Online Companion:
Optimization in Online Content Recommendation Services: Beyond Click-Through Rates

Omar Besbes
Columbia University

Yonatan Gur
Stanford University

Assaf Zeevi
Columbia University

first version: February 28, 2014, last revised: April 8, 2015

A Auxiliary analysis

A.1 Simulation

We conduct the following simulation based on our model estimates (that include approximately 500 articles in each estimation batch). We assume $\ell = 5$, that is, each assortment contains exactly five links, and that prices are uniform, that is, $w(\cdot) = 1$, for each article. We simulate the performance of different recommendation approaches using the following procedure:

Simulation procedure. Inputs: $k \in \{5, 10, 25, 50, 75, 100\}$, and a reader type $u_0 \in \{u_{exp}, u_{inexp}\}$

1. Set batch index $j = 1$
2. Repeat from $\tau = 1$ to $\tau = 1,000$:
   - Construct $\mathcal{X}_0$ by randomly drawing k articles (uniformly) out of those appearing in batch j.
   - From the set of available articles, draw randomly one article (uniformly) to be x_0.
   - Set $T = k - 1$. For all $t = 1, \ldots, T$ follow the update rules: $\mathcal{X}_t = \mathcal{X}_{t-1} \setminus \{x_{t-1}\}$; $u_t = u_{inexp}$ if $u_0 = u_{inexp}$ and $t \leq 1$, otherwise $u_t = u_{exp}$. Based on the estimates of the drawn articles and those of the control parameters in batch j, calculate recommendation schedules that solve: the content recommendation problem (5), the myopic content recommendation problem (6), and the one-step lookahead content recommendation problem (7), obtaining $V^*_{j,\tau} = V^*_1(u_1, \mathcal{X}_1, x_0)$, $V^m_{j,\tau} = V^m_1(u_1, \mathcal{X}_1, x_0)$, and $V^{one}_{j,\tau} = V^{one}_1(u_1, \mathcal{X}_1, x_0)$.
3. Update batch index $j \to j + 1$, and while $j \leq 360$ go back to step 2.
4. Calculate average clicks-per-visit performances:

$$\bar{V}^* = \sum_{j=1}^{360} \sum_{\tau=1}^{1,000} V^*_{j,\tau}; \qquad \bar{V}^m = \sum_{j=1}^{360} \sum_{\tau=1}^{1,000} V^m_{j,\tau}; \qquad \bar{V}^{one} = \sum_{j=1}^{360} \sum_{\tau=1}^{1,000} V^{one}_{j,\tau}.$$

We determine the schedules selected for the various solutions (including the optimal one) by exhaustively searching over potential schedules of recommendations to maximize the objective of each policy given the estimates of candidate articles, and then calculating the expected clicks per visit using these estimates. To reduce the computation time we used the monotonicity of the value function in (5) with respect to the engageability and clickability values of recommended articles to dismiss a set of suboptimal schedules. We repeated the simulation procedure for combinations of $k \in \{5, 10, 25, 50, 75, 100\}$ and $u_0 \in \{u_{exp}, u_{inexp}\}$. The average performances for different sizes of articles' sets are depicted in Figure A-1. Note that Figure A-1 includes also the simulated performance of *adjusted-myopic* recommendations

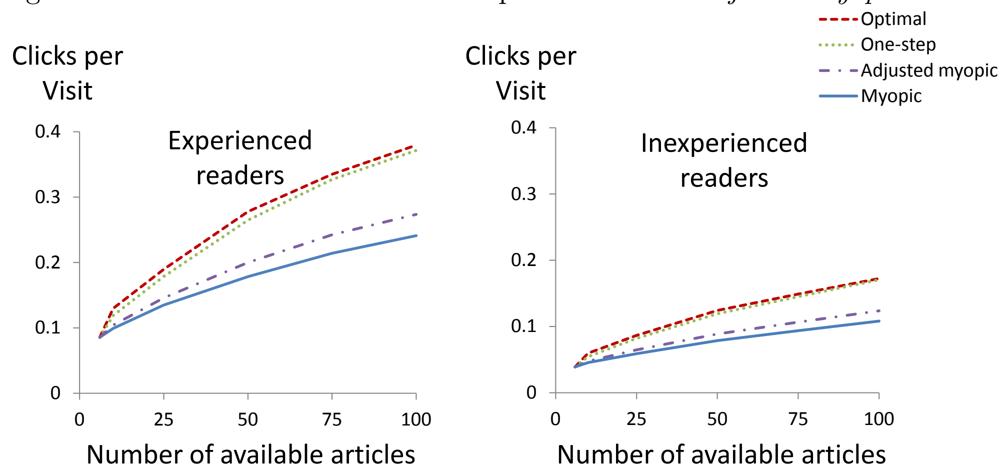

Figure A-1: **The near-optimality of one-step look-ahead recommendations.** (*Left*) The average performance, in clicks per visit, of optimal, one-step lookahead, and adjusted-myopic (defined in §4), and myopic recommendations, for readers that recently clicked on internal recommendations. (*Right*) The average performances for readers that did not click recently on an internal recommendation.

discussed in §4. We note that its simulated performance is only slightly below that of an index policy that maximizes $\gamma(x) \cdot [1 + \beta(x)]$; to avoid overload the latter does not appear in Figure A-1 (its elements are not observed online). Figure A-1 shows that while optimal recommendations that account for the whole future path of readers may generate an increase of approximately 50 percent in clicks per visit compared to myopic recommendations, the major part of this performance gap may be captured by one-step lookahead recommendations. While readers that are familiar with internal recommendations and have clicked on those before tend to generate approximately twice as many clicks per visit compared to readers that did not click on internal recommendations recently, the significant impact of one-step look-ahead recommendations is similar for both "experienced" as well as "inexperienced" readers.

A.2 Robustness of the model

We consider two model variants that use additional information (and thus a larger number of parameters). Repeating the predictive analysis from §2.4, we demonstrate that enriching our model by considering more information does not necessarily increase the established predictive power. This may result from various reasons, one of which is overfitting the data due to the larger number of parameters.

We first consider a *path-rich* model, correspond to the model in (1) with $\phi_{u,x,y}(A)$ defined by:

$$\phi^{pr}_{u,x,y}(A) \; = \; \exp\left\{\alpha + \beta_x + \gamma_y + \mu_{x,y} + \theta'_u + \lambda_{p(y,A)}\right\}, \tag{A-1}$$

where β_x, γ_y, $\mu_{x,y}$, and $\lambda_{p(y,A)}$ are defined as in §2.2, and θ'_u captures the length of the current session: the number of articles viewed during the current visit including the current host article. In particular, $\theta'_u \in \{1, 2, 3, 4\}$ where a value of 1 implies a session length of 1 (the current host article is the first article that is viewed in the session), a value of 2 implies a session length of 2 (the current host article is the second article that is viewed in the session), a value of 3 implies a session length of 3, and a value of 4 implies a session length of 4 or more. The path-rich model aims to capture the reader type at a finer level (relative to the nominal model in §2.2) based on information on the length of the current session.

The second model we consider is the following *context-rich* model, in which $\phi_{u,x,y}(A)$ is defined by:

$$\phi^{cr}_{u,x,y}(A) \; = \; \exp\left\{\alpha + \beta_x + \gamma_y + \mu'_{x,y} + \theta_u + \lambda_{p(y,A)}\right\}, \tag{A-2}$$

where β_x, γ_y, θ_u, and $\lambda_{p(y,A)}$ are defined as in §2.2, and $\mu'_{x,y}$ are entries of a 9 by 9 (non-symmetric) topic transition matrix that use classification of articles into 9 major topics. For example, $\mu'_{1,4}$ is the transition entry from the topic "news" into the topic "sport," and is used whenever the host article x is classified as "news" and the recommended article y is classified as "sport". While the number of parameters in this model is clearly higher, this model aims to better identify the cross effect $\mu_{x,y}$' relative to the indicator $\mu_{x,y}$ that is used in the nominal model defined in §2.2 to identify only the cases in which the host and the recommended article belong to the same sub-topic (recall that in the nominal model defined in §2.2, we use a classification into 84 sub-topics; for example, both the sub-topics "news: politics" and "news: crime" fall under the topic "news" that is used here).

We repeat the predictive analysis described in §2.4 to construct the (out-of-sample) ROC curves of the path-rich and the context-rich model. Figure A-2 depicts the ROC curves of the considered model relative to the one of the nominal model. Figure A-2 demonstrates that both of the considered models do not achieve a better predictive power than the one achieved by the nominal model in §2.4. The area under the curve (AUC) of the path-rich model is the same as the one of the nominal model; the two ROC curves are not statistically distinguishable. The AUC of the context-rich model is lower than the one of the nominal model; while accounting for more information, this model achieves lower predictive power than the one achieved by the nominal model in §2.4, which is probably due to overfitting.

Overfitting and in-sample predictive analysis. Since in each two-hour batch we estimate (in the nominal model) approximately 1,000 parameters, a natural issue to be concerned about is that of overfitting. To investigate the extent of overfitting, we tested the predictive power of the nominal model in-sample as well (that is, tested each set of estimators along the batch over which these estimators were produced). We depict in Figure A-2 the in-sample ROC curve of the nominal model and observe that it is similar to the out-of-sample one, which is reassuring. While the in-sample AUC of this model is slightly higher at 0.76, it is important to recognize that: first, of course, the in-sample predictive power of any model is expected to be higher than an out-of-sample one; second, clickability and engageability

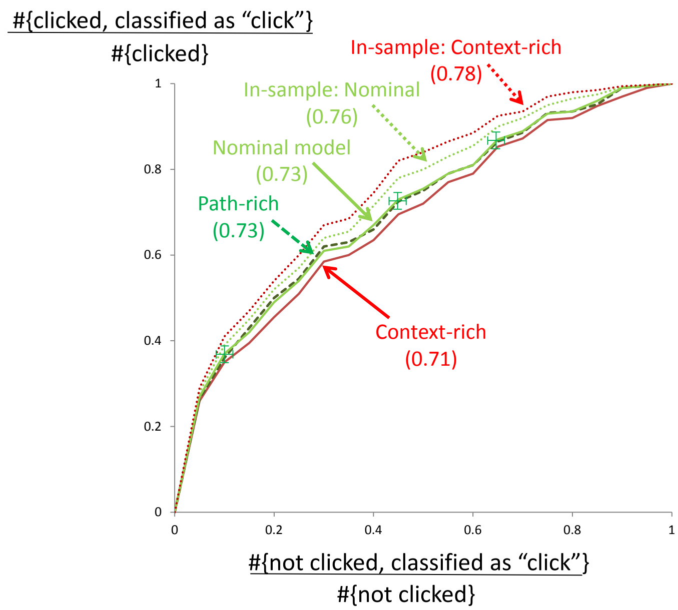

Figure A-2: **Alternative models: predictive analysis** . The plot depicts the out-of-sample ROC curve generated by the basic model defined in §2.2, together with the path-rich model and the context-rich model. In addition, the in-sample ROC curves are added for our model as well as the context-rich model. The area under each curve (AUC) appears in parentheses. All standard errors (with respect to both axes) are smaller than 0.02; three illustrative examples are depicted on the "nominal model" curve.

are time varying, and therefore predictive power may be lost as time goes by.

In Figure A-2, we also depict the in-sample ROC curve associated with the context-rich model. We observe that there is a larger gap between the in-sample AUC (0.78) and and the out-of-sample AUC (0.71). This, in conjunction with the fact that the context-rich model has lower out-of-sample predictive power than the nominal model, may indicate that the context-rich model suffers from overfitting in this context. Indeed, it includes roughly 10% more parameters compared to the nominal model.

A.3 Contrasting topics in the content space

We illustrate how topics may be visualized and understood through the lens of the content space. Consider a topic τ with associated articles $\{x_i^\tau\}_{i=1}^N$ that are assumed to be positioned in the content space according to a bivariate normal distribution with a mean $\mu = (\mu_\beta, \mu_\gamma)$ and a covariance matrix $\Sigma = diag\,(\sigma_\beta, \sigma_\gamma)$. One may represent τ though confidence sets, by estimating:

$$\hat{\mu}_\beta^\tau \;=\; \frac{1}{N}\sum_{i=1}^{N}\beta_i, \quad \hat{\sigma}_\beta^\tau \;=\; \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\beta_i - \hat{\mu}_\beta^\tau\right)^2}; \qquad \hat{\mu}_\gamma^\tau \;=\; \frac{1}{N}\sum_{i=1}^{N}\gamma_i, \quad \hat{\sigma}_\gamma^\tau \;=\; \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\gamma_i - \hat{\mu}_\gamma^\tau\right)^2}.$$

Figure A-3 depicts confidence sets of two topics: *architecture* and *celebrities*. While the celebrities category might lead to more instantaneous clicks, it does not dominate architecture in terms of future clicks.

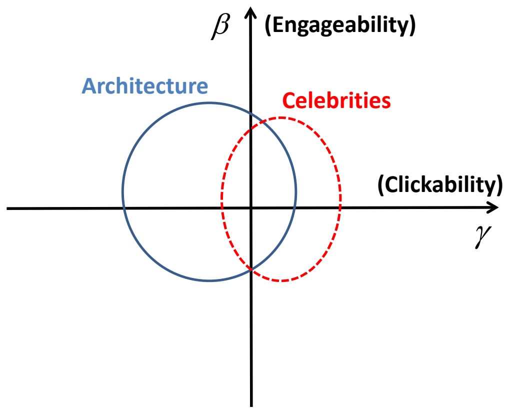

Figure A-3: **Contrasting topics in the content space.** The 90% confidence sets of the topics architecture and celebrities are illustrated. The corresponding estimates (obtained using our whole data set) are $\hat{\mu}_{\beta}^{arc} = 0.29$, $\hat{\mu}_{\gamma}^{arc} = -0.47$, $\hat{\sigma}_{\beta}^{arc} = 1.13$, $\hat{\sigma}_{\gamma}^{arc} = 1.06$; and $\hat{\mu}_{\beta}^{cel} = 0.13$, $\hat{\mu}_{\gamma}^{cel} = 0.72$, $\hat{\sigma}_{\beta}^{cel} = 0.88$, $\hat{\sigma}_{\gamma}^{cel} = 1.01$.

B Proofs

Proof of Proposition 1. We establish that the CRP is NP-hard by showing that the Hamiltonian path problem (HPP), a known NP problem (Gary and Johnson 1979) can be reduced to a special case of the CRP. We denote by $\mathcal{G}(\mathcal{V},\mathcal{E})$ a directed graph, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of arcs. An arc connecting one node v with another node v' is denoted by $e_{v,v'}$. When $v \in \mathcal{V}$ is connected to $v' \in \mathcal{V}$, one has $e_{v,v'} \in \mathcal{E}$. Given a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, the HPP is to determine whether there exists a connected path of arcs in $\mathcal{E}$, that visits all the vertices in $\mathcal{V}$ exactly once.

We next show that the HPP can be reduced to a special case of the CRP. Fix a general, directed graph $\mathcal{G}(\mathcal{V},\mathcal{E})$, and consider the following special case of the CRP, with a fixed x_0, in which:

- $T = |\mathcal{X}_0| - \ell$.
- $\mathcal{X}_1 = \mathcal{V}$, with $w(x) = 1$ for each article in $x \in \mathcal{X}_1$.
- $u_t = u_0 = u$ for all $t = 1,\dots,T$ (the reader type is fixed, and in particular, independent of the length of her path and on the articles she visits along her path).
- $\ell = 1$, i.e., every recommendation consists of a single link. Whenever fan article x hosts a recommendation A that includes a link to article y, we denote for simplicity $\mathbb{P}_{u,x,y}(A) = \mathbb{P}_{u,x}(y)$.
- $\mathbb{P}_{u,x}(y) \in \{0,1\}$ for all $y \in \mathcal{X}_t$, for all $t = 1,\dots,T$ (the click probabilities for any potential recommended link are binary at any epoch). In particular, for any $x, y \in \mathcal{X}_1$ we set:

$$\mathbb{P}_{u,x}(y) \;=\; \begin{cases} 1 & \text{if } e_{x,y} \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases}$$

- $\mathbb{P}_{u,x_0}(y) = 1$ for all $y \in \mathcal{X}_1$ (the first link is clicked, regardless of the selected recommendation).

Then, given the landing article $x_0 \in \mathcal{X}_0$, the CRP takes the following form:

$$V_t^*(u,\mathcal{X}_t,x_{t-1}) \;=\; \max_{x_t \in \mathcal{X}_t} \left\{ \mathbb{P}_{u,x_{t-1}}(x_t)\left(1 + V_{t+1}^*(u,\mathcal{X}_{t+1},x_t)\right)\right\},$$

for $t = 1,\dots,T-1$, and

$$V_T^*(u, \mathcal{X}_T, x_{T-1}) \;=\; \max_{x_T \in \mathcal{X}_T} \left\{ \mathbb{P}_{u,x_{T-1}}(x_T) \right\}.$$

To complete the reduction argument, we observe that there exists a connected path of arcs in $\mathcal{E}$ that visits any vertex in $\mathcal{V}$ exactly once if and only if $V_t^*(u, \mathcal{X}_1, x_0) = T$, and therefore, by obtaining a solution to the CRP one solves the HPP. Since the HPP is NP-hard, the CRP must be NP-hard as well. This concludes the proof. ■

Proof of Proposition 2. To analyze the performance gap between an optimal recommendations and a sequence of myopic recommendations, we focus on a special case of the CRP in which: $T = |\mathcal{X}_0| - \ell$; $u_t = u_0 = u$ for all $t = 1, \dots, T$ (the reader type is fixed, and independent of the length of her path and on the articles she visits along her path); $\ell = 1$, i.e., every recommendation consists of a single link; and $w(x) = 1$ for any available article x. Whenever an article x hosts a recommendation A that includes a link to article y, we denote for simplicity $\mathbb{P}_{u,x,y}(A) = \mathbb{P}_{u,x}(y)$. Then, the CRP in (5) can be written as:

$$V_t^*(u, \mathcal{X}_t, x_{t-1}) \;=\; \max_{x_t \in \mathcal{X}_t)} \left\{ \mathbb{P}_{u,x_{t-1}}(x_t) \left(1 + V_{t+1}^*(u, \mathcal{X}_{t+1}, x_t)\right) \right\},$$

for $t = 1, \dots, T-1$, and

$$V_T^*(u, \mathcal{X}_T, x_{T-1}) \;=\; \max_{x_T \in \mathcal{X}_T} \left\{ \mathbb{P}_{u,x_{T-1}}(x_T) \right\}.$$

Fix $u \in \mathcal{U}$ and $\varepsilon \in (0, 1/2)$. Consider the next construction of a set of available articles and a set of transition probabilities $(\mathcal{X}_0, \mathcal{P}_0^u) \in \mathcal{G}_{T+1}$, with a selection $x_0 \in \mathcal{X}_0$, in which there exists a sequence of articles $x_0, \dots, x_T$, such that:

$$\mathbb{P}_{u,x}(y) \;=\; \begin{cases} 1/2 - \varepsilon & \text{if } x = x_0 \text{ and } y = x_1 \\ 1 & \text{if } x = x_{t-1} \text{ and } y = x_t \text{ for some } t \in \{2, \dots, T\} \\ 1/2 + \varepsilon & \text{if } x = x_0 \text{ and } y = x_T \\ 0 & \text{otherwise.} \end{cases}$$

Then, the optimal schedule of recommendation is to recommend article x_t at epoch t, generating $(1/2 - \varepsilon)T$ expected clicks. Moreover, any myopic schedule of recommendation will begin with recommending x_T at epoch $t = 1$, generating $(1/2 + \varepsilon)$ expected clicks. Therefore, one has:

$$\inf_{(\mathcal{X}_0, \mathcal{P}_0^u) \in \mathcal{G}_{T+1}, x_0 \in \mathcal{X}_0, u \in \mathcal{U}} \left\{ \frac{V_1^m(u, \mathcal{X}_1, x_0)}{V_1^*(u, \mathcal{X}_1, x_0)} \right\} \;\le\; \frac{\frac{1}{2} + \varepsilon}{\left(\frac{1}{2} - \varepsilon\right) T} \;\to 0 \quad \text{as} \quad T \to \infty,$$

which concludes the proof. ■

Proof of Proposition 3. We consider a special case of the CRP in which: $u_t = u_0 = u$ for all $t = 1, \dots, T$; $\ell = 1$, i.e., every recommendation consists of a single link; and $w(x) = 1$ for any available article x. Whenever a an article x hosts a recommendation A that includes a link to article y, we denote for simplicity $\mathbb{P}_{u,x,y}(A) = \mathbb{P}_{u,x}(y)$. Recall that the set $\mathcal{X}$ is defined by $\mathcal{X} = \{(\gamma, \beta) \;:\; -1 \le \gamma \le 1,\; -1 \le \beta \le 1,\; \beta \le 2 - \varepsilon - \gamma\}$, for some $\varepsilon \in [0, 1]$. The set $\mathcal{X}$ is depicted in Figure B-4. In addition, let the efficient frontier set $\mathcal{X}^*$ be defined as:

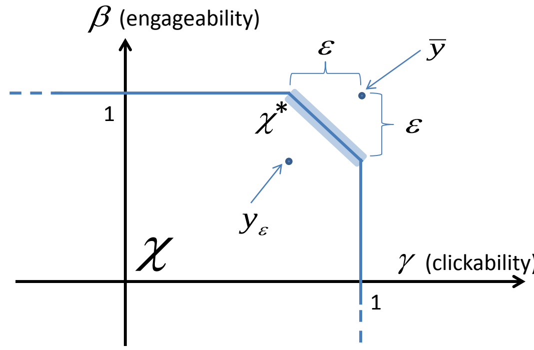

Figure B-4: **Convex set of available articles.** The optimal recommendation schedule is dominated by a policy that recommends $\bar{y}$ at any epoch (if $\bar{y}$ was an available article). The best one-step look-ahead schedule dominates a policy that recommends y_ε at each epoch.

$$\mathcal{X}^* = \{(\gamma, \beta) \;:\; -1 \leq \gamma \leq 1,\; -1 \leq \beta \leq 1,\; \beta = 2 - \varepsilon - \gamma\}\,.$$

The set $\mathcal{X}^*$ is depicted in Figure B-4. Consider the one-step look-ahead policy, defined by

$$V_t^{one}(u, \mathcal{X}, x_{t-1}) \;=\; \mathbb{P}_{u,x_{t-1}}(x_t)\left(1 + V_{t+1}^{one}(u, \mathcal{X}, x_t)\right),$$

for $t = 1, \ldots, T-1$, where

$$x_t \in \arg\max_{y \in \mathcal{X}} \left\{ \mathbb{P}_{u,x_{t-1}}(y) \left(1 + \max_{x_{t+1} \in \mathcal{X}} \left\{ \mathbb{P}_{u,y}(x_{t+1}) \right\} \right) \right\},$$

for $t = 1, \ldots, T-1$, and where the last recommendation is simply myopic. For each $t \in \{1, \ldots, T\}$ we denote by γ_t and β_t the clickability and the engageability values of x_t, the article which is recommended at step t. We denote the point $(1-\varepsilon, 1-\varepsilon)$ by y_ε (see Figure B-4). Since for any $u \in \mathcal{U}$, $V_t^{one}(u, \mathcal{X}, x_{t-1})$ is increasing in γ_t and β_t for all $t \in \{1, \ldots, T\}$, any point that is selected by the one-step lookahead policy belongs to the set $\mathcal{X}^*$. Moreover,

$$V_1^{one}(u, \mathcal{X}, x_0) \;\geq\; \mathbb{P}_{u,x_0}(y_\varepsilon) \prod_{t=2}^{T} \left(1 + \mathbb{P}_{u,y_\varepsilon}(y_\varepsilon)\right),$$

for any $u \in \mathcal{U}$ and $x_0 \in \mathcal{X}$. In words, the one-step look-ahead policy performs at least as well as a policy that selects y_ε at each epoch. Next, consider the optimal recommendation schedule, defined by

$$V_t^*(u, \mathcal{X}, x_{t-1}) \;=\; \max_{y \in \mathcal{X}} \left\{ \mathbb{P}_{u,x_{t-1}}(x_t)\left(1 + V_{t+1}^*(u, \mathcal{X}, x_t)\right) \right\},$$

for $t = 1, \ldots, T-1$, where the last recommendation is myopic. We denote the point $(1, 1)$ by $\bar{y}$ (see Figure B-4). Clearly, $\bar{y}$ does not belong to $\mathcal{X}$. Moreover, since $V_1^*(u, \mathcal{X}, x_0)$ is increasing in γ_t and β_t for all $t \in \{1, \ldots, T\}$, one has that

$$V_1^*(u, \mathcal{X}, x_0) \;\leq\; \mathbb{P}_{u,x_0}(\bar{y}) \prod_{t=2}^{T} \left(1 + \mathbb{P}_{u,\bar{y}}(\bar{y})\right),$$

for any $u \in \mathcal{U}$ and $x_0 \in \mathcal{X}$. In words, the optimal recommendation schedule performs at most as well as a policy that selects $\bar{y}$ at each epoch. Moreover, one has:

$$
\begin{aligned}
\frac{\mathbb{P}_{u,x_0}(y_\varepsilon)}{\mathbb{P}_{u,x_0}(\bar{y})} &= \frac{e^{\alpha+\theta_u+\beta_{x_0}+1-\varepsilon}}{1+e^{\alpha+\theta_u+\beta_{x_0}+1-\varepsilon}} \cdot \frac{1+e^{\alpha+\theta_u+\beta_{x_0}+1}}{e^{\alpha+\theta_u+\beta_{x_0}+1}} \\
&= e^{-\varepsilon} \cdot \frac{1+e^{\alpha+\theta_u+\beta_{x_0}+1}}{1+e^{\alpha+\theta_u+\beta_{x_0}+1-\varepsilon}} \;\geq\; e^{-\varepsilon},
\end{aligned} \tag{B-3}
$$

for any $u \in \mathcal{U}$ and $x_0 \in \mathcal{X}$. In addition, we have:

$$
\begin{aligned}
\frac{\mathbb{P}_{u,y_\varepsilon}(y_\varepsilon)}{\mathbb{P}_{u,\bar{y}}(\bar{y})} &= \frac{e^{\alpha+\theta_u+2-2\varepsilon}}{1+e^{\alpha+\theta_u+2-2\varepsilon}} \cdot \frac{1+e^{\alpha+\theta_u+2}}{e^{\alpha+\theta_u+2}} \\
&= e^{-2\varepsilon} \cdot \frac{1+e^{\alpha+\theta_u+2}}{1+e^{\alpha+\theta_u+2-2\varepsilon}} \;\geq\; e^{-2\varepsilon},
\end{aligned} \tag{B-4}
$$

for any $u \in \mathcal{U}$. Therefore, one has for any $u \in \mathcal{U}$, $x_0 \in \mathcal{X}$, and $\delta \geq \delta_\varepsilon$:

$$
\begin{aligned}
\frac{V_1^{one}(u,\mathcal{X},x_0)}{V_1^{*}(u,\mathcal{X},x_0)} &\geq \frac{\mathbb{P}_{u,x_0}(y_\varepsilon)\prod_{t=2}^{T}\left(1+\mathbb{P}_{u,y_\varepsilon}(y_\varepsilon)\right)}{\mathbb{P}_{u,x_0}(\bar{y})\prod_{t=2}^{T}\left(1+\mathbb{P}_{u,\bar{y}}(\bar{y})\right)} \\
&\overset{(a)}{\geq} e^{-2\varepsilon} \cdot \prod_{t=2}^{T}\left(\frac{1+\mathbb{P}_{u,y_\varepsilon}(y_\varepsilon)}{1+\mathbb{P}_{u,\bar{y}}(\bar{y})}\right) \\
&\overset{(b)}{\geq} e^{-2\varepsilon} \cdot \prod_{t=2}^{T}\left(\frac{1+e^{-2\varepsilon}\mathbb{P}_{u,\bar{y}}(\bar{y})}{1+\mathbb{P}_{u,\bar{y}}(\bar{y})}\right) \\
&\overset{(c)}{\geq} e^{-2\varepsilon} \cdot \prod_{t=2}^{T}\left(\frac{1+e^{-2\varepsilon}\bar{p}}{1+\bar{p}}\right) \;=\; e^{-2\varepsilon} \cdot \left(\frac{1+e^{-2\varepsilon}\bar{p}}{1+\bar{p}}\right)^{T-1},
\end{aligned}
$$

where: (a) holds by (B-3); (b) holds by (B-4); and (c) holds since $\mathbb{P}_{u,\bar{y}}(\bar{y}) \leq \bar{p}$ for all $u \in \mathcal{U}$. This concludes the proof. ■

References

Gary, M. R. and D. S. Johnson (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman and Company, New York.