||_{P_{\theta_0}} =& ||\widehat{\mathbb{E}}_n[\mathbf{E}|\mathbf{H}_2]^T \widehat{\mathbb{E}}_n[\mathbf{Y}|\mathbf{H}_2, A_2] - \mathbb{E}[\mathbf{E}|\mathbf{H}_2]^T \mathbb{E}[\mathbf{Y}|\mathbf{H}_2, A_2]||_{P_{\theta_0}} \\
=& ||(\widehat{\mathbb{E}}_n[\mathbf{E}|\mathbf{H}_2] - \mathbb{E}[\mathbf{E}|\mathbf{H}_2])^T \widehat{\mathbb{E}}_n[\mathbf{Y}|\mathbf{H}_2, A_2] \\
&+ \mathbb{E}[\mathbf{E}|\mathbf{H}_2]^T (\widehat{\mathbb{E}}_n(\mathbf{Y}|\mathbf{H}_2, A_2) - \mathbb{E}(\mathbf{Y}|\mathbf{H}_2, A_2))||_{P_{\theta_0}} \\
\leq& ||\widehat{\mathbb{E}}_n[\mathbf{E}|\mathbf{H}_2] - \mathbb{E}[\mathbf{E}|\mathbf{H}_2]||_{P_{\theta_0}}^T ||\widehat{\mathbb{E}}_n[\mathbf{Y}|\mathbf{H}_2, A_2]||_\infty \\
&+ ||\mathbb{E}[\mathbf{E}|\mathbf{H}_2]||_\infty^T ||\widehat{\mathbb{E}}_n(\mathbf{Y}|\mathbf{H}_2, A_2) - \mathbb{E}(\mathbf{Y}|\mathbf{H}_2, A_2)||_{P_{\theta_0}},
\end{aligned}$$

where the last inequality uses Minkowski's inequality. As both $\mathbb{E}[\mathbf{E}|\mathbf{H}_2]$ and $\widehat{\mathbb{E}}_n[\mathbf{Y}|\mathbf{H}_2, A_2]$ are assumed to be strictly bounded, we have by (V1), (V2) and Slutsky's theorem that the term following the inequality above converges to zero in probability. We also note the obvious property that if $Y_n \to_p 0$ and $|X_n| < Y_n$, then $X_n \to_p 0$ as well. Therefore $||\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)||_{P_{\theta_0}} \to_p 0$ as well. Let $\mathcal{A}_2$ be the set of all possible values that could be in $\mathcal{A}_{\mathbf{H}_2}$ for some $\mathbf{H}_2 \in \mathcal{H}_2$. Note that, by (V4):

$$\begin{aligned}
&\sqrt{\mathbb{E}_{\mathbf{H}_2 \sim P_{\theta_0}, A_2 \sim \mathcal{U}(\mathcal{A}_{\mathbf{H}_2})} \left[\left(\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2) \right)^2 \right]} \\
=& \sqrt{\mathbb{E}_{\mathbf{H}_2, A_2 \sim P_{\theta_0}} \left[\frac{\mathcal{U}(\mathcal{A}_{\mathbf{H}_2})}{\Pr(A_2|\mathbf{H}_2)} \left(\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2) \right)^2 \right]} \\
\leq& 1/\sqrt{|\mathcal{A}_2|\epsilon} \sqrt{\mathbb{E}_{\mathbf{H}_2, A_2 \sim P_{\theta_0}} \left[\left(\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2) \right)^2 \right]}.
\end{aligned}$$

As $||\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)||_{P_{\theta_0}} \to_p 0$, $||\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)||_{P_{\theta_0}(\mathbf{H}_2)\mathcal{U}(\mathcal{A}_{\mathbf{H}_2})} \to_p 0$ as well. It is also easy to see that $\left|\left| \max_{A_2} |\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)| \right|\right|_{P_{\theta_0}}^2 \leq |\mathcal{A}_2| \times ||\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)||_{P_{\theta_0}(\mathbf{H}_2)\mathcal{U}(\mathcal{A}_{\mathbf{H}_2})}^2$, which means that $\left|\left| \max_{A_2} |\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)| \right|\right|_{P_{\theta_0}} \to_p 0$.

Note that:

$$
\begin{aligned}
&\big|\big| \max_{A_2} |\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)| \big|\big|^2_{P_{\theta_0}} \\
&= \mathbb{E}\left[\max_{A_2} |\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)|^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\max_{A_2} |\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)|^2 \big| \mathbf{H}_1, A_1\right)\right] \\
&\geq \mathbb{E}\left[\mathbb{E}^2\left(\max_{A_2} |\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)| \big| \mathbf{H}_1, A_1\right)\right] \\
&= \big|\big|\mathbb{E}\left(\max_{A_2} |\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)| \big| \mathbf{H}_1, A_1\right)\big|\big|^2_{P_{\theta_0}} \\
&\geq \big|\big|\mathbb{E}\left(\max_{A_2} \widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - \max_{A_2} Q_2(\mathbf{H}_2, A_2) \big| \mathbf{H}_1, A_1\right)\big|\big|^2_{P_{\theta_0}}.
\end{aligned}
$$

The first inequality comes from Jensen's inequality. As $\big|\big| \max_{A_2} |\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)| \big|\big|_{P_{\theta_0}} \to_p 0$, $\big|\big|\mathbb{E}\left[\max_{A_2} \widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - \max_{A_2} Q_2(\mathbf{H}_2, A_2)| \big| \mathbf{H}_1, A_1\right]\big|\big|_{P_{\theta_0}} \to_p 0$ as well.

Note that, by Minkowski's inequality:

$$
\begin{aligned}
||\widehat{Q}_{n,1} - Q_1||_{P_{\theta_0}} &= ||\widehat{\mathbb{E}}_n[\max_{A_2} \widehat{Q}_{n,2}(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1] - \mathbb{E}[\max_{A_2} Q_2(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1]||_{P_{\theta_0}} \\
\leq\ & ||\widehat{\mathbb{E}}_n[\max_{A_2} \widehat{Q}_{n,2}(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1] - \mathbb{E}[\max_{A_2} \widehat{Q}_{n,2}(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1]||_{P_{\theta_0}} \\
& + ||\mathbb{E}[\max_{A_2} \widehat{Q}_{n,2}(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1] - \mathbb{E}[\max_{A_2} Q_2(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1]||_{P_{\theta_0}}.
\end{aligned}
$$

As $\big|\big|\mathbb{E}\left[\max_{A_2} \widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - \max_{A_2} Q_2(\mathbf{H}_2, A_2)| \big| \mathbf{H}_1, A_1\right]\big|\big|_{P_{\theta_0}} \to_p 0$, we have by (V3) and Slutsky's theorem that $||\widehat{Q}_{n,1} - Q_1||_{P_{\theta_0}} \to_p 0$.

We have thus far proven or assumed the following:

$$
\begin{aligned}
&\big|\big|\widehat{Q}_{n,1}(\mathbf{H}_1, A_1) - Q_1(\mathbf{H}_1, A_1)\big|\big|_{P_{\theta_0}} \to_p 0, \\
&\big|\big|\widehat{\mathbb{E}}_n[\max_{A_2} \widehat{Q}_{n,2}(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1] - \mathbb{E}[\max_{A_2} \widehat{Q}_{n,2}(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1]\big|\big|_{P_{\theta_0}} \to_p 0, \text{ and} \\
&\left|\left|\mathbb{E}\left[\max_{A_2} |\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)| \big| \mathbf{H}_1, A\right]\right|\right|_{P_{\theta_0}} \to_p 0.
\end{aligned}
$$

Therefore, as $\Pr(A_1|\mathbf{H}_1) > \epsilon$ almost surely by (V4), it is easy to see that:

$$\big|\big|\max_{A_1}|\widehat{Q}_{n,1}(\mathbf{H}_1, A_1) - Q_1(\mathbf{H}_1, A_1)|\big|\big|_{P_{\theta_0}} \to_p 0, \tag{1}$$

$$\Bigg|\Bigg|\max_{A_1}\big|\widehat{\mathbb{E}}_n[\max_{A_2}\widehat{Q}_{n,2}(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1] - \mathbb{E}[\max_{A_2}\widehat{Q}_{n,2}(\mathbf{H}_2, A_2)|\mathbf{H}_1, A_1]\big|\Bigg|\Bigg|_{P_{\theta_0}} \to_p 0, \tag{2}$$

$$\Bigg|\Bigg|\max_{A_1}\mathbb{E}\left[\max_{A_2}|\widehat{Q}_{n,2}(\mathbf{H}_2, A_2) - Q_2(\mathbf{H}_2, A_2)|\Big|\mathbf{H}_1, A_1\right]\Bigg|\Bigg|_{P_{\theta_0}} \to_p 0. \tag{3}$$

Finally, note that:

$$\begin{aligned}
&V(\pi^*) - V(\hat{\pi}_n) \\
&= \big|\big|\max_{A_1} Q_1(\mathbf{H}_1, A_1) - \mathbb{E}[Q_2(\mathbf{H}_2, \hat{\pi}_{n,2}(\mathbf{H}_2))|\mathbf{H}_1, \hat{\pi}_{n,1}(\mathbf{H}_1)]\big|\big|_{P_{\theta_0}} \\
&= \big|\big|\max_{A_1} Q_1(\mathbf{H}_1, A_1) - \widehat{Q}_{n,1}(\mathbf{H}_1, \pi_1^*(A_1)) \\
&\quad + \widehat{Q}_{n,1}(\mathbf{H}_1, \pi_1^*(A_1)) - \mathbb{E}[Q_2(\mathbf{H}_2, \hat{\pi}_{n,2}(\mathbf{H}_2))|\mathbf{H}_1, \hat{\pi}_{n,1}(\mathbf{H}_1)]\big|\big|_{P_{\theta_0}} \\
&\leq \big|\big|\max_{A_1} Q_1(\mathbf{H}_1, A_1) - \widehat{Q}_{n,1}(\mathbf{H}_1, \pi_1^*(A_1) \\
&\quad + \widehat{Q}_{n,1}(\mathbf{H}_1, \hat{\pi}_{n,1}(A_1)) - \mathbb{E}[Q_2(\mathbf{H}_2, \hat{\pi}_{n,2}(\mathbf{H}_2))|\mathbf{H}_1, \hat{\pi}_{n,1}(\mathbf{H}_1)]\big|\big|_{P_{\theta_0}} \\
&\leq \big|\big|Q_1(\mathbf{H}_1, \pi_1^*(A_1)) - \widehat{Q}_{n,1}(\mathbf{H}_1, \pi_1^*(A_1))\big|\big|_{P_{\theta_0}} \\
&\quad + \big|\big|\widehat{Q}_{n,1}(\mathbf{H}_1, \hat{\pi}_{n,1}(A_1)) - \mathbb{E}[Q_2(\mathbf{H}_2, \hat{\pi}_{n,2}(\mathbf{H}_2))|\mathbf{H}_1, \hat{\pi}_{n,1}(\mathbf{H}_1)]\big|\big|_{P_{\theta_0}} \\
&\leq \big|\big|Q_1(\mathbf{H}_1, \pi_1^*(A_1)) - \widehat{Q}_{n,1}(\mathbf{H}_1, \pi_1^*(A_1))\big|\big|_{P_{\theta_0}} \\
&\quad + \big|\big|\widehat{\mathbb{E}}_n[\widehat{Q}_2(\mathbf{H}_2, \hat{\pi}_{n,2}(\mathbf{H}_2))|\mathbf{H}_1, \hat{\pi}_{n,1}(\mathbf{H}_1)] - \mathbb{E}[\widehat{Q}_2(\mathbf{H}_2, \hat{\pi}_{n,2}(\mathbf{H}_2))|\mathbf{H}_1, \hat{\pi}_{n,1}(\mathbf{H}_1)]\big|\big|_{P_{\theta_0}} \\
&\quad + \big|\big|\mathbb{E}[\widehat{Q}_2(\mathbf{H}_2, \hat{\pi}_{n,2}(\mathbf{H}_2))|\mathbf{H}_1, \hat{\pi}_{n,1}(\mathbf{H}_1)] - \mathbb{E}[Q_2(\mathbf{H}_2, \hat{\pi}_{n,2}(\mathbf{H}_2))|\mathbf{H}_1, \hat{\pi}_{n,1}(\mathbf{H}_1)]\big|\big|_{P_{\theta_0}}.
\end{aligned}$$

The first inequality uses the fact that the term inside the $||\cdot||_{P_{\theta_0}}$ operator on the second line is equal to $\mathbb{E}_{\pi^*}[U|\mathbf{H}_1] - \mathbb{E}_{\hat{\pi}_n}[U|\mathbf{H}_1]$, which is always non-negative, and the fact that $\hat{Q}_{n,1}(\mathbf{H}_1, \hat{\pi}_{n,1}(A_1)) \geq \hat{Q}_{n,1}(\mathbf{H}_1, \pi^*(\mathbf{H}_1))$ by definition of $\hat{\pi}_n$. The second and third inequalities comes from Minkowski's inequality. Therefore, by equations (1), (2) and (3), and Slutsky's theorem, we can thus conclude that $V(\pi^*) - V(\hat{\pi}_n) \to_p 0$. □

Theorem C.3. *$\hat{\theta}_n \to_p \theta_0$ and $||\widehat{\mathbb{E}}_n[\mathbf{E}|\mathbf{H}_2] - \mathbb{E}[\mathbf{E}|\mathbf{H}_2]||_{P_{\theta_0}} \to_p 0$ provided with probability one that $\Pr_{\theta_0}(\mathbf{H}_3|\mathbf{V}) = \Pr(\mathbf{H}_3|\mathbf{V})$ for some $\theta_0 \in \Theta$, $\Pr_{\theta_0}(\mathbf{H}_3) \neq \Pr_\theta(\mathbf{H}_3)$ for all $\theta \neq \theta_0$ and $g(\mathbf{H}_3) > c$ for some $c > 0$. Moreover, $\sqrt{n}(\hat{\theta} - \theta_0) \to_d \mathcal{N}(0, I(\theta_0)^{-1})$ provided $I(\theta_0)$ is non-singular.*

Proof. To show that $\hat{\theta}_n \to_p \theta_0$, we require (C1) $\Pr(\mathbf{H}_3|\mathbf{E}) = p_{\theta_0}(\mathbf{H}_3|\mathbf{E})$ for some $\theta_0 \in \Theta$; (C2) $p_\theta(\mathbf{H}_3|\mathbf{E})$ is continuous in θ; (C3) $p_\theta(\mathbf{H}_3|\mathbf{E}) < F(\mathbf{H}_3|\mathbf{E})$ for some $\mathbb{E}_{\theta_0}[F(\mathbf{H}_3|\mathbf{E})] < \infty$; (C4) $|\log p_\theta(\mathbf{H}_3)| \leq F(\mathbf{H}_3)$ for some $\mathbb{E}_{\theta_0}[F(\mathbf{H}_3)] < \infty$; (C5) $p_{\theta_0}(\mathbf{H}_3) \neq M_\theta(\mathbf{H}_3)$ for all $\theta \neq \theta_0$. (C1) and (C5) is assumed. It is easy to verify that $\Pr(\mathbf{W}_1|\mathbf{V}, \beta)$, $\Pr(\mathbf{W}_1^R|\mathbf{V}, \lambda)$, $\Pr(B_1|\mathbf{E}^T\mathbf{X}_2, \alpha)$, $\Pr(\mathbf{W}_2|\mathbf{V}, \beta)$, $\Pr(\mathbf{W}_2^R|\mathbf{V}, \lambda)$ and $\Pr(B_2|\mathbf{E}^T\mathbf{Y}, \alpha)$ are all continuous wrt θ. Let $f_\theta(\mathbf{H}_3|\mathbf{E})$ be the product of these terms and note that $f_\theta(\mathbf{H}_3) = \mathbb{E}_E[f_\theta(\mathbf{H}_3|\mathbf{E})]$ and $p_\theta(\mathbf{H}_3|\mathbf{E}) = f_\theta(\mathbf{H}_3|\mathbf{E})g(\mathbf{H}_3)$ As the product of continuous functions is continuous, (C2) is therefore satisfied. Moreover, $(\mathbf{W}_t, \mathbf{W}_t^R, B_t)_{1 \leq t \leq 2}$ is categorical and it is easy to verify that $p_\theta(\mathbf{H}_3|\mathbf{E}) \leq 1$ always, which means (C3) is satisfied.

Showing (C4) is a bit trickier. Recall that we assume some large pre-specified $B < \infty$ such that $||\theta||_\infty \leq B$ and some small $\epsilon > 0$ such that $\alpha_{1,t}, \lambda_t, \alpha_{0,j+1,t} - \alpha_{0,j,t} \geq \epsilon$, for all $\theta \in \Theta$. Then it can be seen that for all $V \in \mathbb{R}^2$, $\min_{\theta, \mathbf{H}_3 \in \Theta \times \mathcal{H}_3} \Pr(W_{tk}|\mathbf{V}, \theta) \geq 1 - \sigma(B + B|\mathbf{V}|) = \sigma\left[-B(|\mathbf{V}| + 1)\right], \forall (t,k) \in \{1,2\} \times \{1, ..., 12\}$, $\min_{\theta, \mathbf{H}_3 \in \Theta \times \mathcal{H}_3} \Pr(\mathbf{W}_t^R|\mathbf{E}^R, \theta) \geq \exp(-3B)/6 \equiv B_\lambda > 0$ and $\min_{\theta, \mathbf{H}_3 \in \Theta \times \mathcal{H}_3} \Pr(B_t|\mathbf{E}^T\mathbf{X}_{t+1}) \geq \min(1 - \sigma(B), \sigma(\epsilon - 10B), \sigma(B) - \sigma(B - \epsilon)) \equiv B_\alpha > 0$. Recall we also assume $\min_{\mathbf{H}_3 \in \mathcal{H}_3} g(\mathbf{H}_3) > B_X$ for some small $B_X > 0$. Therefore, $1 \geq \log p_\theta(\mathbf{H}_3) \geq \log B_\lambda^2 B_\alpha^2 \int_{\mathbb{R}^2} \left[\sigma\left(-B(|\mathbf{V}| + 1)\right)\right]^{24} d\mathbf{V} + \log B_X$.

Let $C = \int_{\mathbb{R}^2} \left[\sigma\left(-B(|\mathbf{V}| + 1)\right)\right]^{24} d\mathbf{V} < \infty$. By the triangle inequality, we can conclude that for all $\mathbf{H}_3, \Theta \in \Theta \times \mathcal{H}_3$, $|\log p_\theta(\mathbf{H}_3)| < |\log B_\lambda^2 B_\alpha^2 C| + |\log B_X| < \infty$. Thus by choosing $F(\mathbf{H}_3) = |\log B_\lambda^2 B_\alpha^2 C| + |\log B_X|$, assumption (C4) is satisfied.

We also wish to show $||\widehat{\mathbb{E}}_n[\mathbf{E}|\mathbf{H}_2] - \mathbb{E}[\mathbf{E}|\mathbf{H}_2]||_{P_{\theta_0}} \to 0$. Let $\mathbf{H}_{3,D} = (\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_1^R, \mathbf{W}_2^R, B_1, B_R)$ be the components of $\mathbf{H}_3$ dependent on $\mathbf{E}$ and $\mathbf{H}_{3,I} = (\mathbf{X}_1, \mathbf{X}_2, A_1, A_2, \mathbf{Y})$ be the components conditionally independent of $\mathbf{E}$. Note that $\mathbf{H}_3 = \mathbf{H}_{3,D} \cup \mathbf{H}_{3,I}$, $f_\theta(\mathbf{H}_3|\mathbf{E})$ is a function of only $\mathbf{E}$ and $\mathbf{H}_{3,D}$. Note that $\widehat{\mathbb{E}}_n[\mathbf{E}|\mathbf{H}_2] = \mathbb{E}_{\hat{\theta}_n}[\mathbf{E}|\mathbf{H}_2]$ where $\mathbb{E}_\theta[\mathbf{E}|\mathbf{H}_2] = \frac{\int_{\mathcal{E}} \mathbf{E} p_\theta(\mathbf{H}_2|\mathbf{E}) \Pr(\mathbf{E}) d\mathbf{E}}{\int_{\mathcal{E}} p_\theta(\mathbf{H}_2) \Pr(\mathbf{E}) d\mathbf{E}} = \frac{\int_{\mathcal{E}} \mathbf{E} f_\theta(\mathbf{H}_2|\mathbf{E}) \Pr(\mathbf{E}) d\mathbf{E}}{\int_{\mathcal{E}} f_\theta(\mathbf{H}_2|\mathbf{E}) \Pr(\mathbf{E}) d\mathbf{E}}$, $f_\theta(\mathbf{H}_2|\mathbf{E}) = \sum_{\mathbf{H}_{3,D} \in \mathcal{H}_{3,D}(\mathbf{H}_{2,D})} f_\theta(\mathbf{H}_3|\mathbf{E})$, $\mathcal{H}_{3,D}(\mathbf{h}_{2,D})$ is the set of $\mathbf{H}_{3,D}$ where $\mathbf{H}_{2,D} = \mathbf{h}_{2,D}$ and does not depend on θ, and $\mathbf{H}_{2,D} = (\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_1^R, \mathbf{W}_2^R, B_1)$. This is a finite sum, and each element $f_\theta(\mathbf{H}_3|\mathbf{E})$ of this sum is continuous in θ. Therefore, $f_\theta(\mathbf{H}_2|\mathbf{E})$ is continuous in θ. As $f_\theta(\mathbf{H}_2|\mathbf{E}) < 1$, we have that both the numerator and denominator is continuous in θ by the dominated convergence theorem. Then by the continuous mapping theorem, $\mathbb{E}_{\hat{\theta}_n}[\mathbf{E}|\mathbf{H}_2] \to_p \mathbb{E}_\theta[\mathbf{E}|\mathbf{H}_2]$, point-wise for almost every $\mathbf{E}, \mathbf{H}_2 \in \mathcal{E} \times \mathcal{H}_2$. Note that $||\widehat{\mathbb{E}}_n[\mathbf{E}|\mathbf{H}_2] - \mathbb{E}[\mathbf{E}|\mathbf{H}_2]||^2_{P_{\theta_0}} = \sum_{\mathbf{H}_2 \in \mathcal{H}_{2,D}} \left[\left(\mathbb{E}_{\hat{\theta}_n}[\mathbf{E}|\mathbf{H}_{2,D}] - \mathbb{E}[\mathbf{E}|\mathbf{H}_{2,D}] \right)^2 \right]$ is a finite sum of terms. Thus applying the continuous mapping theorem and Slutsky's theorem, we obtain our desired result.

To show that $\sqrt{n}(\hat{\theta} - \theta_0) \to_d \mathcal{N}(0, I(\theta_0)^{-1})$, we require for all $\theta, \theta_1, \theta_2 \in \mathcal{N}_\epsilon(\theta_0)$ and with probability one (N1) $I(\theta_0)$ is non-singular; (N2) $|p_{\theta_1}(\mathbf{H}_3|\mathbf{E}) - p_{\theta_2}(\mathbf{H}_3|\mathbf{E})| \le F(\mathbf{H}_3|\mathbf{E})|\theta_1 - \theta_2|$ for some $\mathbb{E}_{\theta_0,\mathbf{E}}[F^2(\mathbf{H}_3|\mathbf{E})] < \infty$; (N3) $p_\theta(\mathbf{H}_3) > c$ for some $c > 0$; (N4) $p_\theta(\mathbf{H}_3|\mathbf{E})$ is continuous differentiable; (N5) $||\nabla_\theta p_\theta(\mathbf{H}_3|\mathbf{E})||_\infty < G(\mathbf{H}_3|\mathbf{E})$ for some $\mathbb{E}_{\theta_0,\mathbf{E}}[G^2(\mathbf{H}_3|\mathbf{E})] < \infty$. (N1) is satisfied by assumption, and (N3) was demonstrated previously.

To prove the remaining conditions, we need to derive the gradient of the log-likelihood. Using elementary probability, calculus and linear algebra, we can derive the relevant marginal probabilities and their partial derivatives in closed-form:

$$\begin{aligned}
\Pr(W_{1,k}|\mathbf{V},\beta) =& \sigma(\beta_{0,k,1}+\beta_{1,k,1}^T\mathbf{V})^{W_{1k}}(1-\sigma(\beta_{0,k,1}+\beta_{1,k,1}^T\mathbf{V}))^{1-W_{1k}},\\
\nabla_{\beta_{1,k,1}}\Pr(W_{1,k}|\mathbf{V},\beta) =& (2W_{1k}-1)\sigma'(\beta_{0,k,1}+\beta_{1,k,1}^T\mathbf{V})V,\\
\Pr(\mathbf{W}_1^R|\mathbf{V},\lambda) =& \frac{\exp(-\lambda_1 T(\mathbf{W}_1^R,\mathbf{E}^R))}{\sum_{\mathbf{v}\in\mathcal{P}}\exp(-\lambda_1 T(\mathbf{v},\mathbf{E}^R))},\\
\nabla_{\lambda_1}\Pr(\mathbf{W}_1^R|\mathbf{V},\lambda) =& \frac{\left[\sum_{\mathbf{v}\in\mathcal{P}}\exp(-\lambda_1 T(\mathbf{v},\mathbf{E}^R))(T(\mathbf{v},\mathbf{E}^R)-T(\mathbf{W}_1^R,\mathbf{E}^R)\right]}{\left(\sum_{\mathbf{v}\in\mathcal{P}}\exp(-\lambda_1 T(\mathbf{v},\mathbf{E}^R))\right)^2}\exp(-\lambda_1 T(\mathbf{W}_1^R,\mathbf{E}^R)),\\
\Pr(B_1) =& \sigma(\alpha_{0,B_1,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{X}_2)^{1-I(B_1=7)}-I(B_1\neq 1)\sigma(\alpha_{0,B_1-1,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{X}_2),\\
\nabla_{\alpha_{1,1}}\Pr(B_1) =& -\left[I(B_1\neq 7)\sigma'(\alpha_{0,B_1,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{X}_2)+I(B_1\neq 1)\sigma'(\alpha_{0,B_1-1,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{X}_2)\right](\mathbf{E}^T\mathbf{X}_2).
\end{aligned}$$

Other partial derivatives can be derived similarly:

$$\begin{aligned}
\nabla_{\beta_{0,k,1}}\Pr(W_{1,k}|\mathbf{V},\beta) =& (2W_{1k}-1)\sigma'(\beta_{0,k,1}+\beta_{1,k,1}^T\mathbf{V}),\\
\nabla_{\beta_{1,k,2}}\Pr(W_{2,k}|\mathbf{V},\beta) =& (2W_{2k}-1)\sigma'(\beta_{0,k,2}+\beta_{1,k,2}^T\mathbf{V})V,\\
\nabla_{\beta_{0,k,2}}\Pr(W_{2,k}|\mathbf{V},\beta) =& (2W_{2k}-1)\sigma'(\beta_{0,k,2}+\beta_{1,k,2}^T\mathbf{V}),\\
\nabla_{\lambda_2}\Pr(\mathbf{W}_2^R|\mathbf{V},\lambda) =& \frac{\left[\sum_{\mathbf{v}\in\mathcal{P}}\exp(-\lambda_2 T(\mathbf{v},\mathbf{E}^R))(T(\mathbf{v},\mathbf{E}^R)-T(\mathbf{W}_2^R,\mathbf{E}^R)\right]}{\left(\sum_{\mathbf{v}\in\mathcal{P}}\exp(-\lambda_2 T(\mathbf{v},\mathbf{E}^R))\right)^2}\exp(-\lambda_1 T(\mathbf{W}_2^R,\mathbf{E}^R)),\\
\nabla_{\alpha_{0,k,1}}\Pr(B_1) =& I(B_1=k)\sigma'(\alpha_{0,k,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{X}_2)-I(B_1=k-1)\sigma'(\alpha_{0,k,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{X}_2),\\
=& \left[1_k(B_1)-1_{k-1}(B_1)\right]\sigma'(\alpha_{0,k,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{X}_2),\\
\nabla_{\alpha_{1,2}}\Pr(B_2) =& -\left[I(B_2\neq 7)\sigma'(\alpha_{0,B_2,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{Y})+I(B_2\neq 1)\sigma'(\alpha_{0,B_2-1,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{Y})\right](\mathbf{E}^T\mathbf{Y}),\\
\nabla_{\alpha_{0,k,2}}\Pr(B_2) =& \left[1_k(B_2)-1_{k-1}(B_2)\right]\sigma'(\alpha_{0,k,1}-\alpha_{1,1}\mathbf{E}^T\mathbf{Y}).
\end{aligned}$$

Let $\beta_{\cdot,k,t}=(\beta_{0,k,t},\beta_{1,k,t}^T)^T$, $\mathbf{V}^*=(1,\mathbf{V}^T)^T$, $\mathbf{W}_t^{(-k)}=(\mathbf{W}_{t,1},...,\mathbf{W}_{t,k-1},\mathbf{W}_{t,k+1},...,\mathbf{W}_{t,12})^T$. Note that $3-t=2$ if $t=1$ and $3-t=1$ if $t=2$ (i.e. $3-t$ is the temporal inverse). We now have:

$$
\begin{aligned}
f_\theta(\mathbf{H}_3|\mathbf{V}) &= \Pr_\theta(\mathbf{W}_1|\mathbf{V}) \Pr_\theta(\mathbf{W}_2|\mathbf{V}) \Pr_\theta(\mathbf{W}_1^R|\mathbf{V}) \Pr(\mathbf{W}_2^R|\mathbf{V}) \Pr_\theta(B_1|\mathbf{E}^T\mathbf{X}_2) \Pr_\theta(B_2|\mathbf{E}^T\mathbf{Y}), \\
\nabla_{\beta_{\cdot,k,t}} f_\theta(\mathbf{H}_3|\mathbf{V}) &= \Pr_\theta(\mathbf{W}_1^R|\mathbf{V}) \Pr_\theta(B_1|\mathbf{E}^T\mathbf{X}_2) \Pr_\theta(\mathbf{W}_2^R|\mathbf{V}) \Pr_\theta(B_2|\mathbf{E}^T\mathbf{Y}) \\
&\quad \times \Pr_\theta(\mathbf{W}_{3-t}|\mathbf{V}) \Pr_\theta(\mathbf{W}_t^{(-k)}|\mathbf{V})(2W_{t,k}-1)\sigma'(\beta_{0,k,\cdot}^T\mathbf{V}^*)\mathbf{V}^*, \\
\nabla_{\lambda_t} f_\theta(\mathbf{H}_3|\mathbf{V}) &= \Pr_\theta(\mathbf{W}_1|\mathbf{V}) \Pr_\theta(B_1|\mathbf{E}^T\mathbf{X}_2) \Pr_\theta(\mathbf{W}_2|\mathbf{V}) \Pr_\theta(B_2|\mathbf{E}^T\mathbf{Y}) \Pr_\theta(\mathbf{W}_{3-t}^R|\mathbf{V}) \\
&\quad \times \exp(-\lambda_t T(\mathbf{W}_t^R, \mathbf{E}^R)) \frac{\left[\sum_{\mathbf{v}\in\mathcal{P}} \exp(-\lambda_t T(\mathbf{v}, \mathbf{E}^R))(T(\mathbf{v}, \mathbf{E}^R) - T(\mathbf{W}_t^R, \mathbf{E}^R)\right]}{\left(\sum_{\mathbf{v}\in\mathcal{P}} \exp(-\lambda_t T(\mathbf{v}, \mathbf{E}^R))\right)^2}, \\
\nabla_{\alpha_{0,k,t}} f_\theta(\mathbf{H}_3|\mathbf{V}) &= \Pr_\theta(\mathbf{W}_1|\mathbf{V}) \Pr_\theta(\mathbf{W}_1^R|\mathbf{V}) \Pr_\theta(\mathbf{W}_2|\mathbf{V}) \Pr_\theta(\mathbf{W}_2^R|\mathbf{V}) \Pr_\theta(B_{3-t}|\mathbf{E}^T\mathbf{X}_2)\,(1_k(B_t) - 1_{k-1}(B_t)) \\
&\quad \times \sigma'(\alpha_{0,k,t} - \alpha_{1,t}\mathbf{E}^T\mathbf{X}_{t+1}), \\
\nabla_{\alpha_{1,t}} f_\theta(\mathbf{H}_3|\mathbf{V}) &= \Pr_\theta(\mathbf{W}_1|\mathbf{V}) \Pr_\theta(\mathbf{W}_1^R|\mathbf{V}) \Pr_\theta(\mathbf{W}_2|\mathbf{V}) \Pr_\theta(\mathbf{W}_2^R|\mathbf{V}) \Pr_\theta(B_{3-t}|\mathbf{E}^T\mathbf{X}_2) \\
&\quad \times \left[I(B_t \neq 7)\sigma'(\alpha_{0,B_t,1} - \alpha_{1,1}\mathbf{E}^T\mathbf{Y}) - I(B_t \neq 1)\sigma'(\alpha_{0,B_t-1,1} - \alpha_{1,1}\mathbf{E}^T\mathbf{Y})\right](\mathbf{E}^T\mathbf{Y}).
\end{aligned}
$$

We can see that $\nabla_\theta f_\theta(\mathbf{H}_3|\mathbf{V})$ is continuous in θ, and thus so is $\nabla_\theta p_\theta(\mathbf{H}_3|\mathbf{V}) = g(\mathbf{H}_3)\nabla_\theta f_\theta(\mathbf{H}_3|\mathbf{V})$. Thus (N4) is satisfied.

We can also derive an upper bound for $||\nabla_\theta f_\theta(\mathbf{H}_3|\mathbf{V})||_\infty$. As $|\Pr(\mathbf{W}_t^R|\theta)|$, $\Pr(B_t|\theta)|$, and $|\Pr(\mathbf{W}_t|\theta) \leq 1|$, $\forall\theta \in \Theta$ and $|\sigma'(x)| \leq 1$, $\forall x \in \mathbb{R}$, we have that $|\nabla_{\beta_{\cdot,k,t}} f_\theta(\mathbf{H}_3|\mathbf{V})| \leq |\mathbf{V}^*|$. As $|\exp(-\lambda_t T(\mathbf{W}_t^R, \mathbf{E}^R)| < 1$ and $|T(\mathbf{v}, \mathbf{E}^R)|$, and $|T(\mathbf{W}_t^R, \mathbf{E}^R)| \leq 3$ for $v \in \mathcal{P}$, $\exp(-\lambda_t T(\mathbf{v}, \mathbf{E}^R) = 1$ for some $v \in \mathcal{P}$ and $|\mathcal{P}| = 6$, we have that $|\nabla_{\lambda_t} f_\theta(\mathbf{H}_3|\mathbf{V})| \leq 18$. As $|\mathbf{E}^T X| < 10$, $|\nabla_{\alpha_{0,k,t}} f_\theta(\mathbf{H}_3|\mathbf{V})| \leq 1$ and $|\nabla_{\alpha_{1,t}} f_\theta(\mathbf{H}_3|\mathbf{V})| \leq 20$, we then have that $||\nabla_\theta f_\theta(\mathbf{H}_3|\mathbf{V})||_\infty \leq \max(|\mathbf{V}^*|, 20)$. Thus (N5) is satisfied, and by the Leibniz integral rule, $\nabla_\theta \log p_\theta(\mathbf{H}_3) = \nabla_\theta \left[\log \int_{\mathbb{R}^2} f_\theta(\mathbf{H}_3|\mathbf{V}) \Pr(\mathbf{V}) d\mathbf{V} + \log g(\mathbf{H}_3)\right] = (f_\theta(\mathbf{H}_3))^{-1} \int_{\mathbb{R}^2} \nabla_\theta f_\theta(\mathbf{H}_3|\mathbf{V}) d\mathbf{V}$.

It remains to show (N2). By the mean value theorem, an everywhere-differentiable function $f : \mathcal{X} \to \mathbb{R}$ with bounded first derivatives will be Lipschitz continuous over $\mathcal{X}$ with Lipschitz constant L upper-bounded as $\sup_{x\in\mathcal{X}} |f'(x)|$ (Shalev-Shwartz and Ben-David, 2014). Then:

$$
\begin{aligned}
\Pr(W_{1,k}|\mathbf{V},\theta) =& W_{tk}\sigma(\beta_{0,k,t}+\beta_{1,k,t}^T\mathbf{V})+(1-W_{tk})(1-\sigma(\beta_{0,k,t}+\beta_{1,k,t}^T\mathbf{V}))\\
|\Pr(W_{t,k}|\mathbf{V},\theta^{(1)})-\Pr(W_{t,k}|\mathbf{V},\theta^{(2)})| \leq& W_{tk}\left|\sigma(\beta_{\cdot,k,t}^{(1)T}\mathbf{V}^*)-\sigma(\beta_{\cdot,k,t}^{(2)T}\mathbf{V}^*)\right|\\
&+(1-W_{tk})\left|\sigma(\beta_{\cdot,k,t}^{(2)T}\mathbf{V}^*)-\sigma(\beta_{\cdot,k,t}^{(1)T}\mathbf{V}^*)\right|\\
\leq& W_{tk}|\beta_{\cdot,k,t}^{(1)T}\mathbf{V}^*-\beta_{\cdot,k,t}^{(2)T}\mathbf{V}^*|+(1-W_{tk})|\beta_{\cdot,k,t}^{(2)T}\mathbf{V}^*-\beta_{\cdot,k,t}^{(1)T}\mathbf{V}^*|\\
=&|(\beta_{\cdot,k,t}^{(2)}-\beta_{\cdot,k,t}^{(1)})^T\mathbf{V}^*|\\
\leq&||\mathbf{V}^*||_2||\beta_{\cdot,k,t}^{(2)}-\beta_{\cdot,k,t}^{(1)}||_2\leq||\mathbf{V}^*||_2||\theta^{(2)}-\theta^{(1)}||_2.
\end{aligned}
$$

The first inequality comes from the triangle inequality. The second equality comes from the fact that the sigmoid function is everywhere-differentiable and $|\sigma'(x)|\leq 1, \forall x\in\mathbb{R}$, making it Lipschitz with constant $L=1$. The third inequality comes from the Cauchy-Schwartz inequality.

Moreover:

$$
\begin{aligned}
\Pr(B_t|\mathbf{E}^T\mathbf{X}_{t+1},\theta) =& I(B_t=7)\\
&+I(B_t\neq 7)\sigma(\alpha_{0,B_t,1}-\alpha_{1,t}\mathbf{E}^T\mathbf{X}_{t+1})-I(B_t\neq 1)\sigma(\alpha_{0,B_t-1,1}-\alpha_{1,t}\mathbf{E}^T\mathbf{X}_{t+1}),\\
|\Pr(B_t|\mathbf{E}^T\mathbf{X}_{t+1},\theta^{(1)})&-\Pr(B_t|\mathbf{E}^T\mathbf{X}_{t+1},\theta^{(2)})|\\
\leq& I(B_t\neq 7)\left|\sigma(\alpha_{0,B_t,t}^{(1)}-\alpha_{1,t}^{(1)}\mathbf{E}^T\mathbf{X}_{t+1})-\sigma(\alpha_{0,B_t,t}^{(2)}-\alpha_{1,t}^{(2)}\mathbf{E}^T\mathbf{X}_{t+1})\right|\\
&+I(B_t\neq 1)\left|\sigma(\alpha_{0,B_t-1,t}^{(1)}-\alpha_{1,t}^{(1)}\mathbf{E}^T\mathbf{X}_{t+1})-\sigma(\alpha_{0,B_t-1,t}^{(2)}-\alpha_{1,t}^{(2)}\mathbf{E}^T\mathbf{X}_{t+1})\right|\\
\leq&|(\alpha_{0,B_t,t}^{(1)}-\alpha_{0,B_t,t}^{(2)})+(\alpha_{1,t}^{(2)}-\alpha_{1,t}^{(1)})\mathbf{E}^T\mathbf{X}_{t+1}|\\
&+|(\alpha_{0,B_t-1,t}^{(1)}-\alpha_{0,B_t-1,t}^{(2)})+(\alpha_{1,t}^{(2)}-\alpha_{1,t}^{(1)})\mathbf{E}^T\mathbf{X}_{t+1}|\\
\leq&|\alpha_{0,B_t,t}^{(1)}-\alpha_{0,B_t,t}^{(2)}|+|\alpha_{1,t}^{(1)}-\alpha_{1,t}^{(2)}|\mathbf{E}^T\mathbf{X}_{t+1}+|\alpha_{0,B_t-1,t}^{(1)}-\alpha_{0,B_t-1,t}^{(2)}|+|\alpha_{1,t}^{(1)}-\alpha_{1,t}^{(2)}|\mathbf{E}^T\mathbf{X}_{t+1}\\
\leq&2||\alpha_{0,\cdot,t}^{(1)}-\alpha_{0,\cdot,t}^{(2)}||_2+20|\alpha_{1,t}^{(1)}-\alpha_{1,t}^{(2)}|\\
\leq&22||\theta^{(2)}-\theta^{(1)}||_2.
\end{aligned}
$$

The first inequality uses the triangle inequality for absolute values. The second inequality uses the fact that the sigmoid has Lipschitz constant $L = 1$. The third inequality uses the triangle inequality again. The fourth inequality uses the fact that $|\mathbf{E}^T\mathbf{X}_{t+1}| \leq 10, t \in \{1, 2\}$.

Note that $f : [0, \infty) \to [0, 1]$ defined by $f(x) = \exp(-x)$ is Lipschitz with constant $L = \sup_{x \in [0,\infty)} \{f'(x)\} \leq 1$. Moreover, note that the tuple $\mathcal{T} = (T(\mathbf{v}, \mathbf{E}^R) : v \in \mathcal{P}) = (0, 1, 1, 2, 2, 3)$ is equivalent for all $E^R \in \mathcal{P}$. Finally, note that $|\exp(-\lambda_t T(\mathbf{W}_1^R, \mathbf{E}^R))|, |[\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t T(\mathbf{v}, \mathbf{E}^R))]^{-1}| \leq 1$, $T(x, y) \in \{0, 1, 2, 3\}$, and $\text{length}(\mathcal{T}) = 6$. Putting this all together we have:

$$\nabla_{\lambda_t} \left(\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t T(\mathbf{v}, \mathbf{E}^R)) \right)^{-1} = \frac{\sum_{T \in \mathcal{T}} \exp(-\lambda_t T) T}{\left(\sum_{T \in \mathcal{T}} \exp(-\lambda_t T) \right)^2} \leq 18$$

Thus $f : [0, \infty) \to \mathbb{R}$ defined as $f(x) = 1/\sum_{T \in \mathcal{T}} \exp(-xT)$ is Lipschitz with constant $L \leq 18$. Then:

$$\begin{aligned}
\Pr(\mathbf{W}_t^R|\mathbf{V}, \lambda_t) =& \frac{\exp(-\lambda_t T(\mathbf{W}_t^R, \mathbf{E}^R))}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t T(\mathbf{v}, \mathbf{E}^R))} \text{ and} \\
|\Pr(\mathbf{W}_t^R|\mathbf{V}, \lambda_t^{(2)}) - \Pr(\mathbf{W}_t^R|\mathbf{V}, \lambda_t^{(1)})| =& \left| \frac{\exp(-\lambda_t^{(2)} T(\mathbf{W}_t^R, \mathbf{E}^R))}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t^{(2)} T(\mathbf{v}, \mathbf{E}^R))} - \frac{\exp(-\lambda_t^{(1)} T(\mathbf{W}_t^R, \mathbf{E}^R))}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t^{(1)} T(\mathbf{v}, \mathbf{E}^R))} \right| \\
\leq& \left| \frac{\exp(-\lambda_t^{(2)} T(\mathbf{W}_t^R, \mathbf{E}^R))}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t^{(2)} T(\mathbf{v}, \mathbf{E}^R))} - \frac{\exp(-\lambda_t^{(1)} T(\mathbf{W}_t^R, \mathbf{E}^R))}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t^{(2)} T(\mathbf{v}, \mathbf{E}^R))} \right| \\
&+ \left| \frac{\exp(-\lambda_t^{()} T(\mathbf{W}_t^R, \mathbf{E}^R))}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t^{(2)} T(\mathbf{v}, \mathbf{E}^R))} - \frac{\exp(-\lambda_t^{(1)} T(\mathbf{W}_t^R, \mathbf{E}^R))}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t^{(1)} T(\mathbf{v}, \mathbf{E}^R))} \right| \\
\leq& \left| \exp(-\lambda_t^{(2)} T(\mathbf{W}_t^R, \mathbf{E}^R)) - \exp(-\lambda_t^{(1)} T(\mathbf{W}_t^R, \mathbf{E}^R)) \right| \\
&+ \left| \frac{1}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t^{(2)} T(\mathbf{v}, \mathbf{E}^R))} - \frac{1}{\sum_{\mathbf{v} \in \mathcal{P}} \exp(-\lambda_t^{(1)} T(\mathbf{v}, \mathbf{E}^R))} \right|
\end{aligned}$$

$$
\begin{aligned}
&\leq |\lambda_t^{(2)} T(\mathbf{W}_t^R, \mathbf{E}^R) - \lambda_t^{(1)} T(\mathbf{W}_t^R, \mathbf{E}^R)| + 18|\lambda_t^{(2)} - \lambda_t^{(1)}| \\
&\leq 21|\lambda_t^{(2)} - \lambda_t^{(1)}| \\
&\leq 21||\theta^{(2)} - \theta^{(1)}||.
\end{aligned}
$$

As the product of Lipschitz continuous functions is also Lipschitz continuous with the Lipschitz constant being the sum of those of the functions being multiplied (Shalev-Shwartz and Ben-David, 2014), $f_\theta(\mathbf{H}_3|\mathbf{V})$ is Lipschitz continuous in $\theta \in \Theta$ with constant $L \leq 24||\mathbf{V}^*||_2 + 43$, and thus so is $p_\theta(\mathbf{H}_3|\mathbf{V})$. Thus condition (N2) is satisfied, concluding the proof. $\square$

D Supplementary Tables and Figures

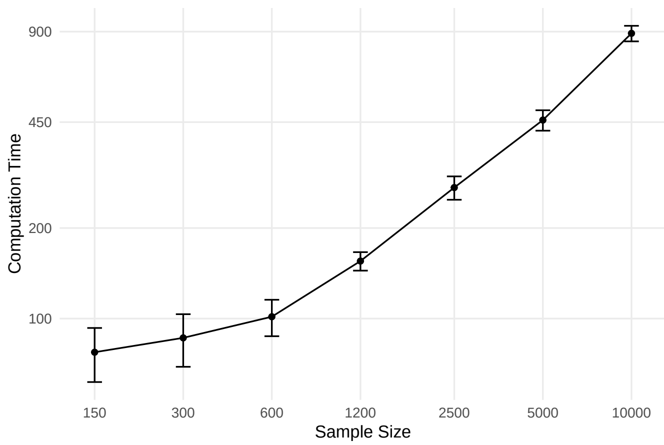

Figure D.1: Model-fitting time per sample size. Shown is the mean across seeds with standard error bars for each sample size.

References

Precup, D., Sutton, R. S., and Singh, S. (2000). Eligibility traces for off-policy policy evaluation. In *Proceedings of the 17th International Confrence on Machine Learning (ICML 2000)*, page 80.

Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Zhao, Y.-Q., Zeng, D., Laber, E. B., and Kosorok, M. R. (2015). New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598.