Supplement to ‘Compound Sequential Change-point Detection in Parallel Data Streams’

Yunxiao Chen

London School of Economics and Political Science

Xiaoou Li

University of Minnesota

Contents

A Notations

For the readers' convenience, we provide a list of notations below. We will also restate these notation when they first appear in the proof.

- $X_{S,t}$ for some set $S \subset \{1, \cdots, K\}$: $X_{S,t} = (X_{k,t})_{k \in S}$.
- $X_{k,s:t}$: $X_{k,s:t} = (X_{k,r})_{s \leqslant r \leqslant t}$.
- $\mathbb{T}^{\mathbf{A}}$: an arbitrary sequential procedure.
- $S_t^{\mathbf{A}}$: The set of active streams at time t given by procedure $\mathbb{T}^{\mathbf{A}}$.
- $\mathcal{F}_t^{\mathbf{A}}$: σ-field of information obtained up to time t following $\mathbb{T}^{\mathbf{A}}$.
- $W_{k,t}^{\mathbf{A}}$: posterior probability $\mathbb{P}(\tau_k < t | \mathcal{F}_t^{\mathbf{A}})$ at the k-th stream following $\mathbb{T}^{\mathbf{A}}$ at time t.
- $W_{S,t}^{\mathbf{A}}$ for some set $S \subset \{1, \cdots, K\}$: $W_{S,t}^{\mathbf{A}} = (W_{k,t}^{\mathbf{A}})_{k \in S}$.
- $\mathbb{T}^*$: the proposed sequential procedure.
- S_t^*, $\mathcal{F}_t^*$, $W_{k,t}^*$, $W_{S,t}^*$ are defined similarly for procedure $\mathbb{T}^*$.
- $\mathbb{T}^{\mathbf{AP}_{t_0}}$: the sequential procedure that takes the same steps as $\mathbb{T}^{\mathbf{A}}$ up to time t_0 (meaning $S_t^{\mathbf{AP}_{t_0}} = S_t^{\mathbf{A}}$ for $1 \leqslant t \leqslant t_0$) and updates by Algorithm 1 from time $t_0 + 1$ and onward.
- $S_t^{\mathbf{AP}_{t_0}}$, $\mathcal{F}_t^{\mathbf{AP}_{t_0}}$, $W_{k,t}^{\mathbf{AP}_{t_0}}$, $W_{S,t}^{\mathbf{AP}_{t_0}}$ are defined similarly for procedure $\mathbb{T}^{\mathbf{AP}_{t_0}}$.
- $\overset{d}{=}$: equal in distribution.
- $\boldsymbol{\varnothing}$: a vector with zero length.
- dim: length of a vector, where $\dim(\boldsymbol{\varnothing}) = 0$.
- $Z \sim N(0,1)$: the notation '$\sim$' means that the left side follows the distribution on the right side.

B Proof Sketch

In this section, we discuss the main steps and techniques for proving Theorem 6 through an induction argument. Its proof is involved, relying on some monotone coupling results on stochastic processes living in a special partially ordered space. In what follows, we give a sketch of the proof to provide more insights into the proposed procedure. When $s = 0$, it is trivial that (5.2) holds. The induction is to show that for any $\mathbb{T}^{\mathbf{A}} \in \mathcal{T}_\alpha$ and any t_0, (5.2) holds for $s = s_0 + 1$, assuming that it holds for $s \leqslant s_0$. The induction step is proved by the following three steps.

1. Show that $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$ is 'better' than $\mathbb{T}^{\mathbf{A}}$ conditional on $\mathcal{F}_{t_0}^{\mathbf{A}}$.
2. Show that $\mathbb{T}^{\mathbf{AP}_{t_0}}$ is 'better' than $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$ conditional on $\mathcal{F}_{t_0}^{\mathbf{A}}$.
3. Show that $\mathbb{T}^{\mathbf{AP}_{t_0}}$ is 'better' than $\mathbb{T}^{\mathbf{A}}$ conditional on $\mathcal{F}_{t_0}^{\mathbf{A}}$ by combining the first two steps.

Here, we say a procedure is 'better than' the other, if its conditional expectation of the size of index set at time t_0+s+1 is no less than that of the other, given the information filtration $\mathcal{F}_{t_0}^{\mathbf{A}}$. Roughly, we prove the first step by replacing t_0 with t_0+1 in the induction assumption and taking conditional expectation given $\mathcal{F}_{t_0}^{\mathbf{A}}$, and prove the third step by combining the first and second steps. The main technical challenge lies in the second step, for which we develop several technical tools. Among these tools, an important one is the following monotone coupling result regarding a special partial order relationship.

We define a partially ordered space $(\mathcal{S}_{\mathrm{o}}, \preccurlyeq)$ as follows. Let

$$\mathcal{S}_{\mathrm{o}} = \bigcup_{k=1}^{K} \left\{ \mathbf{v} = (v_1, \cdots, v_k) \in [0,1]^k : 0 \leqslant v_1 \leqslant \cdots v_k \leqslant 1 \right\} \cup \{\varnothing\},$$

where $\varnothing$ represents a vector with zero length. For $\mathbf{u} \in \mathcal{S}_{\mathrm{o}}$, let $\dim(\mathbf{u})$ be the length of the vector $\mathbf{u}$.

Definition 1. *For $\mathbf{u}, \mathbf{v} \in \mathcal{S}_{\mathrm{o}}$, we say $\mathbf{u} \preccurlyeq \mathbf{v}$ if $\dim(\mathbf{u}) \geqslant \dim(\mathbf{v})$ and $u_i \leqslant v_i$ for $i = 1, ..., \dim(\mathbf{v})$. In addition, we say $\mathbf{u} \preccurlyeq \varnothing$ for any $\mathbf{u} \in \mathcal{S}_{\mathrm{o}}$.*

To emphasize the dependence on the sequential procedure, we use $S_t^{\mathbf{A}}$ and $\mathcal{F}_t^{\mathbf{A}}$ to denote the index set and the information filtration at time t given by the sequential procedure $\mathbb{T}^{\mathbf{A}}$. We further define $W_{k,t}^{\mathbf{A}} = \mathbb{P}\left(\tau_k < t \mid \mathcal{F}_t^{\mathbf{A}}\right)$. Similarly, we define the index set $S_t^{\mathbf{AP}_{t_0}}$, information filtration $\mathcal{F}_t^{\mathbf{AP}_{t_0}}$, and posterior probability $W_{k,t}^{\mathbf{AP}_{t_0}}$ given by the sequential procedure $\mathbb{T}^{\mathbf{AP}_{t_0}}$. For any vector $\mathbf{v} = (v_1, \cdots, v_m)$, we use the notation $[\mathbf{v}] = (v_{(1)}, \cdots, v_{(m)})$ for its order statistic. In addition, let $[\varnothing] = \varnothing$.

Proposition B.1. *Let $\{x_t, s_t, 1 \leqslant t \leqslant t_0\}$ be any sequence in the support of the stochastic process $\left\{(X_{k,t})_{k \in S_t^{\boldsymbol{A}}}, S_t^{\boldsymbol{A}}, 1 \leqslant t \leqslant t_0\right\}$ following a sequential procedure $\mathbb{T}^{\boldsymbol{A}} \in \mathcal{T}_\alpha$. Then, there* *exists a coupling of $\mathcal{S}_{\mathrm{o}}$-valued random variables $(\widehat{W}, \widehat{W}')$ such that*

$$\widehat{W} \stackrel{d}{=} \left[\left(W^{\boldsymbol{AP}_{t_0}}_{k,t_0+1} \right)_{k \in S^{\mathbf{AP}_{t_0}}_{t_0+1}} \right] \Bigg| \left\{ (X_{k,t})_{k \in S^{\boldsymbol{A}}_t} = x_t, S^{\boldsymbol{A}}_t = s_t, 1 \leqslant t \leqslant t_0 \right\},$$

$$\widehat{W}' \stackrel{d}{=} \left[\left(W^{\boldsymbol{AP}_{t_0+1}}_{k,t_0+1} \right)_{k \in S^{\boldsymbol{AP}_{t_0+1}}_{t_0+1}} \right] \Bigg| \left\{ (X_{k,t})_{k \in S^{\boldsymbol{A}}_t} = x_t, S^{\boldsymbol{A}}_t = s_t, 1 \leqslant t \leqslant t_0 \right\},$$

and $\widehat{W} \preccurlyeq \widehat{W}'$ a.s., where $\stackrel{d}{=}$ denotes that random variables on both sides are identically distributed.

We clarify that by the above proposition, the resulting $\widehat{W}$ and $\widehat{W}'$ are defined on the same probability space. Let

$$Y_s = \left[\left(W^{\mathbf{AP}_{t_0}}_{k,t_0+s} \right)_{k \in S^{\mathbf{AP}_{t_0}}_{t_0+s}} \right] \in \mathcal{S}_{\mathrm{o}}.$$

Under model $\mathcal{M}_{\mathrm{s}}$, the stochastic process Y_s is stochastically monotone in that the following monotone coupling result holds.

Proposition B.2. *Suppose that model $\mathcal{M}_s$ holds. Then for any $\mathbf{y}, \mathbf{y}' \in \mathcal{S}_{\mathrm{o}}$ such that $\mathbf{y} \preccurlyeq \mathbf{y}'$, there exists a coupling $(\widehat{Y}_s, \widehat{Y}'_s)$, $s = 0, 1, \ldots$, satisfying*

1. *$\{\widehat{Y}_s : s \geqslant 0\}$ has the same distribution as the conditional process $\{Y_s : s \geqslant 0\}$ given $Y_0 = \mathbf{y}$, and $\{\widehat{Y}'_s : s \geqslant 0\}$ has the same distribution as the conditional process $\{Y_s : s \geqslant 0\}$ given $Y_0 = \mathbf{y}'$.*
2. *$\widehat{Y}_s \preccurlyeq \widehat{Y}'_s$, a.s. for all $s \geqslant 0$.*

Moreover, the process $(\widehat{Y}_s, \widehat{Y}'_s)$ does not depend on $\mathbb{T}^{\boldsymbol{A}}$, t_0, or the information filtration $\mathcal{F}^{\boldsymbol{A}}_{t_0}$.

Roughly, Proposition B.1 shows that the sequential procedure $\mathbb{T}^{\mathbf{AP}_{t_0}}$ tends to have a stochastically smaller detection statistic, in terms of the partial order $\preccurlyeq$, than that of $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$ at time $t_0 + 1$, and thus tends to keep more active streams. Proposition B.2 further shows that this trend will be carried over to any future time, including time $t_0 + s + 1$. The second step of induction is proved by formalizing this heuristic.

C Proof of Theorem 6

Theorem 6. *Suppose that model $\mathcal{M}_s$ holds. For any $t_0, s \geqslant 0$ and any sequential detection procedure $\mathbb{T}^{\boldsymbol{A}} \in \mathcal{T}_\alpha$, let $\mathcal{F}^{\boldsymbol{A}}_t$ be the information filtration and $S^{\boldsymbol{A}}_t$ be the set of active streams* *at time t given by $\mathbb{T}^{\boldsymbol{A}}$. Then,*

$$\mathbb{E}\left[|S^{\boldsymbol{A}}_{t_0+s}|\middle|\mathcal{F}^{\boldsymbol{A}}_{t_0}\right] \leqslant \mathbb{E}\left[|S^{\mathbf{AP}_{t_0}}_{t_0+s}|\middle|\mathcal{F}^{\boldsymbol{A}}_{t_0}\right] \ a.s.$$

Proof of Theorem 6. We will prove the theorem by inducting on s.

For the base case ($s = 0$) the theorem is obviously true for all t_0 and all $\mathbb{T}^{\mathbf{A}} \in \mathcal{T}_\alpha$ as the both sides of (5.2) are exactly the same.

We will prove the induction step in the rest of the proof. Assume (5.2) is true for any strategy $\mathbb{T}^{\mathbf{A}} \in \mathcal{T}_\alpha$ and any t_0, for some $s = s_0$. Our goal is to prove that it is also true for any t_0, for $s = s_0 + 1$, using the following steps, where we recall that $\mathbb{T}^{\mathbf{AP}_{t_0}}$ is defined as the sequential procedure that takes the same steps as $\mathbb{T}^{\mathbf{A}}$ up to time t_0 and updates by Algorithm 1 from time $t_0 + 1$ and onward, and the sequential procedure $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$ is defined similarly.

Step 1: comparing $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$ and $\mathbb{T}^{\mathbf{A}}$. For $s = s_0 + 1$, since we assume (5.2) is true for all t_0, we could replace t_0 by $t_0 + 1$ and s by s_0 in (5.2) and arrive at

$$\mathbb{E}\left[\left|S^{\mathbf{A}}_{t_0+s_0+1}\right|\middle|\mathcal{F}^{\mathbf{A}}_{t_0+1}\right] \leqslant \mathbb{E}\left[\left|S^{\mathbf{AP}_{t_0+1}}_{t_0+s_0+1}\right|\middle|\mathcal{F}^{\mathbf{A}}_{t_0+1}\right] \text{ a.s.}$$

Taking conditional expectation $\mathbb{E}\left[\cdot|\mathcal{F}^{\mathbf{A}}_{t_0}\right]$ on both sides, we arrive at

$$\mathbb{E}\left[\left|S^{\mathbf{A}}_{t_0+s_0+1}\right|\middle|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \leqslant \mathbb{E}\left[\left|S^{\mathbf{AP}_{t_0+1}}_{t_0+s_0+1}\right|\middle|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \text{ a.s.} \tag{C.2}$$

Step 2: comparing $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$ and $\mathbb{T}^{\mathbf{AP}_{t_0}}$. First, define a function $\phi_{t,s} : \mathcal{S}_{\mathrm{o}} \to \mathbb{R}$,

$$\phi_{t,s}(\mathbf{u}) = \mathbb{E}\Big[|S^{\mathbf{AP}_t}_{t+s}|\Big|\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_t,t}\big] = \mathbf{u}\Big] = \mathbb{E}\Big[\dim\big(\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s},t+s}\big]\big)\Big|\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_t,t}\big] = \mathbf{u}\Big] \tag{C.3}$$

for $t, s \geqslant 0$. Here, for a set S, and time points s and t, $W^{\mathbf{AP}_t}_{S,s} = \big(W^{\mathbf{AP}_t}_{k,s}\big)_{k\in S}$, where $W^{\mathbf{AP}_t}_{k,s} = \mathbb{P}\left(\tau_k < s\middle|\mathcal{F}^{\mathbf{AP}_t}_s\right)$.

From Proposition B.2, we can see that $\phi_{t,s}(\mathbf{u})$ does not depend on the sequential procedure $\mathbb{T}^{\mathbf{A}}$ and the value of t. Thus, by replacing $\mathbb{T}^{\mathbf{A}}$ with $\mathbb{T}^{\mathbf{AP}_{t_0}}$, t with $t_0 + 1$, and s with s_0 in (E.2), we obtain

$$\phi_{t_0+1,s_0}(\mathbf{u}) = \mathbb{E}\Big[\dim\big(\big[W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1},t_0+s_0+1}\big]\big)\Big|\big[W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1},t_0+1}\big] = \mathbf{u}\Big]. \tag{C.4}$$

Here, to see the superscript of the process in the above equation is $\mathbf{AP}_{t_0}$, we used the fact that if we follow the procedure $\mathbb{T}^{\mathbf{AP}_{t_0}}$ and switch to the proposed procedure at time $t_0 + 1$, then the overall sequential procedure is still $\mathbb{T}^{\mathbf{AP}_{t_0}}$.

Also from Proposition B.2, we can see that for any $\mathbf{u} \preccurlyeq \mathbf{u}' \in \mathcal{S}_{\mathrm{o}}$, there exists a coupling $(\widehat{Y}_s, \widehat{Y}'_s)$ such that $\widehat{Y}_s$ has the same distribution as $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s}, t+s}\big]$ given $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t}, t}\big] = \mathbf{u}$, $\widehat{Y}'_s$ has the same distribution as $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s}, t+s}\big]$ given $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t}, t}\big] = \mathbf{u}'$, and $\widehat{Y}_s \preccurlyeq \widehat{Y}'_s$ a.s. Thus,

$$\phi_{t,s}(\mathbf{u}) = \mathbb{E}\big(\dim(\widehat{Y}_s)\big) \text{ and } \phi_{t,s}(\mathbf{u}') = \mathbb{E}\big(\dim(\widehat{Y}'_s)\big).$$

According to the definition of the partial relationship '$\preccurlyeq$', $\widehat{Y}_s \preccurlyeq \widehat{Y}'_s$ implies $\dim(\widehat{Y}_s) \geqslant \dim(\widehat{Y}'_s)$. Combining this result with the above display, we conclude that $\phi_{t,s}(\mathbf{u}) \geqslant \phi_{t,s}(\mathbf{u}')$ for any $\mathbf{u} \preccurlyeq \mathbf{u}' \in \mathcal{S}_{\mathrm{o}}$.

Next, we write $\mathbb{E}\Big[\big|S^{\mathbf{AP}_{t_0+1}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\Big]$ and $\mathbb{E}\Big[\big|S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\Big]$ in terms of the conditional expectation involving the function $\phi_{t,s}$. We start with $\mathbb{E}\Big[\big|S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\Big]$. By the iterative law of conditional expectation and (C.4), we obtain

$$\begin{aligned}
\mathbb{E}\Big[\big|S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\Big] =& \mathbb{E}\Big[\mathbb{E}\Big\{\big|S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1}\big|\Big|\big[W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1}, t_0+1}\big]\Big\}\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\Big] \\
=& \mathbb{E}\Big[\mathbb{E}\Big\{\dim\big(\big[W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1}, t_0+s_0+1}\big]\big)\Big|\big[W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1}, t_0+1}\big]\Big\}\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\Big] \\
=& \mathbb{E}\Big[\phi_{t_0+1,s_0}\Big(\big[W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1}, t_0+1}\big]\Big)\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\Big].
\end{aligned}$$

According to the definition of the information filtration $\mathcal{F}^{\mathbf{A}}_{t_0}$, we further write the above conditional expectation as

$$\mathbb{E}\Big[\big|S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\Big] = \mathbb{E}\Big[\phi_{t_0+1,s_0}\Big(\big[W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1}, t_0+1}\big]\Big)\Big|\big\{S^{\mathbf{A}}_r, X_{k,r}, k \in S^{\mathbf{A}}_r, 1 \leqslant r \leqslant t_0\big\}\Big]. \tag{C.5}$$

Similarly, we have

$$\mathbb{E}\Big[\big|S^{\mathbf{AP}_{t_0+1}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\Big] = \mathbb{E}\Big[\phi_{t_0+1,s_0}\Big(\big[W^{\mathbf{AP}_{t_0+1}}_{S^{\mathbf{AP}_{t_0+1}}_{t_0+1}, t_0+1}\big]\Big)\Big|\big\{S^{\mathbf{A}}_r, X_{k,r}, k \in S^{\mathbf{A}}_r, 1 \leqslant r \leqslant t_0\big\}\Big]. \tag{C.6}$$

We proceed to a comparison between (C.5) and (C.6). According to Proposition B.1, for each sequence $\{x_r, s_r, 1 \leqslant r \leqslant t_0\}$ that is in the support of the process $\{X_{S^{\mathbf{A}}_r, r}, S^{\mathbf{A}}_r, 1 \leqslant r \leqslant t_0\}$ there exists a coupling $(\widehat{W}, \widehat{W}')$ such that

$$\widehat{W} \overset{d}{=} [W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1}, t_0+1}]\Big|\{X_{S^{\mathbf{A}}_r, r} = x_r, S^{\mathbf{A}}_r = s_r, 1 \leqslant r \leqslant t_0\},$$

$$\widehat{W}' \overset{d}{=} [W^{\mathbf{AP}_{t_0+1}}_{S^{\mathbf{AP}_{t_0+1}}_{t_0+1}, t_0+1}] \Big| \{X_{S^{\mathbf{A}}_r, r} = x_r, S^{\mathbf{A}}_t = s_r, 1 \leqslant r \leqslant t_0\},$$

and

$$\widehat{W} \leqslant \widehat{W}' \text{ a.s.},$$

where '$\overset{d}{=}$' means two random variables on both sides have the same distribution. Thus,

$$\mathbb{E}\left[\phi_{t_0+1,s_0}\Big([W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1}, t_0+1}]\Big)\Big| X_{S^{\mathbf{A}}_r, r} = x_r, S^{\mathbf{A}}_r = s_r, 1 \leqslant r \leqslant t_0\right] = \mathbb{E}\phi_{t_0+1,s_0}\big(\widehat{W}\big) \tag{C.7}$$

and

$$\mathbb{E}\left[\phi_{t_0+1,s_0}\Big([W^{\mathbf{AP}_{t_0+1}}_{S^{\mathbf{AP}_{t_0+1}}_{t_0+1}, t_0+1}]\Big)\Big| X_{S^{\mathbf{A}}_r, r} = x_r, S^{\mathbf{A}}_r = s_r, 1 \leqslant r \leqslant t_0\right] = \mathbb{E}\phi_{t_0+1,s_0}\big(\widehat{W}'\big). \tag{C.8}$$

On the other hand, note that we have shown $\phi_{t_0+1,s_0}(\mathbf{u}) \geqslant \phi_{t_0+1,s_0}(\mathbf{u}')$ for any $\mathbf{u} \leqslant \mathbf{u}' \in \mathcal{S}_{\text{o}}$ and $\widehat{W} \leqslant \widehat{W}'$ a.s. by the coupling. Thus,

$$\phi_{t_0+1,s_0}(\widehat{W}) \geqslant \phi_{t_0+1,s_0}(\widehat{W}') \text{ a.s.}$$

Combining the above inequality with (C.7) and (C.8), we arrive at

$$\begin{aligned}
&\mathbb{E}\left[\phi_{t_0+1,s_0}\Big([W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1}, t_0+1}]\Big)\Big| X_{S^{\mathbf{A}}_r, r} = x_r, S^{\mathbf{A}}_r = s_r, 1 \leqslant r \leqslant t_0\right] \\
\geqslant &\mathbb{E}\left[\phi_{t_0+1,s_0}\Big([W^{\mathbf{A}}_{S^{\mathbf{A}}_{t_0+1}, t_0+1}]\Big)\Big| X_{S^{\mathbf{A}}_r, r} = x_r, S^{\mathbf{A}}_r = s_r, 1 \leqslant r \leqslant t_0\right]
\end{aligned}$$

for each sequence $\{x_r, s_r, 1 \leqslant r \leqslant t_0\}$ that is in the support of the process $\{X_{S^{\mathbf{A}}_r, r}, S^{\mathbf{A}}_r, 1 \leqslant r \leqslant t_0\}$. Comparing the above inequality with (C.5) and (C.6), we conclude that

$$\mathbb{E}\left[\big|S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \geqslant \mathbb{E}\left[\big|S^{\mathbf{AP}_{t_0+1}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \quad a.s. \tag{C.9}$$

Step 3: combining results from Steps 1 and 2. Combining (C.2) and (C.9), we obtain

$$\mathbb{E}\left[\big|S^{\mathbf{A}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \leqslant \mathbb{E}\left[\big|S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1}\big|\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \text{ a.s.}, \tag{C.10}$$

which implies that (5.2) holds for arbitrary $\mathbb{T}^{\mathbf{A}} \in \mathcal{T}_\alpha$, t_0, and $s = s_0 + 1$. This completes the induction. □

Remark C.1. *Proposition B.2 is used in Step 2 of the above proof, where we only use the property that $\widehat{Y}_s \leqslant \widehat{Y}'_s$ is independent of t_0 and $\mathbb{T}^{\boldsymbol{A}}$. The independence between $(\widehat{Y}_s, \widehat{Y}'_s)$ and $\mathcal{F}^{\mathbf{A}}_{t_0}$ is an additional result that further characterizes the coupling process. We did not use* *this additional property directly in the proof.*

D Proof of Theorems 1 and 5

It suffices to prove Theorem 5, as Theorem 1 is straightforwardly implied by Theorem 5.

Theorem 5. *Let $\mathbb{T}^{\boldsymbol{A}} \in \mathcal{T}_\alpha$ be an arbitrary sequential procedure. Further let $\mathbb{T}^{\boldsymbol{AP}_{t_0}}$ and $\mathbb{T}^{\boldsymbol{AP}_{t_0+1}}$ be the switching procedures described above, with switching time t_0 and t_0+1, respectively, for some $t_0 \geqslant 0$. Then, $\mathbb{T}^{\boldsymbol{AP}_{t_0}}, \mathbb{T}^{\boldsymbol{AP}_{t_0+1}} \in \mathcal{T}_\alpha$ and under model $\mathcal{M}_s$*

$$\mathbb{E}\left(\mathrm{U}_t(\mathbb{T}^{\boldsymbol{A}})\right) \leqslant \mathbb{E}\left(\mathrm{U}_t(\mathbb{T}^{\boldsymbol{AP}_{t_0+1}})\right) \leqslant \mathbb{E}\left(\mathrm{U}_t(\mathbb{T}^{\boldsymbol{AP}_{t_0}})\right) \leqslant \mathbb{E}\left(\mathrm{U}_t(\mathbb{T}^{*})\right),$$

for all $t = 1, 2, \cdots$.

Proof of Theorem 5. First, note that $\mathbb{T}^{\mathbf{A}} \in \mathcal{T}_\alpha$ and $\mathbb{T}^{\mathbf{AP}_{t_0}}$ agrees with $\mathbb{T}^{\mathbf{A}} \in \mathcal{T}_\alpha$ up to time t_0. Thus, $\mathbb{T}^{\mathbf{AP}_{t_0}}$ control the LFNR to be no greater than α from time 1 to t_0. Also, according to Proposition 1, $\mathbb{T}^{\mathbf{AP}_{t_0}}$ controls the LFNR at level α from time t_0+1 and onward. Thus, $\mathbb{T}^{\mathbf{AP}_{t_0}} \in \mathcal{T}_\alpha$. Similarly, $\mathbb{T}^{\mathbf{AP}_{t_0+1}} \in \mathcal{T}_\alpha$.

Applying Theorem 6 but replacing t_0 by t_0+1, and taking expectation on both sides of the inequality, we obtain

$$\mathbb{E}|S^{\mathbf{A}}_{t_0+1+s}| \leqslant \mathbb{E}|S^{\mathbf{AP}_{t_0+1}}_{t_0+1+s}|$$

for every $t_0 \geqslant 0$ and $s \geqslant 0$. That is, for every $t \geqslant t_0+1$,

$$\mathbb{E}|S^{\mathbf{A}}_t| \leqslant \mathbb{E}|S^{\mathbf{AP}_{t_0+1}}_t|.$$

For $t < t_0+1$, as $\mathbb{T}^{\mathbf{A}}$ and $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$ share the same index set, we have

$$\mathbb{E}|S^{\mathbf{A}}_t| = \mathbb{E}|S^{\mathbf{AP}_{t_0+1}}_t|.$$

Combining the above inequalities, we obtain

$$\mathbb{E}|S^{\mathbf{A}}_t| \leqslant \mathbb{E}|S^{\mathbf{AP}_{t_0+1}}_t|$$

for all $t \geqslant 0$. This further implies

$$\mathbb{E}\{\mathrm{U}_t(\mathbb{T}^{\mathbf{A}})\} = \sum_{s=1}^{t} \mathbb{E}|S^{\mathbf{A}}_s| \leqslant \sum_{s=1}^{t} \mathbb{E}|S^{\mathbf{AP}_{t_0+1}}_s| = \mathbb{E}\{\mathrm{U}_t(\mathbb{T}^{\mathbf{AP}_{t_0+1}})\}.$$

This proves the inequality for comparing procedures $\mathbb{T}^{\mathbf{A}}$ and $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$. We then compare $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$ and $\mathbb{T}^{\mathbf{AP}_{t_0}}$, based on the same arguments above except that we replace $\mathbb{T}^{\mathbf{A}}$ by $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$, and replace $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$ by $\mathbb{T}^{\mathbf{AP}_{t_0}}$. We obtain

$$\mathbb{E}\{\mathrm{U}_t(\mathbb{T}^{\mathbf{AP}_{t_0+1}})\} \leqslant \mathbb{E}\{\mathrm{U}_t(\mathbb{T}^{\mathbf{AP}_{t_0}})\}$$

for all $t \geqslant 0$.

Finally, we compare $\mathbb{T}^{\mathbf{AP}_{t_0}}$ and $\mathbb{T}^* = \mathbb{T}^{\mathbf{AP}_0}$ using a similar argument, which gives

$$\mathbb{E}\{\mathrm{U}_t(\mathbb{T}^{\mathbf{AP}_{t_0}})\} \leqslant \mathbb{E}\{\mathrm{U}_t(\mathbb{T}^*)\}.$$

□

E Proof of Theorem 4

First, by Theorem 6, we directly see that $\mathbb{E}\{|S_t^{\mathbf{A}}|\} \leqslant \mathbb{E}\{|S_t^*|\}$, for any sequential procedure $\mathbb{T}^{\mathbf{A}} \in \mathcal{T}_\alpha$. Thus, $\mathbb{E}(\mathrm{CD}_t(\mathbb{T}^*)) = K - \mathbb{E}|S_t^*| \leqslant \mathbb{E}(\mathrm{CD}_t(\mathbb{T}^{\mathbf{A}}))$, which further implies $\mathbb{E}(\mathrm{CD}_t(\mathbb{T}^*)) = \inf_{\mathbb{T}\in\mathcal{T}_\alpha} \mathbb{E}(\mathrm{CD}_t(\mathbb{T}))$.

We proceed to the analysis of $\mathrm{RL}_t(\mathbb{T})$. By interchanging the order of double summation, we have

$$\mathrm{RL}_t(\mathbb{T}) = \sum_{k=1}^{K}(T_k\wedge\tau_k\wedge t) = \sum_{k=1}^{K}\sum_{s=1}^{t}\mathbb{1}(s\leqslant T_k\wedge\tau_k) = \sum_{s=1}^{t}\sum_{k=1}^{K}\mathbb{1}(s\leqslant T_k\wedge\tau_k) = \sum_{s=1}^{t}\sum_{k\in S_s}\{1-\mathbb{1}(\tau_k<s)\}$$

which leads to

$$\mathbb{E}\{\mathrm{RL}_t(\mathbb{T})\} = \sum_{s=1}^{t}\mathbb{E}\big[\sum_{k\in S_s}\{1-\mathbb{1}(\tau_k<s)\}\big] = \sum_{s=1}^{t}\mathbb{E}\Big[\mathbb{E}\big[\sum_{k\in S_s}\{1-\mathbb{1}(\tau_k<s)\}\big]|\mathcal{F}_s\Big].$$

Recall that $W_{k,s} = \mathbb{P}(\tau_k < s|\mathcal{F}_s)$ and $S_s \in \mathcal{F}_s$. The above display yields

$$\mathbb{E}\{\mathrm{RL}_t(\mathbb{T})\} = \sum_{s=1}^{t}\mathbb{E}\big\{\sum_{k\in S_s}(1-W_{k,s})\big\}.$$

From the above equation, we can see that in order to show $\mathbb{E}\{\mathrm{RL}_t(\mathbb{T}^*)\} = \sup_{\mathbb{T}\in\mathcal{T}_\alpha}\mathbb{E}\{\mathrm{RL}_t(\mathbb{T})\}$, it suffices to show $\mathbb{E}\big\{\sum_{k\in S_t}(1-W_{k,t})\big\}$ is maximized for every $t = 1, 2, \cdots$, which follows directly from the following extension of Theorem 6.

Proposition E.1. *Suppose that model $\mathcal{M}_s$ holds. For any $t_0, s \geqslant 0$ and any sequential detection procedure $\mathbb{T}^{\boldsymbol{A}} \in \mathcal{T}_\alpha$, let $\mathcal{F}_t^{\boldsymbol{A}}$ be the information filtration and $S_t^{\boldsymbol{A}}$ be the set of active* *streams at time t given by $\mathbb{T}^{\boldsymbol{A}}$. Then,*

$$\mathbb{E}\left[\Psi([W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t_0+s},t_0+s}])\Big|\mathcal{F}^{\boldsymbol{A}}_{t_0}\right] \leqslant \mathbb{E}\left[\Psi([W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+s},t_0+s}])\Big|\mathcal{F}^{\boldsymbol{A}}_{t_0}\right] \ \ a.s.,$$

where $\Psi : \mathcal{S}_{\mathrm{o}} \to \mathbb{R}$ is defined as $\Psi(\mathbf{w}) = \sum_{k=1}^{m}(1 - w_k)$ for $\mathbf{w} = (w_1, \cdots, w_m) \in \mathcal{S}_{\mathrm{o}}$.

In the rest of the section, we provide the proof of Proposition E.1.

Proof of Proposition E.1. The proof of Proposition E.1 is similar to that of Theorem 6. We will only state the main differences and omit the repetitive details.

First, by replacing $|S^{\mathbf{A}}_t|$ with $\Psi([W^{\mathbf{A}}_{S^{\mathbf{A}}_t,t}])$ for t taking different values in the proof of Theorem 6, we obtain the following inequality that is similar to (C.2)

$$\mathbb{E}\left[\Psi([W^{\mathbf{A}}_{S^{\mathbf{A}}_{t_0+s_0+1},t_0+s_0+1}])\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \leqslant \mathbb{E}\left[\Psi(W^{\mathbf{AP}_{t_0+1}}_{S_{t_0+s_0+1},t_0+s_0+1}\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \text{ a.s.} \tag{E.1}$$

for all t_0 and s_0. Next, we define a function $\widetilde{\phi}_{t,s} : \mathcal{S}_{\mathrm{o}} \to \mathbb{R}$,

$$\widetilde{\phi}_{t,s}(\mathbf{u}) = \mathbb{E}\Big[\Psi([W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s},t+s}])\Big|[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_t,t}] = \mathbf{u}\Big] \tag{E.2}$$

for $t, s \geqslant 0$. Then, we replace the ϕ with $\widetilde{\phi}$ in the proof of Theorem 6 and obtain the following inequality that is similar to (C.9).

$$\mathbb{E}\left[\Psi([W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1},t_0+s_0+1}])\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \geqslant \mathbb{E}\left[\Psi([W^{\mathbf{AP}_{t_0+1}}_{S^{\mathbf{AP}_{t_0+1}}_{t_0+s_0+1},t_0+s_0+1}])\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \ \ a.s. \tag{E.3}$$

We point out that to arrive at the above inequality, the following property about Ψ is used: $\Psi(\mathbf{w}') \leqslant \Psi(\mathbf{w})$ for any $\mathbf{w}, \mathbf{w}' \in \mathcal{S}_{\mathrm{o}}$ satisfying $\mathbf{w} \leqslant \mathbf{w}'$.

Combining (E.1) and (E.3), we obtain

$$\mathbb{E}\left[\Psi([W^{\mathbf{A}}_{S^{\mathbf{A}}_{t_0+s_0+1},t_0+s_0+1}])\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \leqslant \mathbb{E}\left[\Psi([W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+s_0+1},t_0+s_0+1}])\Big|\mathcal{F}^{\mathbf{A}}_{t_0}\right] \text{ a.s.,}$$

which extends (C.10) and completes the proof. □

F Proof of Propositions B.1 and B.2

The proof of Propositions B.1 and B.2 is involved. We will first introduce some concepts in stochastic ordering, followed by several useful lemmas, and then present the proof of the propositions.

F.1 Stochastic ordering

We first review a few important concepts and classic results on partially ordered spaces. More details about stochastic ordering and coupling can be found in Kamae et al. (1977); Lindvall (1999, 2002); Thorisson (2000).

Definition 2 (Partially Ordered Space (pospace))**.** *A space $(\mathcal{S}, \preccurlyeq)$ is said to be a partially ordered space (or pospace) if $\preccurlyeq$ is a partial order relation over the topological space $\mathcal{S}$ and the set $\{(x, y) \in \mathcal{S}^2 : x \preccurlyeq y\}$ is a closed subset of $\mathcal{S}^2$.*

Definition 3 (Increasing functions over a partially ordered space)**.** *Let $(\mathcal{S}_1, \preccurlyeq_{\mathcal{S}_1})$ and $(\mathcal{S}_2, \preccurlyeq_{\mathcal{S}_2})$ be partially ordered polish spaces. A map $g : \mathcal{S}_1 \to \mathcal{S}_2$ is said to be increasing if $g(u) \preccurlyeq_{\mathcal{S}_2} g(v)$ for all $u \preccurlyeq_{\mathcal{S}_1} v$ with $u, v \in \mathcal{S}_1$.*

Definition 4 (Stochastic ordering of real-valued random variables)**.** *Let X and Y be two random variables, we say X is stochastically less than or equal to Y, if $\mathbb{P}(X \geqslant x) \leqslant \mathbb{P}(Y \geqslant x)$ for all real number x. In this case, we write $X \leqslant_{st} Y$.*

The following statements give some equivalent definitions for $X \leqslant_{st} Y$

Fact 1. *The following statements are equivalent.*

1. *$X \leqslant_{st} Y$.*
2. *For all increasing, bounded, and measurable functions $g : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}(g(X)) \leqslant \mathbb{E}(g(Y))$.*
3. *There exists a coupling $(\widehat{X}, \widehat{Y})$ such that $\widehat{X} \overset{d}{=} X$, $\widehat{Y} \overset{d}{=} Y$, and*

$$\widehat{X} \leqslant \widehat{Y} \ a.s.$$

Here, $\overset{d}{=}$ denotes that the random variables on both sides have an identical distribution.

In particular, the equivalence between 1 and 3 is known as the Strassen's Theorem (Strassen, 1965).

Definition 5 (Stochastic ordering on a partially ordered polish space)**.** *Let $(\mathcal{S}, \preccurlyeq)$ be a partially ordered polish space, and let X and Y be $\mathcal{S}$-valued random variables. We say Y stochastically dominates X, denoted by $X \preccurlyeq_{st} Y$ if for all bounded, increasing, and measurable function $g : \mathcal{S} \to \mathbb{R}$, $\mathbb{E}(g(X)) \leqslant \mathbb{E}(g(Y))$.*

Fact 2 (Strassen's theorem for polish pospace, Theorem 2.4 in Lindvall (2002))**.** *Let $(\mathcal{S}, \preccurlyeq)$ be a polish partially ordered space, and let X and Y be $\mathcal{S}$-valued random variables. Then, $X \preccurlyeq_{st} Y$ if and only if there exists a coupling $(\widehat{X}, \widehat{Y})$ such that $\widehat{X} \overset{d}{=} X$, $\widehat{Y} \overset{d}{=} Y$ and $\widehat{X} \preccurlyeq \widehat{Y}$ a.s.*

Definition 6 (Stochastic dominance for Markov kernels)**.** *Let K and $\widetilde{K}$ be transition kernels for Markov chains over a partially ordered polish space $(\mathcal{S}, \preccurlyeq)$. The transition kernel $\widetilde{K}$ is said to stochastically dominate K if*

$$x \preccurlyeq y \implies K(x, \cdot) \preccurlyeq_{st} \widetilde{K}(y, \cdot).$$

In particular, if the above is satisfied for the same kernel $K = \widetilde{K}$, then we say K is stochastically monotone.

Fact 3 (Strassen's theorem for Markov chains over a polish pospace)**.** *Let $\{X_t\}$ and $\{Y_t\}$ be Markov chains over a partially ordered polish space, $(\mathcal{S}, \preccurlyeq)$, with transition kernels K and $\widetilde{K}$ where $\widetilde{K}$ stochastically dominates K. Then, for all initial points $x_0 \preccurlyeq y_0$, there is a coupling $\{(\widehat{X}_t, \widehat{Y}_t)\}$ of $\{X_t\}$ starting at x_0 and $\{Y_t\}$ starting at y_0 such that*

$$\widehat{X}_t \preccurlyeq \widehat{Y}_t \qquad \forall t \quad a.s.$$

Fact 3 is a special case of Theorem 5.8 in Lindvall (2002).

F.2 Stochastic ordering and Markov chains on $\mathcal{S}_{\mathrm{u}}$ and $\mathcal{S}_{\mathrm{o}}$

In this section, we provide some supporting lemmas regarding properties of the partial order relationship defined in Section B, and show stochastic ordering of several Markov chains. The proof of these lemmas is given in Section F.4.

Recall that in Section B, we define a space

$$\mathcal{S}_{\mathrm{o}} = \bigcup_{k=1}^{K} \left\{ \mathbf{v} = (v_1, \cdots, v_k) \in [0,1]^k : 0 \leqslant v_1 \leqslant \cdots v_k \leqslant 1 \right\} \cup \{\varnothing\}.$$

Here, we also define a space with unordered elements.

$$\mathcal{S}_{\mathrm{u}} = \bigcup_{k=1}^{K} [0,1]^k \cup \{\varnothing\}.$$

We first present a lemma showing that the space $(\mathcal{S}_{\mathrm{o}}, \preccurlyeq)$ is a polish partial order space.

Lemma F.1. $(\mathcal{S}_{\mathrm{o}}, \preccurlyeq)$ *is a partially ordered space. In addition, $\mathcal{S}_{\mathrm{o}}$ is a polish space equipped* *with the metric*

$$d(\mathbf{u},\mathbf{v}) = \begin{cases} \max_{1\leqslant m\leqslant \dim(\mathbf{u})} |u_m - v_m| & \textit{if } \dim(\mathbf{u}) = \dim(\mathbf{v}) \geqslant 1 \\ 0 & \textit{if } \mathbf{u} = \mathbf{v} = \boldsymbol{\varnothing} \\ 2 & \textit{if } \dim(\mathbf{u}) \neq \dim(\mathbf{v}) \end{cases}$$

for $\mathbf{u},\mathbf{v} \in \mathcal{S}_{\mathrm{o}}$.

We define mappings $I_{\mathrm{o}} : \mathcal{S}_{\mathrm{o}} \to \{0,\cdots,K\}$ and $H_{\mathrm{o}} : \mathcal{S}_{\mathrm{o}} \to \mathcal{S}_{\mathrm{o}}$ as follows. For any $\mathbf{u} \in \mathcal{S}_{\mathrm{o}}$, define

$$I_{\mathrm{o}}(\mathbf{u}) = \begin{cases} \sup\left\{n : \sum_{i=1}^{n} u_i \leqslant \alpha n, n \in \{0,...,\dim(\mathbf{u})\}\right\} & \text{if } \dim(\mathbf{u}) \geqslant 1, \mathbf{u} = (u_1,...,u_{\dim(\mathbf{u})}) \\ 0 & \text{if } \dim(\mathbf{u}) = 0, \end{cases} \tag{F.1}$$

and

$$H_{\mathrm{o}}(\mathbf{u}) = \begin{cases} (u_1,\cdots,u_{I_{\mathrm{o}}(\mathbf{u})}) & \text{if } I_{\mathrm{o}}(\mathbf{u}) \geqslant 1, \\ \boldsymbol{\varnothing} & \text{otherwise.} \end{cases}$$

The mapping H_{o} is closely related to the one-step update rule in Algorithm 1, as summarized in the next lemma.

Lemma F.2. *If we input* $(W_{k,t})_{k\in S_t} = \mathbf{u}$ *and an index set* S_t *with* $|S_t| = \dim(\mathbf{u})$ *in Algorithm 1, then the output* S_{t+1} *satisfies*

$$|S_{t+1}| = I_{\mathrm{o}}(\mathbf{u}) \textit{ and } [(W_{k,t})_{k\in S_{t+1}}] = H_{\mathrm{o}}([\mathbf{u}]).$$

Other compound sequential detection rules in $\mathcal{T}_\alpha$ are characterized through the next lemma.

Lemma F.3. $\mathbb{T} = (T_1,\cdots,T_K) \in \mathcal{T}_\alpha$ *if and only if*

$$\mathbb{T} \in \mathcal{T} \textit{ and } \sum_{k=1}^{K} \mathbb{1}(T_k > t) W_{k,t} \leqslant \alpha \cdot \sum_{k=1}^{K} \mathbb{1}(T_k > t) \textit{ for } t = 0,1,2,\cdots$$

The above expression is equivalent to

$$S_{t+1} \textit{ is } \mathcal{F}_t \textit{ measurable },\ S_{t+1} \subset S_t, \sum_{k\in S_{t+1}} W_{k,t} \leqslant \alpha \cdot |S_{t+1}|$$

for $t = 0,1,2,\cdots,$ *and* $T_k = \sup\{t : k \in S_t\}$.

The next lemma compares the second statement in the above lemma with the output of the function H_{o}.

Lemma F.4. *Let* $\mathbf{u} = (u_1, \cdots, u_m) \in \mathcal{S}_{\mathrm{u}}$ *with* $\dim(\mathbf{u}) = m \geqslant 1$. *Let* $k_1, \cdots, k_l \in \{1, \cdots, m\}$ *be distinct and satisfy*

$$\sum_{i=1}^{l} u_{k_i} \leqslant \alpha l.$$

Then, $H_{\mathrm{o}}([\mathbf{u}]) \preccurlyeq [(u_{k_1}, \cdots, u_{k_l})]$. *Moreover, if* $H_{\mathrm{o}}([\mathbf{u}]) = \varnothing$, *then for any* $S \subset \{1, \cdots, m\}$ *with* $|S| \geqslant 1$, $\sum_{i\in S} u_i > \alpha|S|$.

Lemma F.5. *For any* $\mathbf{u} \preccurlyeq \mathbf{v} \in \mathcal{S}_{\mathrm{o}}$, $H_{\mathrm{o}}(\mathbf{u}) \preccurlyeq H_{\mathrm{o}}(\mathbf{v})$. *That is, the mapping* $H_{\mathrm{o}}(\mathbf{u})$ *is increasing in* $\mathbf{u}$.

Next, we present several lemmas on the stochastic ordering of random variables and Markov chains. We start with a simple but useful result regarding the stochastic monotonicity of a likelihood ratio under a mixture model.

Lemma F.6. *Let* $p(x)$ *and* $q(x)$ *be two density functions with respect to some baseline measure* μ *and assume that* $p(\cdot)$ *and* $q(\cdot)$ *have the same support. Let* $L(x) = \frac{q(x)}{p(x)}$ *be the likelihood ratio. For* $\delta \in [0, 1]$, *let* Z_δ *be a random variable with the density function* $\delta q + (1-\delta)p$ *and* $L_\delta = L(Z_\delta)$. *Then, for* $0 \leqslant \delta_1 < \delta_2 \leqslant 1$, *we have*

$$L_{\delta_1} \leqslant_{st} L_{\delta_2}.$$

This result is intuitive: if we have more weights in q for the mixture distribution, then the likelihood ratio will be larger, giving more evidence in favor of q.

Lemma F.7. *Assume model* $\mathcal{M}_s$ *holds. Let* $V_{k,t} = \mathbb{P}\left(\tau_k < t | X_{k,1}, \cdots, X_{k,t}\right)$. *Then,*

$$V_{k,0} = 0 \text{ and } V_{k,t+1} = \frac{q(X_{k,t+1})/p(X_{k,t+1})}{(1-\theta)(1-V_{k,t})/(\theta + (1-\theta)V_{k,t}) + q(X_{k,t+1})/p(X_{k,t+1})}. \tag{F.2}$$

Moreover, $\{V_{k,t}\}_{t=0,1,\cdots}$ *are independent and identically distributed processes for different* k.

Lemma F.8. *Assume model* $\mathcal{M}_s$ *holds. Let* $\delta_{k,t} = \mathbb{P}\left(\tau_k \leqslant t | X_{k,1}, \cdots, X_{k,t}\right)$, *then*

$$\delta_{k,t} = \theta + (1-\theta)V_{k,t},$$

where $V_{k,t}$ *is defined in* (F.2).

Lemma F.9. *Under model $\mathcal{M}_s$, the process $\{V_{1,t}\}_{t\geqslant 0}$ defined in* (F.2) *is a homogeneous Markov chain. In addition, its transition kernel is stochastically monotone. We will later refer to this transition kernel as $K(\cdot,\cdot)$.*

Lemma F.10. *For any $t \geqslant 1$ and $\mathbb{T}^{\boldsymbol{A}}$, $[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t+1}]$ is conditionally independent of $\mathcal{F}^{\boldsymbol{A}}_t$ given $[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t}]$. Moreover, the conditional density of $\big[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t+1}\big]$ at $\mathbf{v}$ given $\big[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t}\big] = \mathbf{u} \in \mathcal{S}_{\mathrm{o}}$ is*

$$\mathbb{K}_a(\mathbf{u},\mathbf{v}) := \begin{cases} \sum_{\pi\in\mathbf{P}_{\dim(\mathbf{u})}} \prod_{l=1}^{\dim(\mathbf{u})} K(u_l, v_{\pi(l)}) & \text{if } \dim(\mathbf{u}) = \dim(\mathbf{v}) \geqslant 1, \\ 1 & \text{if } \dim(\mathbf{u}) = \dim(\mathbf{v}) = 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{P}_m$ denotes the set of all permutations over $\{1,\cdots,m\}$.

Lemma F.11. *For each $\mathbf{u} \in \mathcal{S}_{\mathrm{o}}$ with $\dim(\mathbf{u}) = m \geqslant 1$, generate an $\mathcal{S}_{\mathrm{o}}$-valued random variable V as follows,*

1. *For each $k \in \{1,\cdots,m\}$, generate $Z_k \sim K(u_k,\cdot)$ independently for different k.*
2. *Let $V = [(Z_1, ..., Z_m)]$.*

In addition, if $m = 0$, we let $V = \boldsymbol{\varnothing}$. Then, $V \sim \mathbb{K}_a(\mathbf{u},\cdot)$.

Lemma F.12. *For $\mathbf{u},\mathbf{u}' \in \mathcal{S}_{\mathrm{o}}$ with $\mathbf{u} \preccurlyeq \mathbf{u}'$, we have $\mathbb{K}_a(\mathbf{u},\cdot) \preccurlyeq_{st} \mathbb{K}_a(\mathbf{u}',\cdot)$.*

Lemma F.13. *For any $t, s \geqslant 0$ and $\mathbb{T}^{\boldsymbol{A}}$, $\big[W^{\boldsymbol{A}\mathbf{P}_t}_{S^{\boldsymbol{A}\mathbf{P}_t}_{t+s+1},t+s+1}\big]$ is conditionally independent of $\mathcal{F}^{\boldsymbol{A}\mathbf{P}_t}_{t+s}$ given $\big[W^{\boldsymbol{A}\mathbf{P}_t}_{S^{\boldsymbol{A}\mathbf{P}_t}_{t+s},t+s}\big]$. Moreover, the conditional density of $\big[W^{\boldsymbol{A}\mathbf{P}_t}_{S^{\boldsymbol{A}\mathbf{P}_t}_{t+s+1},t+s+1}\big]$ at $\mathbf{v}$ given $\big[W^{\boldsymbol{A}\mathbf{P}_t}_{S^{\boldsymbol{A}\mathbf{P}_t}_{t+s},t+s}\big] = \mathbf{u} \in \mathcal{S}_{\mathrm{o}}$ is*

$$\mathbb{K}_o(\mathbf{u},\mathbf{v}) := \mathbb{K}_a(H_{\mathrm{o}}(\mathbf{u}),\mathbf{v}) = \begin{cases} \sum_{\pi\in\mathbf{P}_{I_{\mathrm{o}}(\mathbf{u})}} \prod_{l=1}^{I_{\mathrm{o}}(\mathbf{u})} K(H_{\mathrm{o}}(\mathbf{u})_l, v_{\pi(l)}) & \text{if } \dim(\mathbf{v}) = I_{\mathrm{o}}(\mathbf{u}) \geqslant 1 \\ 1 & \text{if } \dim(\mathbf{v}) = I_{\mathrm{o}}(\mathbf{u}) = 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{P}_m$ denotes the set of all permutations over $\{1,\cdots,m\}$.

Remark F.1. *There is a key difference between Lemma F.10 and Lemma F.13, though they may look similar at a first glance. In Lemma F.10, we consider the conditional distribution of $\big[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t+1}\big]$ given $\big[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t}\big]$, where the index set $S^{\boldsymbol{A}}_{t+1}$ is the same for the two random vectors.*

In Lemma F.13, we consider the conditional distribution of $\left[W^{\boldsymbol{AP}_t}_{S^{\mathbf{AP}_t}_{t+s+1},t+s+1}\right]$ *given* $\left[W^{\boldsymbol{AP}_t}_{S^{\mathbf{AP}_t}_{t+s},t+s}\right]$*, where the two random vectors are associated with two different index sets* $S^{\mathbf{AP}_t}_{t+s+1}$ *and* $S^{\mathbf{AP}_t}_{t+s}$*. This difference reflects a key difference between the proposed one-step update rule and an arbitrary procedure.*

Lemma F.14. *For each* $\mathbf{u} \in \mathcal{S}_{\mathrm{o}}$ *and* $m = I_{\mathrm{o}}(\mathbf{u})$*, generate an* $\mathcal{S}_{\mathrm{o}}$*-valued random variable* V *as follows,*

1. *For each* $k \in \{1, \cdots, m\}$*, generate* $Z_k \sim K(H_{\mathrm{o}}(\mathbf{u})_k, \cdot)$ *independently for different* k*.*
2. *Let* $V = [(Z_1, ..., Z_m)]$*.*

In addition, if $m = 0$*, we let* $V = \boldsymbol{\varnothing}$*. Then,* $V \sim \mathbb{K}_o(\mathbf{u}, \cdot)$*.*

F.3 Proof of Propositions B.1 and B.2

Proposition B.1. *Let* $\{x_t, s_t, 1 \leqslant t \leqslant t_0\}$ *be any sequence in the support of the stochastic process* $\left\{(X_{k,t})_{k \in S^{\boldsymbol{A}}_t}, S^{\boldsymbol{A}}_t, 1 \leqslant t \leqslant t_0\right\}$ *following a sequential procedure* $\mathbb{T}^{\boldsymbol{A}} \in \mathcal{T}_\alpha$*. Then, there exists a coupling of* $\mathcal{S}_{\mathrm{o}}$*-valued random variables* $(\widehat{W}, \widehat{W}')$ *such that*

$$
\begin{aligned}
\widehat{W} &\overset{d}{=} \left[\left(W^{\boldsymbol{AP}_{t_0}}_{k,t_0+1}\right)_{k \in S^{\boldsymbol{AP}_{t_0}}_{t_0+1}}\right] \Bigg| \left\{(X_{k,t})_{k \in S^{\boldsymbol{A}}_t} = x_t, S^{\boldsymbol{A}}_t = s_t, 1 \leqslant t \leqslant t_0\right\},\\
\widehat{W}' &\overset{d}{=} \left[\left(W^{\boldsymbol{AP}_{t_0+1}}_{k,t_0+1}\right)_{k \in S^{\boldsymbol{AP}_{t_0+1}}_{t_0+1}}\right] \Bigg| \left\{(X_{k,t})_{k \in S^{\boldsymbol{A}}_t} = x_t, S^{\boldsymbol{A}}_t = s_t, 1 \leqslant t \leqslant t_0\right\},
\end{aligned}
$$

and $\widehat{W} \leqslant \widehat{W}'$ *a.s., where* $\overset{d}{=}$ *denotes that random variables on both sides are identically distributed.*

Proof of Proposition B.1. First, given $\left\{X_{S^{\mathbf{A}}_t} = x_t, S^{\mathbf{A}}_t = s_t, 1 \leqslant t \leqslant t_0\right\}$, $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t_0},t_0}\right]$ is determined. To simplify the notation, we assume $W^{\mathbf{A}}_{S^{\mathbf{A}}_{t_0},t_0} = \mathbf{w}_{t_0} \in \mathcal{S}_{\mathrm{u}}$ given $\left\{X_{S^{\mathbf{A}}_t} = x_t, S^{\mathbf{A}}_t = s_t, 1 \leqslant t \leqslant t_0\right\}$.

In addition, $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t_0+1},t_0}\right]$ is determined by $\left\{X_{S^{\mathbf{A}}_t} = x_t, S^{\mathbf{A}}_t = s_t, 1 \leqslant t \leqslant t_0\right\}$ and the sequential procedure $\mathbb{T}^{\mathbf{A}}$. To simplify the notation, we assume $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t_0+1},t_0}\right] = \mathbf{w}^*_{t_0+1} \in \mathcal{S}_{\mathrm{o}}$ given $\{X_{S^{\mathbf{A}}_t} = x_t, S^{\mathbf{A}}_t = s_t, 1 \leqslant t \leqslant t_0\}$. We clarify that $\mathbf{w}^*_{t_0+1}$ is a deterministic (and measurable) function of x_t, s_t for $1 \leqslant t \leqslant t_0$ (depending on the sequential procedure $\mathbb{T}^{\mathbf{A}}$). According to Lemma F.10 (replacing t by t_0), the conditional distribution of $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t_0+1},t_0+1}\right]$ given $\left\{X_{S^{\mathbf{A}}_t} = x_t, S^{\mathbf{A}}_t = s_t, 1 \leqslant t \leqslant t_0\right\}$ is the same as the conditional distribution given $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t_0+1},t_0}\right] = \mathbf{w}^*_{t_0+1}$. Moreover, the conditional density is $\mathbb{K}_{\mathrm{a}}(\mathbf{w}^*_{t_0+1}, \cdot)$.

We perform a similar analysis by replacing $\mathbb{T}^{\mathbf{A}}$ by $\mathbb{T}^{\mathbf{AP}_{t_0}}$ in the above analysis. We denote $\big[W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1},t_0}\big] = \mathbf{w}_{t_0+1}$ and obtain that the conditional density of $\big[W^{\mathbf{AP}_{t_0}}_{S^{\mathbf{AP}_{t_0}}_{t_0+1},t_0+1}\big]$ given $\big\{X_{S^{\mathbf{A}}_t} = x_t, S^{\mathbf{A}}_t = s_t, 1 \leqslant t \leqslant t_0\big\}$ is $\mathbb{K}_{\mathrm{a}}(\mathbf{w}_{t_0+1}, \cdot)$.

According to the above analysis and Strassen Theorem for pospace (Fact 2), to prove the proposition, it is sufficient to show $\mathbb{K}_{\mathrm{a}}(\mathbf{w}_{t_0+1}, \cdot) \preccurlyeq_{st} \mathbb{K}_{\mathrm{a}}(\mathbf{w}^*_{t_0+1}, \cdot)$. By Lemma F.12, we have $\mathbb{K}_{\mathrm{a}}(\mathbf{u}, \cdot) \preccurlyeq_{st} \mathbb{K}_{\mathrm{a}}(\mathbf{u}', \cdot)$ for any $\mathbf{u} \preccurlyeq \mathbf{u}' \in \mathcal{S}_{\mathrm{o}}$. Thus, it is sufficient to show that $\mathbf{w}_{t_0+1} \preccurlyeq \mathbf{w}^*_{t_0+1}$.

Now we compare $\mathbf{w}_{t_0+1}$ and $\mathbf{w}^*_{t_0+1}$. According to the definition of $\mathbb{T}^{\mathbf{AP}_{t_0}}$ and Lemma F.2, we know $\mathbf{w}_{t_0+1} = H_{\mathrm{o}}([\mathbf{w}_{t_0}])$. There are two cases: 1) $\mathbf{w}^*_{t_0+1} = \varnothing$, and 2) $\mathbf{w}^*_{t_0+1} \neq \varnothing$. We analyze these cases separately. For the first case, $\mathbf{w}_{t_0+1} \preccurlyeq \mathbf{w}^*_{t_0+1}$ by definition of the partial order. For the second case, according to Lemma F.3 and Lemma F.4, we can see that $\mathbf{w}_{t_0+1} = H_{\mathrm{o}}([\mathbf{w}_{t_0}]) \neq \varnothing$. Write $\mathbf{w}_{t_0} = (w_{t_0,1}, \cdots, w_{t_0,m})$ for some m, then $\mathbf{w}^*_{t_0+1}$ can be written as $\mathbf{w}^*_{t_0+1} = (w_{t_0,k_1}, \cdots, w_{t_0,k_l})$ for some distinct $k_1, \cdots, k_l \in \{1, \cdots, m\}$. According to Lemma F.3, for $\mathbb{T}^{\mathbf{A}}$ to control LFNR at time $t_0 + 1$, $\mathbf{w}^*_{t_0+1}$ satisfies $\sum_{i=1}^{l} w_{t_0,k_i} \leqslant \alpha l$. Thus, according to Lemma F.4, $\mathbf{w}_{t_0+1} = H_{\mathrm{o}}([\mathbf{w}_{t_0}]) \preccurlyeq [\mathbf{w}^*_{t_0+1}] = \mathbf{w}^*_{t_0+1}$. □

Proposition B.2. *Suppose that model $\mathcal{M}_s$ holds. Then for any $\mathbf{y}, \mathbf{y}' \in \mathcal{S}_{\mathrm{o}}$ such that $\mathbf{y} \preccurlyeq \mathbf{y}'$, there exists a coupling $(\widehat{Y}_s, \widehat{Y}'_s)$, $s = 0, 1, \ldots$, satisfying*

1. *$\{\widehat{Y}_s : s \geqslant 0\}$ has the same distribution as the conditional process $\{Y_s : s \geqslant 0\}$ given $Y_0 = \mathbf{y}$, and $\{\widehat{Y}'_s : s \geqslant 0\}$ has the same distribution as the conditional process $\{Y'_s : s \geqslant 0\}$ given $Y_0 = \mathbf{y}'$.*
2. *$\widehat{Y}_s \preccurlyeq \widehat{Y}'_s$, a.s. for all $s \geqslant 0$.*

Moreover, the process $(\widehat{Y}_s, \widehat{Y}'_s)$ does not depend on $\mathbb{T}^{\boldsymbol{A}}$, t_0, or the information filtration $\mathcal{F}^{\boldsymbol{A}}_{t_0}$.

Proof of Proposition B.2. Recall $Y_s = \left[\left(W^{\mathbf{AP}_{t_0}}_{k,t_0+s}\right)_{k \in S^{\mathbf{AP}_{t_0}}_{t_0+s}}\right]$. By letting $t = t_0$ in Lemma F.13, we obtain that $\{Y_s\}_{s \geqslant 0}$ is a homogeneous Markov chain, whose transition kernel is $\mathbb{K}_{\mathrm{o}}$, which is independent of the sequential procedure $\mathbb{T}^{\mathbf{A}}$, t_0, and the information filtration $\mathcal{F}^{\mathbf{A}}_{t_0}$. For the rest of the proof, according to Definition 6 and Fact 3, it is sufficient to show that $\mathbb{K}_{\mathrm{o}}$ is stochastically monotone. That is, $\mathbb{K}_{\mathrm{o}}(\mathbf{u}, \cdot) \preccurlyeq_{st} \mathbb{K}_{\mathrm{o}}(\mathbf{u}', \cdot)$ for any $\mathbf{u}, \mathbf{u}' \in \mathcal{S}_{\mathrm{o}}$ with $\mathbf{u} \preccurlyeq \mathbf{u}'$. Thus, it is sufficient to show that for all $\mathbf{u} \preccurlyeq \mathbf{u}'$ there exists a coupling $(\widehat{V}, \widehat{V}')$ such that $\widehat{V} \sim \mathbb{K}_{\mathrm{o}}(\mathbf{u}, \cdot)$, $\widehat{V}' \sim \mathbb{K}_{\mathrm{o}}(\mathbf{u}', \cdot)$ and $\widehat{V} \preccurlyeq \widehat{V}'$ a.s. In what follows, we construct such a coupling.

For $\mathbf{u} \preccurlyeq \mathbf{u}'$ with $\mathbf{u}, \mathbf{u}' \in \mathcal{S}_{\mathrm{o}}$, we know that $H_{\mathrm{u}}(\mathbf{u}) \preccurlyeq H_{\mathrm{u}}(\mathbf{u}')$ by Lemma F.5. By the definition of the partial order, this implies that $\dim(H_{\mathrm{u}}(\mathbf{u}')) \leqslant \dim(H_{\mathrm{u}}(\mathbf{u}))$ and $H_{\mathrm{u}}(\mathbf{u})_k \leqslant H_{\mathrm{u}}(\mathbf{u}')_k$ for each $1 \leqslant k \leqslant \dim(H_{\mathrm{u}}(\mathbf{u}'))$. By Lemma F.9, this further implies

$$K(H_{\mathrm{u}}(\mathbf{u})_k, \cdot) \preccurlyeq_{st} K(H_{\mathrm{u}}(\mathbf{u}')_k, \cdot)$$

for $k = 1, ..., \dim(H_{\mathrm{u}}(\mathbf{u}'))$. Thus, by Strassen's Theorem for random variables (Fact 1), this implies that there exists a coupling $(\widehat{Z}_k, \widehat{Z}'_k)$ such that

$$\widehat{Z}_k \sim K(H_{\mathrm{u}}(\mathbf{u})_k, \cdot), \widehat{Z}'_k \sim K(H_{\mathrm{u}}(\mathbf{u}')_k, \cdot), \text{ and } \widehat{Z}_k \leqslant \widehat{Z}'_k \text{ a.s.}$$

for $k = 1, ..., \dim(H_{\mathrm{u}}(\mathbf{u}'))$. In addition, we choose the coupling so that $(\widehat{Z}_k, \widehat{Z}'_k)$ are independent for different k. For $\dim(H_{\mathrm{u}}(\mathbf{u}')) < k \leqslant \dim(H_{\mathrm{u}}(\mathbf{u}))$, we construct $\widehat{Z}_k \sim K(H_{\mathrm{u}}(\mathbf{u})_k, \cdot)$ so that $\widehat{Z}_k$'s are independent for different k. Let $\widehat{Z} = (\widehat{Z}_1, \cdots, \widehat{Z}_{\dim(H_{\mathrm{u}}(\mathbf{u}))})$ and $\widehat{Z}' = (\widehat{Z}'_1, \cdots, \widehat{Z}'_{\dim(H_{\mathrm{u}}(\mathbf{u}'))})$.

For this coupling, it is easy to verify

$$\dim(\widehat{Z}) \geqslant \dim(\widehat{Z}') \text{ and } \widehat{Z}_k \leqslant \widehat{Z}'_k \text{ for } 1 \leqslant k \leqslant \dim(\widehat{Z}') \;\; a.s.$$

Thus, $[\widehat{Z}] \preccurlyeq [\widehat{Z}']$ a.s. Let $\widehat{V} = [\widehat{Z}]$ and $\widehat{V}' = [\widehat{Z}']$. Then, our coupling $(\widehat{V}, \widehat{V}')$ gives

$$\widehat{V} \preccurlyeq \widehat{V}' \text{ a.s.}$$

On the other hand, by Lemma F.14, we have

$$\widehat{V} \sim \mathbb{K}_{\mathrm{o}}(\mathbf{u}, \cdot) \text{ and } \widehat{V}' \sim \mathbb{K}_{\mathrm{o}}(\mathbf{u}', \cdot).$$

Therefore,

$$\mathbb{K}_{\mathrm{o}}(\mathbf{u}, \cdot) \preccurlyeq_{st} \mathbb{K}_{\mathrm{o}}(\mathbf{u}', \cdot).$$

□

F.4 Proof of supporting lemmas in Section F.2

Lemma F.1. $(\mathcal{S}_{\mathrm{o}}, \preccurlyeq)$ *is a partially ordered space. In addition,* $\mathcal{S}_{\mathrm{o}}$ *is a polish space equipped with the metric*

$$d(\mathbf{u}, \mathbf{v}) = \begin{cases} \max_{1 \leqslant m \leqslant \dim(\mathbf{u})} |u_m - v_m| & \textit{if } \dim(\mathbf{u}) = \dim(\mathbf{v}) \geqslant 1 \\ 0 & \textit{if } \mathbf{u} = \mathbf{v} = \boldsymbol{\varnothing} \\ 2 & \textit{if } \dim(\mathbf{u}) \neq \dim(\mathbf{v}) \end{cases}$$

for $\mathbf{u}, \mathbf{v} \in \mathcal{S}_{\mathrm{o}}$.

Proof of Lemma F.1. First, $\mathcal{S}_{\mathrm{o}}$ is the union of polish spaces $\{\mathbf{u} = (u_1, \cdots, u_m) : 0 \leqslant u_1 \leqslant \cdots \leqslant u_m \leqslant 1\}$ and $\{\boldsymbol{\varnothing}\}$. Thus, it is also a polish space. Second, it is straightforward to verify that $d(\mathbf{u},\mathbf{v})$ is a metric defined over $\mathcal{S}_{\rm o}$.

Now, we verify that the partial order relationship $\preccurlyeq$ is closed over $\mathcal{S}_{\rm o}$. To see this, let $\mathbf{u},\mathbf{v}\in\mathcal{S}_{\rm o}$ satisfying $\mathbf{u}\npreccurlyeq\mathbf{v}$. There are two cases: 1) $\dim(\mathbf{u})<\dim(\mathbf{v})$, or 2) $\dim(\mathbf{u})\geqslant\dim(\mathbf{v})$ and there exists $m\in\{1,\cdots,\dim(\mathbf{v})\}$ such that $u_m>v_m$. Let $B_d(\mathbf{u},\delta)$ and $B_d(\mathbf{v},\delta)$ be d-balls centering at $\mathbf{u}$ and $\mathbf{v}$ with δ chosen according to different cases: $\delta=1/2$ for the first case; and $\delta=\frac{u_m-v_m}{4}$ for the second case. Then, it is easy to verify that for all $\mathbf{u}'\in B_d(\mathbf{u},\delta)$ and $\mathbf{v}'\in B_d(\mathbf{v},\delta)$, we have $\mathbf{u}'\npreccurlyeq\mathbf{v}'$. That is, the partial order relationship $\preccurlyeq$ is closed over $\mathcal{S}_{\rm o}$. $\square$

Lemma F.2. *If we input $(W_{k,t})_{k\in S_t}=\mathbf{u}$ and an index set S_t with $|S_t|=\dim(\mathbf{u})$ in Algorithm 1, then the output S_{t+1} satisfies*

$$|S_{t+1}|=I_{\rm o}(\mathbf{u})\ \textit{and}\ [(W_{k,t})_{k\in S_{t+1}}]=H_{\rm o}([\mathbf{u}]).$$

Proof of Lemma F.2. If $\mathbf{u}=\varnothing$, then $[\mathbf{u}]=\varnothing$ and $|S_t|=0$. This implies $I_{\rm o}([\mathbf{u}])=0$ and $H_{\rm o}([\mathbf{u}])=\varnothing$. In the rest of the proof we assume that $\mathbf{u}\neq\varnothing$. By Step 1 of Algorithm 1, we obtain that $[\mathbf{u}]=(W_{k_1,t},\cdots,W_{k_{|S_t|},t})$ where $S_t=\{k_1,\cdots,k_{|S_t|}\}$ and $W_{k_1,t}\leqslant\cdots W_{k_{|S_t|},t}$. According to Step 2 and 3 of the algorithm and the definition of $I_{\rm o}([\mathbf{u}])$ in (F.1), the largest n making $R_n\leqslant\alpha$ is $I_{\rm o}([\mathbf{u}])$ and $H_{\rm o}([\mathbf{u}])=[(W_{k,t})_{k\in S_{t+1}}]$. $\square$

Lemma F.3. $\mathbb{T}=(T_1,\cdots,T_K)\in\mathcal{T}_\alpha$ *if and only if*

$$\mathbb{T}\in\mathcal{T}\ \textit{and}\ \sum_{k=1}^K\mathbb{1}(T_k>t)W_{k,t}\leqslant\alpha\cdot\sum_{k=1}^K\mathbb{1}(T_k>t)\ \textit{for}\ t=0,1,2,\cdots$$

The above expression is equivalent to

$$S_{t+1}\ \textit{is}\ \mathcal{F}_t\ \textit{measurable}\ ,\ S_{t+1}\subset S_t,\sum_{k\in S_{t+1}}W_{k,t}\leqslant\alpha\cdot|S_{t+1}|$$

for $t=0,1,2,\cdots,$ and $T_k=\sup\{t:k\in S_t\}$.

Proof of Lemma F.3. By definition and the $\mathcal{F}_t$ measurability of S_{t+1},

$$\mathrm{LFNR}_{t+1}(\mathbb{T})=\mathbb{E}\left[\frac{\sum_{k\in S_{t+1}}\mathbb{1}(\tau_k<t)}{|S_{t+1}|\vee 1}\Big|\mathcal{F}_t\right]=\frac{\sum_{k\in S_{t+1}}\mathbb{P}(\tau_k<t|\mathcal{F}_t)}{|S_{t+1}|\vee 1}=\frac{\sum_{k\in S_{t+1}}W_{k,t}}{|S_{t+1}|\vee 1}.$$

Thus, $\mathbb{T}\in\mathcal{T}_\alpha$ if and only if

$$\frac{\sum_{k\in S_{t+1}}W_{k,t}}{|S_{t+1}|\vee 1}\leqslant\alpha\ \text{a.s.},$$

which is equivalent to

$$\sum_{k \in S_{t+1}} W_{k,t} \leqslant \alpha |S_{t+1}| \;\; a.s.,$$

for every t. $\square$

Lemma F.4. *Let* $\mathbf{u} = (u_1, \cdots, u_m) \in \mathcal{S}_{\mathrm{u}}$ *with* $\dim(\mathbf{u}) = m \geqslant 1$*. Let* $k_1, \cdots, k_l \in \{1, \cdots, m\}$ *be distinct and satisfy*

$$\sum_{i=1}^{l} u_{k_i} \leqslant \alpha l.$$

Then, $H_{\mathrm{o}}([\mathbf{u}]) \preccurlyeq [(u_{k_1}, \cdots, u_{k_l})]$*. Moreover, if* $H_{\mathrm{o}}([\mathbf{u}]) = \varnothing$*, then for any* $S \subset \{1, \cdots, m\}$ *with* $|S| \geqslant 1$*,* $\sum_{i \in S} u_i > \alpha |S|$*.*

Proof of Lemma F.4. We first prove the 'Moreover' part of the lemma by contradiction. If on the contrary $H_{\mathrm{o}}([\mathbf{u}]) = \varnothing$ and there exists a non-empty set $S \subset \{1, \cdots, m\}$ such that $\sum_{i \in S} u_i \leqslant \alpha |S|$, then there exists $i \in S$ such that $u_i \leqslant \alpha$. This further implies $[\mathbf{u}]_1 \leqslant u_i \leqslant \alpha$ and $I_{\mathrm{o}}([\mathbf{u}]) \geqslant 1$, which contracts with the assumption $H_{\mathrm{o}}([\mathbf{u}]) = \varnothing$.

We proceed to the proof of the rest of the lemma. We first prove that l in the lemma satisfies $l \leqslant I_{\mathrm{o}}([\mathbf{u}])$. To see this, recall that $([\mathbf{u}]_1, \cdots, [\mathbf{u}]_m)$ is the order statistic of $(u_1, \cdots, u_m)$. Thus,

$$\sum_{i=1}^{l} [\mathbf{u}]_i \leqslant \sum_{i=1}^{l} u_{k_i} \leqslant \alpha l. \tag{F.3}$$

Recall $I_{\mathrm{o}}([\mathbf{u}]) = \sup\{n : \sum_{i=1}^{n} [\mathbf{u}]_i \leqslant \alpha n, n \in \{0, \cdots, m\}\}$. Thus, (F.3) implies $l \leqslant I_{\mathrm{o}}([\mathbf{u}])$.

Next, we prove that $H_{\mathrm{o}}([\mathbf{u}]) \preccurlyeq [(u_{k_1}, \cdots, u_{k_l})]$. Without loss of generality, assume $u_{k_1}, \cdots, u_{k_l}$ are ordered. That is, $u_{k_1} \leqslant \cdots \leqslant u_{k_l}$ and $[(u_{k_1}, \cdots, u_{k_l})] = (u_{k_1}, \cdots, u_{k_l})$. Then, according to the definition of the order statistic $[\mathbf{u}]$, we have $[\mathbf{u}]_i \leqslant u_{k_i}$ for $i = 1, \cdots, l$. Recall $H_{\mathrm{o}}([\mathbf{u}]) = ([\mathbf{u}]_1, \cdots, [\mathbf{u}]_{I_{\mathrm{o}}(\mathbf{u})})$. This implies $H_{\mathrm{o}}([\mathbf{u}]) \preccurlyeq [(u_{k_1}, \cdots, u_{k_l})]$. $\square$

Lemma F.5. *For any* $\mathbf{u} \preccurlyeq \mathbf{v} \in \mathcal{S}_{\mathrm{o}}$*,* $H_{\mathrm{o}}(\mathbf{u}) \preccurlyeq H_{\mathrm{o}}(\mathbf{v})$*. That is, the mapping* $H_{\mathrm{o}}(\mathbf{u})$ *is increasing in* $\mathbf{u}$*.*

Proof of Lemma F.5. If $\mathbf{v} = \varnothing$, then $H_{\mathrm{o}}(\mathbf{v}) = \varnothing$ and $H_{\mathrm{o}}(\mathbf{u}) \preccurlyeq \varnothing = H_{\mathrm{o}}(\mathbf{v})$ by the definition of the partial order. In the rest of the proof we assume $\dim(\mathbf{v}) \geqslant 1$ and $\mathbf{v} = (v_1, \cdots, v_{\dim(\mathbf{v})})$. As we assumed $\mathbf{u} \preccurlyeq \mathbf{v}$, this implies $\dim(\mathbf{u}) \geqslant \dim(\mathbf{v}) \geqslant 1$. We further denote $\mathbf{u} = (u_1, \cdots, u_{\dim(\mathbf{u})})$

We first show that if $\sum_{i=1}^{L+1} v_i \leqslant \alpha(L+1)$ for some L, then $\sum_{i=1}^{L} v_i \leqslant \alpha L$. That is, $(\sum_{i=1}^{L} v_i)/L$ is increasing in L. To see this, consider two cases. If $v_{L+1} \leqslant \alpha$, then $v_1 \leqslant \cdots \leqslant v_L \leqslant \alpha$ and thus $\sum_{i=1}^{L} v_i \leqslant \alpha L$. If $v_{L+1} > \alpha$, then $\sum_{i=1}^{L} v_i \leqslant \sum_{i=1}^{L+1} v_i - \alpha \leqslant \alpha L$. This result implies that $\sum_{i=1}^{L} v_i \leqslant \alpha L$ for all $1 \leqslant L \leqslant I_{\mathrm{o}}(\mathbf{v})$.

Now we show that $I_{\rm o}(\mathbf{u}) \geqslant I_{\rm o}(\mathbf{v})$ by contradiction. If on the contrary $I_{\rm o}(\mathbf{u}) < I_{\rm o}(\mathbf{v})$, then $I_{\rm o}(\mathbf{u}) + 1 \leqslant I_{\rm o}(\mathbf{v}) \leqslant \dim(\mathbf{v})$ and

$$\sum_{i=1}^{I_{\rm o}(\mathbf{u})+1} u_i \leqslant \sum_{i=1}^{I_{\rm o}(\mathbf{u})+1} v_i \leqslant \alpha(I_{\rm o}(\mathbf{u}) + 1).$$

This contradicts with the definition of $I_{\rm o}(\mathbf{u})$. Therefore, $I_{\rm o}(\mathbf{u}) \geqslant I_{\rm o}(\mathbf{v})$.

We proceed to showing $H_{\rm o}(\mathbf{u}) \preccurlyeq H_{\rm o}(\mathbf{v})$. By the definition of $H_{\rm o}$, we have $H_{\rm o}(\mathbf{u}) = (u_1, \cdots, u_{I_{\rm o}(\mathbf{u})})$ and $H_{\rm o}(\mathbf{v}) = (v_1, \cdots, v_{I_{\rm o}(\mathbf{v})})$. Since we assume $\mathbf{u} \preccurlyeq \mathbf{v}$, we have $u_i \leqslant v_i$ for all $i = 1, \cdots, I_{\rm o}(\mathbf{v})$. This shows that $H_{\rm o}(\mathbf{u}) \preccurlyeq H_{\rm o}(\mathbf{v})$.

□

Lemma F.6. *Let $p(x)$ and $q(x)$ be two density functions with respect to some baseline measure μ and assume that $p(\cdot)$ and $q(\cdot)$ have the same support. Let $L(x) = \frac{q(x)}{p(x)}$ be the likelihood ratio. For $\delta \in [0,1]$, let Z_δ be a random variable with the density function $\delta q + (1-\delta)p$ and $L_\delta = L(Z_\delta)$. Then, for $0 \leqslant \delta_1 < \delta_2 \leqslant 1$, we have*

$$L_{\delta_1} \leqslant_{st} L_{\delta_2}.$$

Proof of Lemma F.6. Let g be a bounded increasing function. Then,

$$\begin{aligned}
&\mathbb{E}g(L_{\delta_2}) - \mathbb{E}g(L_{\delta_1}) \\
=&\mathbb{E}_{Z\sim\delta_2 q+(1-\delta_2)p}\, g\big(L(Z)\big) - \mathbb{E}_{Z\sim\delta_1 q+(1-\delta_1)p}\, g\big(L(Z)\big) \\
=&\delta_2\mathbb{E}_{Z\sim q}\, g\big(L(Z)\big) + (1-\delta_2)\mathbb{E}_{Z\sim p}\, g\big(L(Z)\big) \\
&- \left\{\delta_1\mathbb{E}_{Z\sim q}\, g\big(L(Z)\big) + (1-\delta_1)\mathbb{E}_{Z\sim p}\, g\big(L(Z)\big)\right\} \\
=&(\delta_2-\delta_1)\left\{\mathbb{E}_{Z\sim q}\, g\big(L(Z)\big) - \mathbb{E}_{Z\sim p}\, g\big(L(Z)\big)\right\}.
\end{aligned}$$

Note that $L(Z) = q(Z)/p(Z)$ and $\mathbb{E}_{Z\sim q}\, g\big(L(Z)\big) = \mathbb{E}_{Z\sim p}\left\{L(Z)g\big(L(Z)\big)\right\}$. Thus, the above display can be further written as

$$\mathbb{E}g(L_{\delta_2}) - \mathbb{E}g(L_{\delta_1}) = (\delta_2-\delta_1)\mathbb{E}_{Z\sim p}\left\{\big(L(Z)-1\big)g\big(L(Z)\big)\right\}.$$

For notational simplicity, let $Y = L(Z)$ with $Z \sim p$. Then, $\mathbb{E}(Y) = 1$ and the above display implies

$$\mathbb{E}g(L_{\delta_2}) - \mathbb{E}g(L_{\delta_1}) = (\delta_2-\delta_1)\mathbb{E}\left\{(Y-1)g(Y)\right\} = (\delta_2-\delta_1)\mathbb{E}\left\{(Y-1)(g(Y)-g(1))\right\} \geqslant 0.$$

The last inequality in the above display is due to the fact that $(Y-1)(g(Y)-g(1)) \geqslant 0$

for all increasing function g. We remark that it is also a special case of Harris inequality (Harris, 1960). $\square$

Lemma F.7. *Assume model $\mathcal{M}_s$ holds. Let $V_{k,t} = \mathbb{P}\left(\tau_k < t | X_{k,1}, \cdots, X_{k,t}\right)$. Then,*

$$V_{k,0} = 0 \text{ and } V_{k,t+1} = \frac{q(X_{k,t+1})/p(X_{k,t+1})}{(1-\theta)(1-V_{k,t})/(\theta + (1-\theta)V_{k,t}) + q(X_{k,t+1})/p(X_{k,t+1})}. \tag{F.2}$$

Moreover, $\{V_{k,t}\}_{t=0,1,\cdots}$ are independent and identically distributed processes for different k.

Proof of Lemma F.7. First, it is easy to see that $\{V_{k,s}\}_{s\geqslant 0}$ are independent and identically distributed processes for different k. For the rest of the proof, it is sufficient to prove the lemma for $k=1$. For the ease of exposition, we use the notation $X_{k,s:t}$ to denote $(X_{k,r})_{s\leqslant r\leqslant t}$. First, $\mathbb{P}\left(\tau_1 < 0 | X_{1,1:0}\right) = \mathbb{P}(\tau_1 < 0) = 0 = V_0$. Thus, it is sufficient to verify the update rule for $V_{1,t}$. A direct calculation gives

$$\begin{aligned}
&\mathbb{P}(\tau_1 \leqslant t-1 | X_{1,1:t}) \\
=&\frac{\sum_{s=0}^{t-1}\mathbb{P}(\tau_1 = s)\prod_{r=1}^{s} p(X_{1,r})\prod_{r=s+1}^{t} q(X_{1,r})}{\sum_{s=0}^{t-1} P(\tau_1 = s)\prod_{r=1}^{s} p(X_{1,r})\prod_{r=s+1}^{t} q(X_{1,r}) + P(\tau_1 \geqslant t)\prod_{r=1}^{t} p(X_{1,r})} \\
=&\frac{\sum_{s=0}^{t-1}\theta(1-\theta)^s L_{1,(s+1):t}}{\sum_{s=0}^{t-1}\theta(1-\theta)^s L_{1,(s+1):t} + (1-\theta)^t} \\
=&\frac{Q_{1,t}}{Q_{1,t} + (1-\theta)^t}
\end{aligned}$$

where we write $L_{k,(s+1):t} := \prod_{r=s+1}^{t}\frac{q(X_{k,r})}{p(X_{k,r})}$, the likelihood ratio between $p(\cdot)$ and $q(\cdot)$ based on the data $X_{1,(s+1):t}$, and $Q_{1,t} = \sum_{s=0}^{t-1}\theta(1-\theta)^s L_{1,(s+1):t}$. Then,

$$Q_{1,t} = \frac{(1-\theta)^t\mathbb{P}(\tau_1 \leqslant t-1 | X_{1,1:t})}{1 - \mathbb{P}(\tau_1 \leqslant t-1 | X_{1,1:t})}.$$

Note that

$$Q_{1,t+1} = \sum_{s=0}^{t}\theta(1-\theta)^s L_{1,(s+1):t+1} = q(X_{1,t+1})/p(X_{1,t+1})\left\{\theta(1-\theta)^t + Q_{1,t}\right\}.$$

Thus,

$$
\begin{aligned}
&\mathbb{P}(\tau_1 \leqslant t | X_{1,1:t+1}) \\
=& \frac{Q_{1,t+1}}{Q_{1,t+1} + (1-\theta)^{t+1}} \\
=& \frac{q(X_{1,t+1})/p(X_{1,t+1}) \left\{\theta(1-\theta)^t + Q_{1,t}\right\}}{q(X_{1,t+1})/p(X_{1,t+1}) \left\{\theta(1-\theta)^t + Q_{1,t}\right\} + (1-\theta)^{t+1}} \\
=& \frac{q(X_{1,t+1})/p(X_{1,t+1})}{q(X_{1,t+1})/p(X_{1,t+1}) + (1-\theta)/\left\{\theta + (1-\theta)^{-t} Q_{1,t}\right\}} \\
=& \frac{q(X_{1,t+1})/p(X_{1,t+1})}{q(X_{1,t+1})/p(X_{1,t+1}) + (1-\theta)/\left\{\theta + \frac{\mathbb{P}(\tau_1 \leqslant t-1 | X_{1,1:t})}{1-\mathbb{P}(\tau_1 \leqslant t-1 | X_{1,1:t})}\right\}}.
\end{aligned}
$$

We complete the proof by simplifying the above result. $\square$

Lemma F.8. *Assume model $\mathcal{M}_s$ holds. Let $\delta_{k,t} = \mathbb{P}(\tau_k \leqslant t | X_{k,1}, \cdots, X_{k,t})$, then*

$$\delta_{k,t} = \theta + (1-\theta) V_{k,t},$$

where $V_{k,t}$ is defined in (F.2).

Proof of Lemma F.8. By symmetry, it is sufficient to prove the lemma for $k = 1$. Recall $L_{k,(s+1):t} = \prod_{r=s+1}^{t} \frac{q(X_{k,r})}{p(X_{k,r})}$ and $Q_{k,t} = \sum_{s=0}^{t-1} \theta(1-\theta)^s L_{1,(s+1):t}$.

A direct calculation using Bayes formula gives

$$
\begin{aligned}
\delta_{k,t} =& \frac{\sum_{s=0}^{t-1} \mathbb{P}(\tau_1 = s) \prod_{r=1}^{s} p(X_{1,r}) \prod_{r=s+1}^{t} q(X_{1,r}) + P(\tau_1 = t) \prod_{r=1}^{t} p(X_{1,r})}{\sum_{s=0}^{t-1} P(\tau_1 = s) \prod_{r=1}^{s} p(X_{1,r}) \prod_{r=s+1}^{t} q(X_{1,r}) + P(\tau_1 \geqslant t) \prod_{r=1}^{t} p(X_{1,r})} \\
=& \frac{\sum_{s=0}^{t-1} \theta(1-\theta)^s L_{1,(s+1):t} + \theta(1-\theta)^t}{\sum_{s=0}^{t-1} \theta(1-\theta)^s L_{1,(s+1):t} + (1-\theta)^t} \\
=& \frac{Q_{1,t} + \theta(1-\theta)^t}{Q_{1,t} + (1-\theta)^t} \\
=& V_{1,t} + \theta(1 - V_{1,t}) \\
=& \theta + (1-\theta) V_{1,t}.
\end{aligned}
$$

$\square$

Lemma F.9. *Under model $\mathcal{M}_s$, the process $\{V_{1,t}\}_{t \geqslant 0}$ defined in* (F.2) *is a homogeneous Markov chain. In addition, its transition kernel is stochastically monotone. We will later refer to this transition kernel as $K(\cdot, \cdot)$.*

Proof of Lemma F.9. We first study the conditional distribution of $X_{1,t+1}$ given $V_{1,0}, \cdots, V_{1,t}$. According to the change point model $\mathcal{M}_\mathrm{s}$, we know that $X_{1,t+1}$ is conditionally independent of $V_{1,0}, \cdots, V_{1,t}$ given the event $\{\tau_1 \leqslant t\}$. That is, given $V_{1,0}, \cdots, V_{1,t}$, the conditional density function of $X_{1,t+1}$ is $\delta_{1,t} q(x) + (1 - \delta_{1,t}) p(x)$, which depends on $X_{1,1}, \cdots, X_{1,t}$ only through $V_{1,t}$.

Let the function $L(x) := q(x)/p(x)$ and let $L_{k,t+1} := q(X_{k,t+1})/p(X_{k,t+1})$. Then, $L_{1,t+1} = L(X_{1,t+1})$, whose conditional distribution given $V_{1,0}, \cdots, V_{1,t}$ only depends on $V_{1,t}$. According to the iteration (F.2), this implies that the process $\{V_{1,t}\}_{t \geqslant 0}$ is a Markov process. Note that $\delta_{1,t}$ and the iteration (F.2) depend on t only through $V_{1,t}$. Thus, this Markov chain is a homogeneous Markov chain. We now show that its transition kernel is stochastically monotone.

Let $\delta(x) = \theta + (1 - \theta)x$. For $x \in (0, 1)$, we consider the following steps of generating a random variable $V(x)$.

1. Generate $Z(x)$ with the density $\delta(x) q(\cdot) + (1 - \delta(x)) p(\cdot)$.

2. Let
$$V(x) = \frac{L(Z(x))}{L(Z(x)) + (1 - \theta)(1 - x)/(\theta + (1 - \theta)x)}.$$

From the iteration (F.2) and $X_{1,t+1} | V_t = x \sim (1 - \delta(x)) q(\cdot) + \delta(x) p(\cdot)$, we can see that $V(x)$ has the same distribution as that of $V_{1,t+1}$ given $V_{1,t} = x$. In other words, $V(x)$ has the density function $K(x, \cdot)$.

Now we show that $K(x, \cdot) \leqslant_{st} K(x', \cdot)$ for any $0 < x \leqslant x' < 1$ by coupling. Specifically, since $\delta(x)$ is increasing in x, $\delta(x) \leqslant \delta(x')$. Then, by Lemma F.6, we know $L(Z(x)) \leqslant_{st} L(Z(x'))$. According to the Strassen Theorem for random variables (Fact 1), there exists a coupling $(\widehat{L}, \widehat{L}')$, such that $\widehat{L} \overset{d}{=} L(Z(x))$, $\widehat{L}' \overset{d}{=} L(Z(x'))$ and $\widehat{L} \leqslant \widehat{L}'$ a.s. Then, let $\widehat{V} = \frac{\widehat{L}}{\widehat{L} + (1-\theta)(1-x)/(\theta + (1-\theta)x)} \overset{d}{=} V(x)$ and $\widehat{V}' = \frac{\widehat{L}'}{\widehat{L}' + (1-\theta)(1-x')/(\theta + (1-\theta)x')} \overset{d}{=} V(x')$.

Because $\widehat{L} \leqslant \widehat{L}'$ and $x \leqslant x'$,

$$\begin{aligned}
\widehat{V} &= \frac{\widehat{L}}{\widehat{L} + (1 - \theta)(1 - x)/(\theta + (1 - \theta)x)} \\
&\leqslant \frac{\widehat{L}'}{\widehat{L}' + (1 - \theta)(1 - x)/(\theta + (1 - \theta)x)} \\
&\leqslant \frac{\widehat{L}'}{\widehat{L}' + (1 - \theta)(1 - x')/(\theta + (1 - \theta)x')} \\
&= \widehat{V}' \quad a.s.
\end{aligned}$$

That is, $\widehat{V} \leqslant \widehat{V}'$ a.s., and $(\widehat{V}, \widehat{V}')$ is a coupling of $(V(x), V(x'))$. Thus, $V(x) \leqslant_{st} V(x')$ and so is $K(x, \cdot) \leqslant_{st} K(x', \cdot)$. $\square$

Lemma F.10. *For any $t \geqslant 1$ and $\mathbb{T}^{\boldsymbol{A}}$, $[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t+1}]$ is conditionally independent of $\mathcal{F}^{\boldsymbol{A}}_t$ given $[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t}]$. Moreover, the conditional density of $\left[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t+1}\right]$ at $\mathbf{v}$ given $\left[W^{\boldsymbol{A}}_{S^{\boldsymbol{A}}_{t+1},t}\right] = \mathbf{u} \in \mathcal{S}_{\mathrm{o}}$ is*

$$\mathbb{K}_a(\mathbf{u}, \mathbf{v}) := \begin{cases} \sum_{\pi \in \mathbf{P}_{\dim(\mathbf{u})}} \prod_{l=1}^{\dim(\mathbf{u})} K(u_l, v_{\pi(l)}) & \textit{if } \dim(\mathbf{u}) = \dim(\mathbf{v}) \geqslant 1, \\ 1 & \textit{if } \dim(\mathbf{u}) = \dim(\mathbf{v}) = 0, \\ 0 & \textit{otherwise,} \end{cases}$$

where $\mathbf{P}_m$ denotes the set of all permutations over $\{1, \cdots, m\}$.

Proof of Lemma F.10. First, if $\dim(\mathbf{u}) = 0$, then $\mathbf{u} = \boldsymbol{\varnothing}$, and $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t}\right] = \mathbf{u}$ means that $S^{\mathbf{A}}_{t+1} = \varnothing$. Thus, the conditional distribution of $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t+1}\right]$ given $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t}\right] = \mathbf{u}$ is a point mass at $\boldsymbol{\varnothing}$, and $\mathbb{K}_{\mathrm{a}}(\boldsymbol{\varnothing}, \boldsymbol{\varnothing}) = 1$. In the rest of the proof, we focus on the case that $\mathbf{u} \neq \boldsymbol{\varnothing}$.

We start with deriving the conditional density of $W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t+1}$ at $\mathbf{v} \in \mathcal{S}_{\mathrm{u}}$ given $X_{S^{\mathbf{A}}_1,1} = x_1, S^{\mathbf{A}}_1 = s_1, \cdots, X_{S^{\mathbf{A}}_t,t} = x_t, S^{\mathbf{A}}_t = s_t$, $S^{\mathbf{A}}_{t+1} = s_{t+1}$ and $W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t} = \mathbf{u}$ for some $x_1, \cdots, x_t$ and $s_1, \cdots, s_{t+1}$, and $\mathbf{u} \in \mathcal{S}_{\mathrm{u}}$. Clearly, the conditional density is 0 when $\dim(\mathbf{u}) \neq \dim(\mathbf{v})$, and is arbitrary when $\dim(\mathbf{u}) \neq |s_{t+1}|$ (the density of the random variable being conditional on is zero). Thus, we will focus on the case where $\dim(\mathbf{u}) = \dim(\mathbf{v}) = |s_{t+1}| = m$ for some $m \in \{1, \cdots, K\}$, and we will write $\mathbf{u} = (u_1, \cdots, u_m)$ and $\mathbf{v} = (v_1, \cdots, v_m)$.

Note that given $S^{\mathbf{A}}_{t+1} = s_{t+1}, W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t} = \mathbf{u}$, $W^{\mathbf{A}}_{k,t+1}$'s are independent for different $k \in s_{t+1}$. Moreover, given $S^{\mathbf{A}}_{t+1} = s_{t+1}, W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t} = \mathbf{u}$, $W^{\mathbf{A}}_{k,t+1}$ is the same as $V_{k,t+1}$ (defined in (F.2)) for $k \in s_{t+1}$, and is independent of $X_{S^{\mathbf{A}},1} = x_1, S^{\mathbf{A}} = s_1, \cdots, X_{S^{\mathbf{A}},t} = x_t$ and $S^{\mathbf{A}}_t = s_t$. Thus, $W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t}$ is conditionally independent of $\mathcal{F}^{\mathbf{A}}_t$ given $S^{\mathbf{A}}_{t+1} = s_{t+1}, W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t} = \mathbf{u}$, and its conditional density (by Lemma F.9) is

$$\prod_{l=1}^{m} K(u_l, v_l),$$

Because $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t}\right]$ is the order statistic of $W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t}$, we further obtain its conditional density at $\mathbf{v} \in \mathcal{S}_{\mathrm{o}}$ given $S^{\mathbf{A}}_{t+1} = s_{t+1}, W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t} = \mathbf{u}$,

$$\sum_{\pi \in \mathbf{P}_m} \prod_{l=1}^{m} K(u_l, v_{\pi(l)}) = \sum_{\pi \in \mathbf{P}_m} \prod_{l=1}^{m} K([\mathbf{u}]_l, v_{\pi(l)}) = \mathbb{K}_{\mathrm{a}}([\mathbf{u}], \mathbf{v}),$$

for $\mathbf{v} \in \mathcal{S}_{\mathrm{o}}$ with $\dim(\mathbf{v}) = m$. Observe that the above function is independent of s_{t+1} for $|s_{t+1}| = m$ and depend on $\mathbf{u}$ only through its order statistic $[\mathbf{u}]$. Thus, we further conclude that $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t+1}\right]$ is conditionally independent of $\mathcal{F}^{\mathbf{A}}_t$ given $\left[W^{\mathbf{A}}_{S^{\mathbf{A}}_{t+1},t}\right] = \mathbf{u} \in \mathcal{S}_{\mathrm{o}}$ satisfying $\dim(\mathbf{u}) = m$, and its conditional density is $\mathbb{K}_{\mathrm{a}}(\mathbf{u}, \cdot)$. $\square$

Lemma F.11. *For each $\mathbf{u} \in \mathcal{S}_{\mathrm{o}}$ with $\dim(\mathbf{u}) = m \geqslant 1$, generate an $\mathcal{S}_{\mathrm{o}}$-valued random* *variable V as follows,*

1. *For each $k \in \{1, \cdots, m\}$, generate $Z_k \sim K(u_k, \cdot)$ independently for different k.*

2. *Let $V = [(Z_1, ..., Z_m)]$.*

In addition, if $m = 0$, we let $V = \boldsymbol{\varnothing}$. Then, $V \sim \mathbb{K}_a(\mathbf{u}, \cdot)$.

Proof of Lemma F.11. The lemma is obviously true when $m = 0$. When $m \geqslant 1$, let $\mathbf{z} = (z_1, \cdots, z_m)$. By step 1, the joint density for $(Z_1, \cdots, Z_m)$ at $\mathbf{z}$ is

$$\prod_{i=1}^{m} K(u_i, z_i).$$

By step 2, V is the order statistic of $(Z_1, \cdots, Z_m)$. Thus, its density is

$$\sum_{\pi \in \mathbf{P}_m} \prod_{i=1}^{m} K(u_i, z_{\pi(i)}) = \mathbb{K}_{\mathrm{a}}(\mathbf{u}, \mathbf{z}).$$

□

Lemma F.12. *For $\mathbf{u}, \mathbf{u}' \in \mathcal{S}_{\mathrm{o}}$ with $\mathbf{u} \preccurlyeq \mathbf{u}'$, we have $\mathbb{K}_a(\mathbf{u}, \cdot) \preccurlyeq_{st} \mathbb{K}_a(\mathbf{u}', \cdot)$.*

Proof of Lemma F.12. The lemma is obvious if $\mathbf{u}' = \boldsymbol{\varnothing}$. In what follows, we assume $\dim(\mathbf{u}') = m' \geqslant 1$ and $\dim(\mathbf{u}) = m$. Then, $\mathbf{u} \preccurlyeq \mathbf{u}'$ means $m \geqslant m' \geqslant 1$ and $u_l \leqslant u'_l$ for $1 \leqslant l \leqslant m'$. Let $(Z_1, Z'_1), \cdots (Z_m, Z'_m)$ be independent random vectors such that $Z_l \sim K(u_l, \cdot)$, $Z'_l \sim K(u'_l, \cdot)$ and $Z_l \leqslant Z'_l$ a.s. Such random vectors exists because of Strassen Theorem and Lemma F.9 that the kernel $K(\cdot, \cdot)$ is stochastically monotone. In addition, for $m < l \leqslant m'$, let $Z'_l \sim K(u'_l, \cdot)$ be independent random variables.

Let $Z = (Z_1, \cdots, Z_m) \sim \mathbb{K}_{\mathrm{a}}(\mathbf{u}, \cdot)$, $Z' = (Z'_1, \cdots, Z'_{m'})$, $V = [Z]$ and $V' = [Z']$. Then, $V \preccurlyeq V'$ a.s. On the other hand, by Lemma F.11, we have

$$V \sim \mathbb{K}_{\mathrm{a}}(\mathbf{u}, \cdot) \text{ and } V' \sim \mathbb{K}_{\mathrm{a}}(\mathbf{u}', \cdot),$$

and $V \preccurlyeq V'$ a.s. By Fact 2, the existence of such a coupling implies $\mathbb{K}_{\mathrm{a}}(\mathbf{u}, \cdot) \preccurlyeq_{st} \mathbb{K}_{\mathrm{a}}(\mathbf{u}', \cdot)$. □

Lemma F.13. *For any $t, s \geqslant 0$ and $\mathbb{T}^{\boldsymbol{A}}$, $\big[W^{\boldsymbol{A}\mathbf{P}_t}_{S^{\boldsymbol{A}\mathbf{P}_t}_{t+s+1}, t+s+1}\big]$ is conditionally independent of $\mathcal{F}^{\boldsymbol{A}\mathbf{P}_t}_{t+s}$ given $\big[W^{\boldsymbol{A}\mathbf{P}_t}_{S^{\boldsymbol{A}\mathbf{P}_t}_{t+s}, t+s}\big]$. Moreover, the conditional density of $\big[W^{\boldsymbol{A}\mathbf{P}_t}_{S^{\boldsymbol{A}\mathbf{P}_t}_{t+s+1}, t+s+1}\big]$ at $\mathbf{v}$ given*

$\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s},t+s}\big] = \mathbf{u} \in \mathcal{S}_{\mathrm{o}}$ *is*

$$\mathbb{K}_o(\mathbf{u}, \mathbf{v}) := \mathbb{K}_a(H_{\mathrm{o}}(\mathbf{u}), \mathbf{v}) = \begin{cases} \sum_{\pi \in \mathbf{P}_{I_{\mathrm{o}}(\mathbf{u})}} \prod_{l=1}^{I_{\mathrm{o}}(\mathbf{u})} K(H_{\mathrm{o}}(\mathbf{u})_l, v_{\pi(l)}) & \text{if } \dim(\mathbf{v}) = I_{\mathrm{o}}(\mathbf{u}) \geqslant 1 \\ 1 & \text{if } \dim(\mathbf{v}) = I_{\mathrm{o}}(\mathbf{u}) = 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{P}_m$ *denotes the set of all permutations over* $\{1, \cdots, m\}$.

Proof of Lemma F.13. Apply Lemma F.10 by replacing $\mathbb{T}^{\mathbf{A}}$ by $\mathbb{T}^{\mathbf{AP}_t}$ and t by $t+s$, we obtain that $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s+1},t+s+1}\big]$ is conditionally independent of $\mathcal{F}^{\mathbf{AP}_t}_{t+s}$ given $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s+1},t+s}\big]$. On the other hand, according to the one-step update rule in Algorithm 1 and Lemma F.2, we can see that $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s+1},t+s}\big] = H_{\mathrm{o}}\Big(\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s},t+s}\big]\Big)$. Therefore, we further obtain that $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s+1},t+s+1}\big]$ is conditionally independent of $\mathcal{F}^{\mathbf{AP}_t}_{t+s}$ given $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s},t+s}\big]$.

We proceed to derive its conditional density at $\mathbf{v}$ given $\big[W^{\mathbf{AP}_t}_{S^{\mathbf{AP}_t}_{t+s},t+s}\big] = \mathbf{u}$. We first notice that $\dim(\mathbf{v}) = |S^{\mathbf{AP}_t}_{t+s+1}| = I_{\mathrm{o}}(\mathbf{u})$ (by Lemma F.2). Thus, the conditional density is zero when $\dim(\mathbf{v}) \neq I_{\mathrm{o}}(\mathbf{u})$. For $\dim(\mathbf{v}) = I_{\mathrm{o}}(\mathbf{u})$, by Lemma F.10 and the above analysis, the conditional density is

$$\mathbb{K}_{\mathrm{a}}(H_{\mathrm{o}}(\mathbf{u}), \mathbf{v}) = \sum_{\pi \in \mathbf{P}_{I_{\mathrm{o}}(\mathbf{u})}} \prod_{l=1}^{I_{\mathrm{o}}(\mathbf{u})} K(H_{\mathrm{o}}(\mathbf{u})_l, v_{\pi(l)}) = \mathbb{K}_{\mathrm{o}}(\mathbf{u}, \mathbf{v}).$$

This completes the proof of the lemma. $\square$

Lemma F.14. *For each* $\mathbf{u} \in \mathcal{S}_{\mathrm{o}}$ *and* $m = I_{\mathrm{o}}(\mathbf{u})$*, generate an* $\mathcal{S}_{\mathrm{o}}$*-valued random variable* V *as follows,*

1. *For each* $k \in \{1, \cdots, m\}$*, generate* $Z_k \sim K(H_{\mathrm{o}}(\mathbf{u})_k, \cdot)$ *independently for different* k.
2. *Let* $V = [(Z_1, ..., Z_m)]$.

In addition, if $m = 0$*, we let* $V = \varnothing$*. Then,* $V \sim \mathbb{K}_o(\mathbf{u}, \cdot)$.

Proof of Lemma F.14. The lemma is a direct application of Lemma F.11 and $\mathbb{K}_{\mathrm{o}}(\mathbf{u}, \mathbf{v}) = \mathbb{K}_{\mathrm{a}}(H_{\mathrm{o}}(\mathbf{u}), v)$. $\square$

G Proof of Lemma 1 and Propositions 1 - 3

Lemma 1. *Under model $\mathcal{M}_s$, $W_{k,0} = 0$ for $1 \leqslant k \leqslant K$ and $W_{k,t}$ can be computed using the following update rule for $1 \leqslant k \leqslant K$,*

$$W_{k,t+1} = \begin{cases} \frac{q(X_{k,t+1})/p(X_{k,t+1})}{(1-\theta)(1-W_{k,t})/(\theta+(1-\theta)W_{k,t})+q(X_{k,t+1})/p(X_{k,t+1})} & \text{for } 1 \leqslant t \leqslant T_k - 1, \\ W_{k,T_k} & \text{for } t \geqslant T_k. \end{cases}$$

Proof of Lemma 1. For each $k \in S_{t+1}$, according to the independence assumption for model $\mathcal{M}_\mathrm{s}$,

$$W_{k,t+1} = \mathbb{P}(\tau_k < t+1 | \mathcal{F}_{t+1}) = \mathbb{P}(\tau_k < t+1 | X_{k,1:t+1}).$$

On the other hand, according to Lemma F.7, we have

$$\mathbb{P}(\tau_k < t+1 | X_{k,1:t+1}) = \frac{q(X_{k,t+1})/p(X_{k,t+1})}{(1-\theta)(1-W_{k,t})/(\theta + (1-\theta)W_{k,t}) + q(X_{k,t+1})/p(X_{k,t+1})}.$$

Thus, for $k \in S_{t+1}$,

$$W_{k,t+1} = \frac{q(X_{k,t+1})/p(X_{k,t+1})}{(1-\theta)(1-W_{k,t})/(\theta + (1-\theta)W_{k,t}) + q(X_{k,t+1})/p(X_{k,t+1})}. \tag{G.1}$$

Note that $k \in S_{t+1}$ is equivalent to $T_k \geqslant t+1$. Thus, (G.1) holds for $1 \leqslant t \leqslant T_k - 1$. Moreover, for $t \geqslant T_k$,

$$W_{k,t+1} = \mathbb{P}(\tau_k < t+1 | \mathcal{F}_{t+1}) = \mathbb{P}(\tau_k < t | X_{1,1:T_k}, T_k) = W_{k,T_k}.$$

This completes our proof. □

We proceed to the proofs of propositions.

Proposition 1. *Suppose that we obtain the index set S_{t+1} using Algorithm 1, given the index set S_t and information filtration $\mathcal{F}_t$ at time t. Then the LFNR at time $t+1$ satisfies*

$$\mathbb{E}\left(\frac{\sum_{k \in S_{t+1}} \mathbb{1}(\tau_k < t)}{|S_{t+1}| \vee 1} | \mathcal{F}_t\right) \leqslant \alpha.$$

Proof of Proposition 1. First, it is easy to see that S_{t+1} obtained from Algorithm 1 is $\mathcal{F}_t$ measurable. Thus,

$$\mathbb{E}\left(\frac{\sum_{k \in S_{t+1}} \mathbb{1}(\tau_k < t)}{|S_{t+1}| \vee 1} | \mathcal{F}_t\right) = \frac{\sum_{k \in S_{t+1}} W_{k,t}}{|S_{t+1}| \vee 1}.$$

On the other hand, according to the second and third steps of the algorithm,

$$\frac{\sum_{k\in S_{t+1}} W_{k,t}}{|S_{t+1}| \vee 1} = R_n \leqslant \alpha.$$

Therefore, $\mathbb{E}\left(\frac{\sum_{k\in S_{t+1}} \mathbb{1}(\tau_k<t)}{|S_{t+1}|\vee 1}\middle|\mathcal{F}_t\right) \leqslant \alpha$.

$\square$

Proposition 2. *Let $\mathbb{T}^*$ be defined in Algorithm 2. Then, $\mathbb{T}^* \in \mathcal{T}_\alpha$.*

Proof of Proposition 2. This proposition is proved by combining the results of Proposition 1 and Lemma F.3. $\square$

Proposition 3. *Given LFNR level α and information filtration $\mathcal{F}_t$, the index set S_{t+1} given by Algorithm 1 is locally optimal at time $t+1$.*

Proof of Proposition 3. Let S_{t+1} be the index set obtained by Algorithm 1. By Lemma F.2, $|S_{t+1}| = I_{\mathrm{o}}([W_{S_t,t}])$ and $[W_{S_{t+1},t}] = H_{\mathrm{o}}([W_{S_t,t}])$. There are two cases: 1) $|S_{t+1}| = 0$, and 2) $|S_{t+1}| = n \geqslant 1$. For the first case, $[W_{S_{t+1},t}] = \varnothing$. Note that $\mathbb{E}\left(\frac{\sum_{k\in S} \mathbb{1}(\tau_k<t)}{|S|\vee 1}\middle|\mathcal{F}_t\right) = \frac{\sum_{k\in S} W_{k,t}}{|S|\vee 1}$. By the 'Moreover' part of Lemma F.4, we can see that the only set S satisfying $\mathbb{E}\left(\frac{\sum_{k\in S} \mathbb{1}(\tau_k<t)}{|S|\vee 1}\middle|\mathcal{F}_t\right) \leqslant \alpha$ is $S = \varnothing$. That is $|S| = 0$. Thus, $|S_{t+1}| \geqslant |S|$.

For the second case where $|S_{t+1}| = n \geqslant 1$ and any set $|S|$ satisfying $\mathbb{E}\left(\frac{\sum_{k\in S} \mathbb{1}(\tau_k<t)}{|S|\vee 1}\middle|\mathcal{F}_t\right) \leqslant \alpha$, we use Lemma F.4 again and obtain that $[W_{S_{t+1},t}] = H_{\mathrm{o}}([W_{S_t,t}]) \leqslant [W_{S,t}]$. This implies $|S_{t+1}| = \dim([W_{S_{t+1},t}]) \geqslant \dim([W_{S,t}]) = |S|$. $\square$

H Proof of Theorem 2 and Theorem 3

H.1 Proof of Theorem 2

Theorem 2. *Assume that model $\mathcal{M}_s$ holds and Assumption A1 is satisfied. To emphasize the dependence on K, we denote the proposed procedure by $\mathbb{T}_K^*$, the corresponding information filtration at time t by $\mathcal{F}_{K,t}^*$, and the index set at time t by $S_{K,t}^*$. Then, the following results hold for each $t \geqslant 1$.*

1. *$\lim_{K\to\infty} \widehat{\lambda}_{K,t} = \lambda_t$ a.s., where $\widehat{\lambda}_{K,t} = \max\left\{W_{k,t} : k \in S_{K,t+1}^*\right\}$ is the threshold used by $\mathbb{T}_K^*$.*

2. $\lim_{K\to\infty} \mathrm{LFNR}_{t+1}(\mathbb{T}_K^*) = \mathbb{E}\left(V_t \middle| V_s \leqslant \lambda_s, 0 \leqslant s \leqslant t\right)$, *a.s. Moreover,*

$$\mathbb{E}\left(V_t \middle| V_s \leqslant \lambda_s, 0 \leqslant s \leqslant t\right) = \begin{cases} 1-(1-\theta)^t, & t < \frac{\log(1-\alpha)}{\log(1-\theta)}, \\ \alpha, & t \geqslant \frac{\log(1-\alpha)}{\log(1-\theta)}. \end{cases}$$

3. $\lim_{K\to\infty} K^{-1}|S^*_{K,t+1}| = \mathbb{P}\left(V_1 \leqslant \lambda_1, \cdots, V_t \leqslant \lambda_t\right)$ *a.s.*

We start with a lemma that is useful for the proof of Theorem 2. Its proof is provided in Section H.3.

Lemma H.1. *Under model $\mathcal{M}_s$ and Assumption A1, we have the following results.*

1. *For each $t \geqslant 1$, $(V_1, \cdots, V_t)$ has a continuous and strictly positive joint density function over $(0,1)^t$ (with respect to the Lebesgue measure).*

2. *For any $(v_1, \cdots, v_t) \in (0,1)^t$, $\mathbb{P}(V_1 \leqslant v_1, \cdots, V_t \leqslant v_t) > 0$.*

3. *For any $(v_1, \cdots, v_t) \in (0,1)^t$, the conditional distribution of V_{t+1} given $V_1 \leqslant v_1, \cdots, V_t \leqslant v_t$ has a continuous and positive density function over $(0,1)$.*

Proof of Theorem 2. For a sufficiently large t_0 ($t_0 > t$), let $\mathbb{P}^*$ denote the probability measure for $(V_1, \cdots, V_{t_0})$, and let $\mathbb{Q}$ be an arbitrary probability measure for a t_0-dimensional random vector. We define several mappings iteratively as follows. We initialize the mapping $\Lambda_0(\mathbb{Q}) = 1$ for every $\mathbb{Q}$. Then, for $t \geqslant 1$, define

$$\begin{aligned} D_t(\lambda, \mathbb{Q}) &= \mathbb{Q}\left(V_t \leqslant \lambda, \mathbf{V}_{t-1} \leqslant \mathbf{\Lambda}_{t-1}(\mathbb{Q})\right), \\ N_t(\lambda, \mathbb{Q}) &= \mathbb{E}_{\mathbb{Q}}\left[V_t \mathbb{1}\left\{V_t \leqslant \lambda, \mathbf{V}_{t-1} \leqslant \mathbf{\Lambda}_{t-1}(\mathbb{Q})\right\}\right], \\ G_t(\lambda, \mathbb{Q}) &= \frac{N_t(\lambda, \mathbb{Q})}{D_t(\lambda, \mathbb{Q})} = \mathbb{E}_{\mathbb{Q}}\left[V_t | V_t \leqslant \lambda, \mathbf{V}_{t-1} \leqslant \mathbf{\Lambda}_{t-1}(\mathbb{Q})\right], \end{aligned}$$

and

$$\Lambda_t(\mathbb{Q}) = \sup\left\{\lambda : G_t(\lambda, \mathbb{Q}) \leqslant \alpha \text{ and } \lambda \in [0,1]\right\}.$$

In the above equations, we use notation $\mathbf{V}_t = (V_1, \cdots, V_t)$ and $\mathbf{\Lambda}_t(\mathbb{Q}) = (\Lambda_1(\mathbb{Q}), \cdots, \Lambda_t(\mathbb{Q}))$. In addition, $\{\mathbf{V}_t \leqslant \mathbf{\Lambda}_t(\mathbb{Q})\}$ denotes the event $\{V_1 \leqslant \Lambda_1(\mathbb{Q}), \cdots, V_t \leqslant \Lambda_t(\mathbb{Q})\}$.

The next lemma, whose proof is given in Section H.3, provides results about the above mappings. For two probability measures $\mathbb{Q}$ and $\mathbb{Q}'$ for a t_0-dimensional random vector $\mathbf{V}_t$, their sup-norm is defined as $\|\mathbb{Q} - \mathbb{Q}'\|_\infty = \sup_{\mathbf{v}\in\mathbb{R}^{t_0}} |\mathbb{Q}(\mathbf{V}_t \leqslant \mathbf{v}) - \mathbb{Q}'(\mathbf{V}_t \leqslant \mathbf{v})|$. Then, we say a mapping $f(\mathbb{Q}')$ is sup-norm continuous at $\mathbb{Q}' = \mathbb{Q}$ if $\lim_{\delta\to 0} \sup_{\mathbb{Q}':\|\mathbb{Q}'-\mathbb{Q}\|_\infty \leqslant \delta} |f(\mathbb{Q}') - f(\mathbb{Q})| = 0$.

Lemma H.2. *For each $1 \leqslant t \leqslant t_0$, we have the following results.*

1. *For any fixed $\mathbb{Q}$, $G_t(\lambda, \mathbb{Q})$ is non-decreasing in λ. Moreover, $G_t(\lambda, \mathbb{P}^*)$ is strictly increasing in $\lambda \in (0,1]$ under Assumption A1.*

2. *For any fixed $\lambda \in (0,1]$, $D_t(\lambda, \mathbb{Q})$, $N_t(\lambda, \mathbb{Q})$, and $G_t(\lambda, \mathbb{Q})$ are sup-norm continuous in $\mathbb{Q}$ at $\mathbb{Q} = \mathbb{P}^*$ under Assumption A1.*

3. *$\Lambda_t(\mathbb{Q})$ is sup-norm continuous at $\mathbb{Q} = \mathbb{P}^*$ under Assumption A1. In addition, $\Lambda_t(\mathbb{P}^*) > 0$.*

By definition, $\lambda_t = \Lambda_t(\mathbb{P}^*)$, where $\mathbb{P}^*$ denotes the true probability measure of $(V_1, \cdots, V_{t_0})$. On the other hand, define the empirical measure (recall $V_{k,t} = \mathbb{P}(\tau_k < t | X_{k,1}, \cdots, X_{k,t})$)

$$\mathbb{P}_K = \frac{1}{K} \sum_{k=1}^{K} \delta_{(V_{k,1}, \cdots, V_{k,t_0})}.$$

It is not hard to verify that

$$\widehat{\lambda}_{K,t} = \Lambda_t(\mathbb{P}_K).$$

Now we are able to prove the first part of theorem. Let

$$\mathcal{C} = \left\{(-\infty, \mathbf{x}] : \mathbf{x} \in \mathbb{R}^{t_0}\right\}$$

where $(-\infty, \mathbf{x}]$ denotes the set $(-\infty, x_1] \times \cdots \times (-\infty, x_{t_0}]$. It is known that $\mathcal{C}$ is a Vapnik-Čhervonenkis class and thus, $\lim_{K \to \infty} \sup_{C \in \mathcal{C}} |\mathbb{P}_K(\mathbf{V}_{t_0} \in C) - \mathbb{P}^*(\mathbf{V}_{t_0} \in C)| = 0 \quad a.s.$ (see, e.g., Shorack and Wellner (2009)). In other words,

$$\lim_{K \to \infty} \|\mathbb{P}_K - \mathbb{P}^*\|_\infty = 0 \text{ a.s.} \tag{H.1}$$

This result combined with the third statement of Lemma H.2 implies

$$\lim_{K \to \infty} \Lambda_t(\mathbb{P}_K) = \Lambda_t(\mathbb{P}^*) \quad a.s.$$

That is, $\lim_{K \to \infty} \widehat{\lambda}_{K,t} = \lambda_t$ a.s. This completes our proof for the first statement of the theorem. We proceed to the second and third statements of the theorem. Let

$$J_t(\mathbb{Q}) = \mathbb{E}_{\mathbb{Q}}\left(V_t \mathbb{1}\left\{\mathbf{V}_t \leqslant \mathbf{\Lambda}_t(\mathbb{Q})\right\}\right) \text{ and } H_t(\mathbb{Q}) = \mathbb{Q}\left(\mathbf{V}_t \leqslant \mathbf{\Lambda}_t(\mathbb{Q})\right).$$

We can see that the mapping H_t is the composition of $D_t(\cdot, \mathbb{Q})$ and $\Lambda_t(\mathbb{Q})$. According to Lemma H.1 and Lemma H.2, both mappings are sup-norm continuous at $\mathbb{Q} = \mathbb{P}^*$, and as a result, their composition $H_t(\mathbb{Q})$ is also sup-norm continuous at $\mathbb{Q} = \mathbb{P}^*$. Similarly, according to Lemma H.1 and Lemma H.2, we can also see that $J_t(\mathbb{Q})$ is sup-norm continuous at $\mathbb{Q} = \mathbb{P}^*$.

These results, combined with (H.1), give

$$\lim_{K\to\infty} H_t(\mathbb{P}_K) = H_t(\mathbb{P}^*) \text{ a.s.,} \tag{H.2}$$

and

$$\lim_{K\to\infty} J_t(\mathbb{P}_K) = J_t(\mathbb{P}^*) \text{ a.s.} \tag{H.3}$$

Note that

$$H_t(\mathbb{P}_K) = K^{-1}|S^*_{t+1}| \text{ and } \frac{J_t(\mathbb{P}_K)}{H_t(\mathbb{P}_K)} = \mathbb{E}(\mathrm{FNP}_{t+1}(\mathbb{T})|\mathcal{F}_t). \tag{H.4}$$

(H.2), (H.3), and (H.4) together complete the second and third statements of the theorem.

In the rest of the proof, we show that (4.2) holds.

We first show that for $t \leqslant L := \frac{\log(1-\alpha)}{\log(1-\theta)}$, $\lambda_t = 1$. We show this by induction. For $t = 0$, $\lambda_0 = 1$ by definition. Assume that for some $t \geqslant 1$, $\lambda_0 = \cdots = \lambda_{t-1} = 1$, then

$$G_t(\lambda, \mathbb{P}^*) = \mathbb{E}\left[V_t | V_t \leqslant \lambda, \mathbf{V}_{t-1} \leqslant \mathbf{\Lambda}_{t-1}(\mathbb{P}^*)\right] = \mathbb{E}\left[V_t | V_t \leqslant \lambda\right].$$

In addition, $G_t(1, \mathbb{P}^*) = \mathbb{E}(V_t) = \mathbb{P}(\tau_1 < t) = 1 - (1-\theta)^t \leqslant \alpha$ for $t \leqslant L$. By Lemma H.2, we know that $G_t(\lambda, \mathbb{P}^*)$ is increasing in λ. Thus,

$$\lambda_t = \sup\left\{\lambda : G_t(\lambda, \mathbb{P}^*) \leqslant \alpha \text{ and } \lambda \in [0,1]\right\} = 1.$$

This completes the induction. As a result, for $1 \leqslant t \leqslant L$, $\mathbb{E}\left[V_t | V_t \leqslant \lambda_t, \mathbf{V}_{t-1} \leqslant \boldsymbol{\lambda}_{t-1}\right] = G_t(1, \mathbb{P}^*) = 1 - (1-\theta)^t$.

We proceed to the proof of (4.2) for $t \geqslant L$. Note that $N_t(\lambda, \mathbb{P}^*)$ and $D_t(\lambda, \mathbb{P}^*)$ are continuous in $\lambda \in (0,1)$ (note that $\mathbf{V}_t$ has a joint probability density function by Lemma H.1). Moreover, by Lemma H.2 and Lemma H.1, $D_t(\lambda, \mathbb{P}^*) > 0$ for $\lambda > 0$. Thus, for each t, $G_t(\lambda_t, \mathbb{P}^*) = \alpha$ is equivalent to

$$G_t(1, \mathbb{P}^*) \geqslant \alpha. \tag{H.5}$$

We will show (H.5) $t > L$ by induction. Let $\lfloor L \rfloor$ be the largest integer smaller or equal to L. According to the definition of L, we can see that

$$G_{\lfloor L \rfloor+1}(1, \mathbb{P}^*) = \mathbb{E}(V_{\lfloor L \rfloor+1}) = 1 - (1-\theta)^{\lfloor L \rfloor+1} > \alpha.$$

This proves the base case for the induction.

Assume that for $1 \leqslant s \leqslant t-1$, $G_s(1, \mathbb{P}^*) > \alpha$. Then,

$$G_t(1, \mathbb{P}^*) = \mathbb{E}\left[V_t | \mathbf{V}_{t-1} \leqslant \boldsymbol{\lambda}_{t-1}\right] = \mathbb{E}\left[\mathbb{E}(V_t | X_{1,1:t-1}) | \mathbf{V}_{t-1} \leqslant \boldsymbol{\lambda}_{t-1}\right], \tag{H.6}$$

where $\boldsymbol{\lambda}_{t-1} = (\lambda_1, \cdots, \lambda_{t-1})$. On the other hand,

$$\begin{aligned}
&\mathbb{E}(V_t | X_{1,1:t-1}) \\
=&\mathbb{E}\left[\mathbb{P}\left(\tau_1 < t | X_{1,1:t}\right) | X_{1,1:t-1}\right] \\
=&\mathbb{P}\left(\tau_1 < t | X_{1,1:t-1}\right) \\
=&\mathbb{P}\left(\tau_1 \leqslant t-1 | X_{1,1:t-1}\right) \\
=&\delta_{1,t-1} \\
=&\theta + (1-\theta) V_{t-1},
\end{aligned}$$

where the last two equations are due to Lemma F.8. The above display and (H.6) give

$$G_t(1, \mathbb{P}^*) = \mathbb{E}\left[\theta + (1-\theta) V_{t-1} | \mathbf{V}_{t-1} \leqslant \boldsymbol{\lambda}_{t-1}\right] = \theta + (1-\theta) \mathbb{E}\left[V_{t-1} | \mathbf{V}_{t-1} \leqslant \boldsymbol{\lambda}_{t-1}\right].$$

By induction assumption, we have

$$\mathbb{E}\left[V_{t-1} | \mathbf{V}_{t-1} \leqslant \boldsymbol{\lambda}_{t-1}\right] = \alpha.$$

The above two equations give

$$G_t(1, \mathbb{P}^*) = \theta + (1-\theta)\alpha > \alpha.$$

This completes our proof. □

Remark H.1. *A key observation in the above proof is that $\widehat{\lambda}_{K,t} = \Lambda_t(\mathbb{P}_K)$ while $\lambda_t = \Lambda_t(\mathbb{P}^*)$, where $\mathbb{P}_K$ is the empirical measure and $\mathbb{P}^*$ is the underlying probability measure of the process $\{V_{k,t}\}_{1 \leqslant t \leqslant t_0}$. Thus, to show that $\widehat{\lambda}_{K,t}$ converges to λ_t (i.e., $\Lambda_t(\mathbb{P}_K)$ converges to $\Lambda_t(\mathbb{P}^*)$), it suffices to show that the functional $\Lambda_t(\cdot)$ is continuous and the empirical measure $\mathbb{P}_K$ converges to $\mathbb{P}^*$ in some sense as $K \to \infty$. In the proof, the above heuristics are justified through Vapnik-Čhervonenkis (VC) theory. In particular, as a standard result in VC theory, the empirical measure converges to the underlying measure uniformly over the set $\mathcal{C} = \{(-\infty, \mathbf{x}] : \mathbf{x} \in \mathbb{R}^{t_0}\}$. That is, $\mathbb{P}_K$ converges to $\mathbb{P}^*$ in $\|\cdot\|_\infty$ norm almost surely. The supporting lemma (Lemma H.2) is mainly arguing that the functional of interest is continuous under this norm.*

Moreover, VC theory and theory of empirical processes in general are helpful in under-

standing the convergence of empirical measure over general probability spaces. Based on VC theory, many additional results (e.g., convergence rate) can be developed in addition to the uniform convergence result over the set $\mathcal{C}$ mentioned above. We refer the readers to the book (Shorack and Wellner, 2009) and references therein for a comprehensive review.

H.2 Proof of Theorem 3

Theorem 3. *Suppose that data follow a special case of the model given in Example 1 when $\eta = 1$ and $\tau_0 \sim Geom(\theta)$, and Assumption A2 holds. Let*

$$W_t = \mathbb{P}\left(\tau_0 < t \middle| X_{k,s}, 1 \leqslant k \leqslant K, 1 \leqslant s \leqslant t\right),$$

and

$$T = \min\{t : W_t > \alpha\}.$$

Then, $\mathbb{T}_K^* = (T, \cdots, T)$. Moreover, the following asymptotic results hold.

1. $\lim_{K\to\infty}(T - \tau_0) = 1$ *a.s.,*

2. $\lim_{K\to\infty} \mathrm{LFNR}_{t+1}(\mathbb{T}_K^*) = 0$ *a.s.,*

3. $\lim_{K\to\infty} K^{-1}|S_{K,t+1}^*| = \mathbb{1}(\tau_0 \geqslant t)$ *a.s.*

Proof of Theorem 3. We first note that under the model considered in this theorem, $W_{1,t} = \cdots = W_{K,t} = \mathbb{P}(\tau_0 < t|\mathcal{F}_t)$. Thus, according to $\mathbb{T}^*$, if $W_{1,t} \leqslant \alpha$, then $\sum_{k\in S_t} W_{t,k} \leqslant \alpha|S_t|$, and $S_{t+1} = S_t$. Moreover, if for some t such that $S_t = \{1, \cdots, K\}$ and $W_{1,t+1} > \alpha$, then for any $S \neq \varnothing$, $\sum_{k\in|S|} W_{k,t+1} = W_{k,t+1}|S| > \alpha|S|$, and thus $S_{t+1} = \varnothing$. Thus, $\mathbb{T}^* = (T, \cdots, T)$. In other words, $S_t = \{1, \cdots, K\}$ for $t \leqslant T$ and $S_t = \varnothing$ for $t > T$.

Note that for $t \leqslant T$, $\mathcal{F}_t = \sigma(\{W_{k,s}\}, 1 \leqslant s \leqslant t, 1 \leqslant k \leqslant K)$. Let $\widetilde{W}_{k,t} = \mathbb{P}(\tau_0 < t|X_{k,s}, 1 \leqslant k \leqslant K, 1 \leqslant s \leqslant t)$, which is the conditional probability without deactivating any stream. Then, $W_{k,t} = \widetilde{W}_{k,t}$ for $t \leqslant T$ where we recall $T = \inf\{t : \widetilde{W}_{1,t} > \alpha\}$. We have

$$\begin{aligned}\widetilde{W}_{k,t} =& \frac{\sum_{s=0}^{t-1} \theta(1-\theta)^s \prod_{r=s+1}^{t} \prod_{k=1}^{K} q(X_{k,r})/p(X_{k,r})}{\sum_{s=0}^{t-1} \theta(1-\theta)^s \prod_{r=s+1}^{t} \prod_{k=1}^{K} q(X_{k,r})/p(X_{k,r}) + (1-\theta)^t} \\ =& \frac{\sum_{s=0}^{t-1} \theta(1-\theta)^s \exp\{\sum_{k=1}^{K} l_{k,s,t}\}}{\sum_{s=0}^{t-1} \theta(1-\theta)^s \exp\{\sum_{k=1}^{K} l_{k,s,t}\} + (1-\theta)^t},\end{aligned} \tag{H.7}$$

where we define $l_{k,s,t} = \sum_{r=s+1}^{t} \log(q(X_{k,r})/p(X_{k,r}))$.

For each $u \in \mathbb{Z}_+ \cup \{0\}$, let $A_u = \{\tau_0 = u\}$. By the strong law of large numbers, under Assumption A2,

$$\mathbb{P}\left(\lim_{K\to\infty} \frac{1}{K}\sum_{k=1}^{K} l_{k,s,t} = \mathbb{E}(l_{1,s,t}|\tau_0 = u)\middle| A_u\right) = 1 \tag{H.8}$$

for each $s, t, u \in \mathbb{Z}_+ \cup \{0\}$ with $s < t$. In particular,

$$\mathbb{E}(l_{1,s,t}|\tau_0 = u) = \begin{cases} -(t-s)\mathbb{E}_{Z_1\sim p}\log(p(Z_1)/q(Z_1)) < 0 & \text{if } t \leqslant u \\ \mathbb{E}_{Z_2\sim q}\log(q(Z_2)/p(Z_2)) > 0 & \text{if } t = u+1 \text{ and } s = u. \end{cases}$$

Thus, for each $s < t \leqslant u$ we have

$$\mathbb{P}\left(\lim_{K\to\infty} \sum_{k=1}^{K} l_{k,s,t} = -\infty\middle| A_u\right) = 1, \tag{H.9}$$

and for $t = u+1 = s+1$,

$$\mathbb{P}\left(\lim_{K\to\infty} \sum_{k=1}^{K} l_{k,s,t} = \infty\middle| A_u\right) = 1.$$

According to (H.7), (H.8) and (H.9), we have that for each $t \leqslant u$

$$\mathbb{P}\left(\lim_{K\to\infty} \widetilde{W}_{k,t} = 0\middle| A_u\right) = 1.$$

Moreover, for $t \geqslant u+1$,

$$\mathbb{P}\left(\lim_{K\to\infty} \widetilde{W}_{k,t} = 1\middle| A_u\right) = 1.$$

Combining the above two equations for different $u \in \mathbb{Z}_+ \cup \{0\}$, we arrive at

$$\mathbb{P}\left(\lim_{K\to\infty} \widetilde{W}_{k,t} = \mathbb{1}(t \geqslant \tau_0 + 1)\right) = 1.$$

In other words,

$$\lim_{K\to\infty} \widetilde{W}_{1,t} = \mathbb{1}(t \geqslant \tau_0 + 1) \text{ a.s.}$$

Now we turn to the analysis of $W_{k,t}$ and S_t for the proposed procedure. Let ω be a sample path with $\lim_{K\to\infty} \widetilde{W}_{k,t}(\omega) = \mathbb{1}(t \geqslant \tau_0(\omega) + 1)$ for all $t = 1, 2, \cdots$. Then, there exists $K_0(\omega)$ large enough such that $\widetilde{W}_{1,t}(\omega) < \alpha$ for $t \leqslant \tau_0(\omega)$ and $\widetilde{W}_{1,\tau_0(\omega)+1}(\omega) > \alpha$ for all $K \geqslant K_0(\omega)$. Then, we have $T(\omega) = \inf\{t : \widetilde{W}_{1,t}(\omega) > \alpha\} = \tau_0(\omega) + 1$. Note that the set of such sample path ω has a probability of one. Thus,

$$\lim_{K\to\infty}(T-\tau_0)=1 \text{ and } \lim_{K\to\infty}W_{k,t}=0 \text{ for } t\leqslant\tau_0 \text{ a.s.}$$

This proves the first statement of the theorem. For the second statement, we have

$$\lim_{K\to\infty}\mathbb{E}\left(\mathrm{FNP}_{t+1}(\mathbb{T}^*)|\mathcal{F}_t\right)=\lim_{K\to\infty}\frac{\sum_{k=1}^K \mathbb{1}(T>t)W_{k,t}}{\{\sum_{k=1}^K \mathbb{1}(T>t)\}\vee 1}=\lim_{K\to\infty}W_{k,t}\mathbb{1}(T>t)=0 \text{ a.s.}$$

For the third statement, we have

$$\lim_{K\to\infty}K^{-1}|S_{t+1}|=\lim_{K\to\infty}\mathbb{1}(T>t)=\mathbb{1}(\tau_0\geqslant t) \text{ a.s.}$$

□

H.3 Proof of supporting lemmas in Section H.1

Lemma H.1. *Under model $\mathcal{M}_s$ and Assumption A1, we have the following results.*

1. *For each $t\geqslant 1$, $(V_1,\cdots,V_t)$ has a continuous and strictly positive joint density function over $(0,1)^t$ (with respect to the Lebesgue measure).*
2. *For any $(v_1,\cdots,v_t)\in(0,1)^t$, $\mathbb{P}(V_1\leqslant v_1,\cdots,V_t\leqslant v_t)>0$.*
3. *For any $(v_1,\cdots,v_t)\in(0,1)^t$, the conditional distribution of V_{t+1} given $V_1\leqslant v_1,\cdots,V_t\leqslant v_t$ has a continuous and positive density function over $(0,1)$.*

Proof of Lemma H.1. Note that the second statement of the lemma is obvious given the first statement, and the third statement is a straightforward application of a combination of the first and second statements. Thus, it suffices to show the first statement of the lemma. In what follows, we prove the first statement by induction.

For Z_1 follow the density function $p(\cdot)$, Z_2 follows the density function $q(\cdot)$, let $f_1(\cdot)$ and $f_2(\cdot)$ be the density functions of $q(Z_1)/p(Z_1)$ and $q(Z_2)/p(Z_2)$. By Assumption A1, $f_i(z)>0$ for all $z>0$ and $i=1,2$.

For $t=1$, under the model $\mathcal{M}_\mathrm{s}$, $X_{1,1}$ follows the mixture density $(1-\theta)p(\cdot)+\theta q(\cdot)$. Thus, $q(X_{1,1})/p(X_{1,1})$ has the density function $(1-\theta)f_1+\theta f_2$, which is strictly positive and continuous over $\mathbb{R}_+$. Note that $V_1=\frac{q(X_{1,1})/p(X_{1,1})}{(1-\theta)/\theta+q(X_{1,1})/p(X_{1,1})}$. By standard calculation of density of random variable after transformation, we can see that the density of V_1 is

$$f_{V_1}(v)=\frac{c}{(1-v)^2}\left\{(1-\theta)f_1\left(\frac{cv}{1-v}\right)+\theta f_2\left(\frac{cv}{1-v}\right)\right\}, \tag{H.10}$$

where $c = (1-\theta)/\theta$. This density function is strictly positive and continuous for $v \in (0,1)$.

Assume the induction assumption that the joint density for $(V_1, \cdots, V_t)$, denoted by $f_{V_1,\cdots,V_t}(v_1, \cdots, v_t)$, is strictly positive and continuous over $(0,1)^t$. We proceed to showing $f_{V_1,\cdots,V_{t+1}}(v_1, \cdots, v_{t+1})$ is strictly positive and continuous over $(0,1)^{t+1}$. Recall that $V_{t+1} = \frac{q(X_{t+1,1})/p(X_{t+1,1})}{(1-\theta)(1-V_t)/(\theta+(1-\theta)V_t)+q(X_{1,1})/p(X_{1,1})}$. With a similar derivation as that for (H.10), we have the conditional density of V_{t+1} given $V_1 = v_1, \cdots, V_t = v_t$ is

$$
\begin{aligned}
&f_{V_{t+1}|V_1=v_1,\cdots,V_t=v_t}(v) \\
=&\frac{c_t}{(1-v)^2}\left\{(1-\theta_t) f_1\left(\frac{c_t v}{1-v}\right) + \theta_t f_2\left(\frac{c_t v}{1-v}\right)\right\},
\end{aligned}
$$

where we define $c_t = \frac{(1-\theta)(1-v_t)}{\theta+(1-\theta)v_t} > 0$ and $\theta_t = \mathbb{P}(\tau_1 \leqslant t | V_1 = v_1, \cdots, V_t = v_t) = v_t(1-\theta) + \theta \in (0,1)$. It is easy to see that both c_t and θ_t are continuous in v_t. As a result, $f_{V_{t+1}|V_1=v_1,\cdots,V_t=v_t}(v_{t+1})$ is strictly positive and is continuous in $v_1, \cdots, v_{t+1}$ for $v_1, \cdots, v_{t+1} \in (0,1)$ and so is $f_{V_1,\cdots,V_{t+1}}(v_1, \cdots, v_{t+1}) = f_{V_1,\cdots,V_t}(v_1, \cdots, v_t) f_{V_{t+1}|V_1=v_1,\cdots,V_t=v_t}(v_{t+1})$. This completes our induction and the proof of the lemma. □

Lemma H.2. *For each $1 \leqslant t \leqslant t_0$, we have the following results.*

1. *For any fixed $\mathbb{Q}$, $G_t(\lambda, \mathbb{Q})$ is non-decreasing in λ. Moreover, $G_t(\lambda, \mathbb{P}^*)$ is strictly increasing in $\lambda \in (0,1]$ under Assumption A1.*

2. *For any fixed $\lambda \in (0,1]$, $D_t(\lambda, \mathbb{Q})$, $N_t(\lambda, \mathbb{Q})$, and $G_t(\lambda, \mathbb{Q})$ are sup-norm continuous in $\mathbb{Q}$ at $\mathbb{Q} = \mathbb{P}^*$ under Assumption A1.*

3. *$\Lambda_t(\mathbb{Q})$ is sup-norm continuous at $\mathbb{Q} = \mathbb{P}^*$ under Assumption A1. In addition, $\Lambda_t(\mathbb{P}^*) > 0$.*

Proof of Lemma H.2. For $t = 0, 1, \cdots$ and $\lambda < \lambda'$, let $\widetilde{V}$ be a random variable following the same distribution as $V_t | \mathbf{V}_{t-1} \leqslant \mathbf{\Lambda}_{t-1}(\mathbb{Q})$. Then, by the definition of conditional expectation, we have

$$
\begin{aligned}
&G_t(\lambda', \mathbb{Q}) - G_t(\lambda, \mathbb{Q}) \\
=&Z^{-1}\left[\mathbb{E}_{\mathbb{Q}}\left(\widetilde{V}\mathbb{1}\left\{\widetilde{V} \leqslant \lambda'\right\}\right)\mathbb{Q}\left(\widetilde{V} \leqslant \lambda\right) - \mathbb{E}_{\mathbb{Q}}\left(\widetilde{V}\mathbb{1}\left\{\widetilde{V} \leqslant \lambda\right\}\right)\mathbb{Q}\left(\widetilde{V} \leqslant \lambda'\right)\right] \\
=&Z^{-1}\left[\mathbb{E}_{\mathbb{Q}}\left(\widetilde{V}\mathbb{1}\left\{\lambda < \widetilde{V} \leqslant \lambda'\right\}\right)\mathbb{Q}\left(\widetilde{V} \leqslant \lambda\right) - \mathbb{E}_{\mathbb{Q}}\left(\widetilde{V}\mathbb{1}\left\{\widetilde{V} \leqslant \lambda\right\}\right)\mathbb{Q}\left(\lambda < \widetilde{V} \leqslant \lambda'\right)\right]
\end{aligned}
$$

where $Z = \mathbb{Q}\left(\widetilde{V} \leqslant \lambda\right)\mathbb{Q}\left(\widetilde{V} \leqslant \lambda'\right)$. Let $\widetilde{V}'$ be an independent copy of $\widetilde{V}$, then the above display implies

$$
\begin{aligned}
& G_t(\lambda', \mathbb{Q}) - G_t(\lambda, \mathbb{Q}) \\
= & Z^{-1}\left[\mathbb{E}_{\mathbb{Q}}\left(\widetilde{V}'\mathbb{1}\left\{\lambda < \widetilde{V}' \leqslant \lambda', \widetilde{V} \leqslant \lambda\right\}\right) - \mathbb{E}_{\mathbb{Q}}\left(\widetilde{V}\mathbb{1}\left\{\lambda < \widetilde{V}' \leqslant \lambda', \widetilde{V} \leqslant \lambda\right\}\right)\right] \\
= & Z^{-1}\mathbb{E}_{\mathbb{Q}}\left[\left(\widetilde{V}' - \widetilde{V}\right)\mathbb{1}\left\{\lambda < \widetilde{V}' \leqslant \lambda', \widetilde{V} \leqslant \lambda\right\}\right],
\end{aligned} \tag{H.11}
$$

Because $\left(\widetilde{V}' - \widetilde{V}\right)\mathbb{1}\left\{\lambda < \widetilde{V}' \leqslant \lambda', \widetilde{V} \leqslant \lambda\right\} \geqslant 0$, $G_t(\lambda', \mathbb{Q}) - G_t(\lambda, \mathbb{Q}) \geqslant 0$ from the above display.

In what follows, we use induction to prove the rest of the lemma. Namely, for $\lambda \in (0,1)$, we will prove the following statements for $t = 1, 2, \cdots, t_0$:

$$G_t(\lambda, \mathbb{P}^*) \text{ is strictly increasing in } \lambda; \tag{H.12}$$

$$D_t(\lambda, \mathbb{Q}), N_t(\lambda, \mathbb{Q}), \text{ and } G_t(\lambda, \mathbb{Q}) \text{ are sup-norm continuous at } \mathbb{Q} = \mathbb{P}^*; \tag{H.13}$$

$$\Lambda_t(\mathbb{Q}) \text{ is sup-norm continuous at } \mathbb{Q} = \mathbb{P}^*. \tag{H.14}$$

We start with the base case that $t = 1$. In this case, the conditional distribution $V_1|\mathbf{V}_0 \leqslant \mathbf{\Lambda}_0(\mathbb{Q})$ is the same as the unconditional distribution of V_1 for any $\mathbb{Q}$. According to Lemma H.1, V_1 has a strictly positive and continuous density function over $(0,1)$ under $\mathbb{P}^*$. Thus, $\mathbb{P}^*\left(\left(\widetilde{V}' - \widetilde{V}\right)\mathbb{1}\left\{\lambda < \widetilde{V}' \leqslant \lambda', \widetilde{V} \leqslant \lambda\right\} \geqslant 0\right) > 0$ for $\widetilde{V}$ and $\widetilde{V}'$ are identically distributed as V_1. According to (H.11), $G_1(\lambda', \mathbb{P}^*) - G_1(\lambda, \mathbb{P}^*) > 0$. That is, $G_1(\lambda, \mathbb{P}^*)$ is strictly increasing in λ. This proves the base case for (H.12). For (H.13) and (H.14) the proof of the base cases is similar to that of the induction given below. Thus, we omit the proof for their base cases here.

Now we assume that (H.12), (H.13), and (H.14) hold for $t = 1, 2, \cdots, s-1$. We proceed to prove these equations for $t = s$. First, note that $V_t|\mathbf{V}_{t-1} \leqslant \mathbf{\Lambda}_{t-1}(\mathbb{P}^*)$ has a continuous and strictly positive density function over $(0,1)$. Thus, (H.12) is proved by combining (H.11) with similar arguments as those for the base case where $t = 1$.

Proof of (H.13) for $t = s$. By the induction assumption, $\Lambda_1(\mathbb{Q}), \cdots, \Lambda_{s-1}(\mathbb{Q})$ is sup-norm continuous in $\mathbb{Q}$ at $\mathbb{Q} = \mathbb{P}^*$. This implies that $(\lambda, \mathbf{\Lambda}_{s-1}(\mathbb{Q}))$, a vector-valued mapping, is also sup-norm continuous in $\mathbb{Q}$ at $\mathbb{Q} = \mathbb{P}^*$. On the other hand, $(\lambda, \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)) \in (0,1]^s$ by induction assumptions, and $\mathbf{V}_t$ has a continuous joint probability cumulative function at $(\lambda, \mathbf{\Lambda}_{s-1}(\mathbb{P}^*))$ (by Lemma H.1). Combining these results, we can see that $\mathbb{P}^*(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q}))$ is sup-norm continuous at $\mathbb{Q} = \mathbb{P}^*$.

Now we analyze the mapping $D_s(\lambda, \mathbb{Q}) = \mathbb{Q}\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right)$.

$$\begin{aligned}
&\left|D_s(\lambda, \mathbb{Q}) - D_s(\lambda, \mathbb{P}^*)\right| \\
=& \left|\mathbb{Q}\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) - \mathbb{P}^*\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)\right)\right| \\
\leqslant& \left|\mathbb{Q}\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) - \mathbb{P}^*\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right)\right| \\
&+ \left|\mathbb{P}^*\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) - \mathbb{P}^*\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)\right)\right| \\
\leqslant& \|\mathbb{Q} - \mathbb{P}^*\|_\infty \\
&+ \left|\mathbb{P}^*\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) - \mathbb{P}^*\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)\right)\right|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\limsup_{\|\mathbb{Q}-\mathbb{P}^*\|_\infty \to 0} \left|D_s(\lambda, \mathbb{Q}) - D_s(\lambda, \mathbb{P}^*)\right| \\
=& \lim_{\|\mathbb{Q}-\mathbb{P}^*\|_\infty \to 0} \|\mathbb{Q} - \mathbb{P}^*\|_\infty \\
&+ \lim_{\|\mathbb{Q}-\mathbb{P}^*\|_\infty \to 0} \left|\mathbb{P}^*\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) - \mathbb{P}^*\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)\right)\right| \\
=& 0.
\end{aligned}$$

That is, $D_s(\lambda, \mathbb{Q})$ is sup-norm continuous at $\mathbb{P}^*$. Moreover, by Lemma H.1 and $(\lambda, \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)) \in (0,1]^s$, we have $D_s(\lambda, \mathbb{P}^*) > 0$. This further implies that $D_s(\lambda, \mathbb{Q})^{-1}$ is also sup-norm continuous at $\mathbb{P}^*$.

We proceed to the analysis of $N_s(\lambda, \mathbb{Q})$. We have

$$\begin{aligned}
N_s(\lambda, \mathbb{Q}) =& \mathbb{E}_{\mathbb{Q}}\left[V_s \mathbb{1}\left\{V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right\}\right] \\
=& \mathbb{E}_{\mathbb{Q}}\left[\int_0^1 \mathbb{1}\{r < V_s\} dr \mathbb{1}\left\{V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right\}\right] \\
=& \int_0^1 \mathbb{Q}\left(r < V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) dr \\
=& \mathbb{Q}\left(V_s \leqslant \lambda, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) \\
&- \int_0^\lambda \mathbb{Q}\left(V_s \leqslant r, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) dr \\
=& D_s(\lambda, \mathbb{Q}) - \int_0^\lambda D_s(r, \mathbb{Q}) dr.
\end{aligned} \tag{H.15}$$

We have already shown that the first term $D_s(\lambda, \mathbb{Q})$ on the right-hand side of the above display is sup-norm continuous at $\mathbb{P}^*$. We take a closer look at the second term,

$$
\begin{aligned}
&\left|\int_0^\lambda D_s(r,\mathbb{Q})dr - \int_0^\lambda D_s(r,\mathbb{P}^*)dr\right| \\
\leqslant &\int_0^\lambda \left|\mathbb{Q}\left(V_s \leqslant r, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) - \mathbb{P}^*\left(V_s \leqslant r, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right)\right| dr \\
&+ \int_0^\lambda \left|\mathbb{P}^*\left(V_s \leqslant r, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) - \mathbb{P}^*\left(V_s \leqslant r, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)\right)\right| dr \\
\leqslant &\|\mathbb{Q} - \mathbb{P}^*\|_\infty \\
&+ \int_0^\lambda \left|\mathbb{P}^*\left(V_s \leqslant r, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) - \mathbb{P}^*\left(V_s \leqslant r, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)\right)\right| dr.
\end{aligned}
\tag{H.16}
$$

Since $\mathbf{\Lambda}_{s-1}(\mathbb{Q})$ is sup-norm continuous at $\mathbb{Q} = \mathbb{P}^*$, for any $\varepsilon > 0$, there exists $\delta > 0$ such that $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$ implies $\|\mathbf{\Lambda}_{s-1}(\mathbb{Q}) - \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)\| \leqslant \varepsilon$. Then, for each $r \in [0,1]$ and $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$, $\|(r, \mathbf{\Lambda}_{s-1}(\mathbb{Q})) - (r, \mathbf{\Lambda}_{s-1}(\mathbb{P}^*))\| \leqslant \varepsilon$, and

$$
\begin{aligned}
&\sup_{\|\mathbb{Q}-\mathbb{P}^*\|\leqslant\delta} \left|\mathbb{P}^*\left(V_s \leqslant r, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{Q})\right) - \mathbb{P}^*\left(V_s \leqslant r, \mathbf{V}_{s-1} \leqslant \mathbf{\Lambda}_{s-1}(\mathbb{P}^*)\right)\right| \\
\leqslant &\sup_{\|\mathbf{v}_s-\mathbf{v}'_s\|\leqslant\varepsilon, \mathbf{v}_s, \mathbf{v}'_s\in[0,1]^s} \left|\mathbb{P}^*\left(\mathbf{V}_s \leqslant \mathbf{v}_s\right) - \mathbb{P}^*\left(\mathbf{V}_s \leqslant \mathbf{v}'_s\right)\right|.
\end{aligned}
\tag{H.17}
$$

By Lemma H.1, $\mathbf{V}_s$ has a continuous density function. Thus, its cumulative distribution function, $\mathbb{P}^*\left(\mathbf{V}_s \leqslant \mathbf{v}_s\right)$, is continuous over $[0,1]^s$. As $[0,1]^s$ is compact, this continuity implies that the cumulative distribution is also uniformly continuous over $[0,1]^s$. That is, for any ϵ_1 small enough, there is $\epsilon > 0$, such that

$$
\sup_{\|\mathbf{v}_s-\mathbf{v}'_s\|\leqslant\varepsilon, \mathbf{v}_s, \mathbf{v}'_s\in[0,1]^s} \left|\mathbb{P}^*\left(\mathbf{V}_s \leqslant \mathbf{v}_s\right) - \mathbb{P}^*\left(\mathbf{V}_s \leqslant \mathbf{v}'_s\right)\right| \leqslant \varepsilon_1.
$$

Combine the above inequality with (H.16) and (H.17), we can see that for any $\varepsilon_1 > 0$, there is $0 < \delta < \varepsilon_1$ such that for $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$,

$$
\left|\int_0^\lambda D_s(r,\mathbb{Q})dr - \int_0^\lambda D_s(r,\mathbb{P}^*)dr\right| \leqslant \delta + \varepsilon_1 \leqslant 2\varepsilon_1.
$$

Therefore, $\int_0^\lambda D_s(r,\mathbb{Q})dr$ is sup-norm continuous at $\mathbb{Q} = \mathbb{P}^*$. This result, combined with (H.15), shows that $N_s(\lambda,\mathbb{Q})$ is sup-norm continuous at $\mathbb{Q} = \mathbb{P}^*$.

Finally, the sup-norm continuity of $G_s(\lambda,\mathbb{Q})$ is implied by that of $D_s(\lambda,\mathbb{Q})^{-1}$ and $N_s(\lambda,\mathbb{Q})$ for $\lambda \in (0,1]$.

Proof of (H.14) **for** $t = s$**.** Recall $\Lambda_s(\mathbb{Q}) = \sup\left\{\lambda : G_s(\lambda,\mathbb{Q}) \leqslant \alpha \text{ and } \lambda \in [0,1]\right\}$. We discuss two cases.

Case 1: $\Lambda_s(\mathbb{P}^*) = 1$**.** For any sufficiently small $\varepsilon > 0$, by the strict increasing property of $G_s(\lambda, \mathbb{P}^*)$ there exists $\varepsilon_1 > 0$ such that $G_s(\lambda', \mathbb{P}^*) < G_s(\Lambda_s(\mathbb{P}^*), \mathbb{P}^*) - 2\varepsilon_1$ for all $\lambda' \leqslant \Lambda_s(\mathbb{P}^*) - \varepsilon$. On the other hand, according to the sup-norm continuity of $G_s(\Lambda_s(\mathbb{P}^*) - \varepsilon, \mathbb{Q})$ at $\mathbb{Q} = \mathbb{P}^*$, there exists $\delta > 0$ such that $|G_s(\Lambda_s(\mathbb{P}^*) - \varepsilon, \mathbb{Q}) - G_s(\Lambda_s(\mathbb{P}^*) - \varepsilon, \mathbb{P}^*)| \leqslant \varepsilon_1$ for all $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$. Then, for all $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$ and $\lambda' \leqslant \Lambda_s(\mathbb{P}^*) - \varepsilon$, we have

$$
\begin{aligned}
& G_s(\lambda', \mathbb{Q}) \\
\leqslant & G_s(\Lambda_s(\mathbb{P}^*) - \varepsilon, \mathbb{Q}) \\
\leqslant & G_s(\Lambda_s(\mathbb{P}^*) - \varepsilon, \mathbb{P}^*) + |G_s(\Lambda_s(\mathbb{P}^*) - \varepsilon, \mathbb{Q}) - G_s(\Lambda_s(\mathbb{P}^*) - \varepsilon, \mathbb{P}^*)| \\
\leqslant & G_s(\Lambda_s(\mathbb{P}^*) - \varepsilon, \mathbb{P}^*) + \varepsilon_1 \\
\leqslant & G_s(\Lambda_s(\mathbb{P}^*), \mathbb{P}^*) - \varepsilon_1 \\
\leqslant & \alpha - \varepsilon_1.
\end{aligned}
$$

This implies $1 - \varepsilon = \Lambda_s(\mathbb{P}^*) - \varepsilon \leqslant \Lambda_s(\mathbb{Q}) \leqslant 1$ for all $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$.

Case 2: $\Lambda_s(\mathbb{P}^*) < 1$**.** Using similar arguments as those for the Case 1, we arrive at that for any $\varepsilon > 0$ there exists $\delta > 0$ such that $\Lambda_s(\mathbb{P}^*) - \varepsilon \leqslant \Lambda_s(\mathbb{Q})$ for all $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$. We proceed to an upper bound of $\Lambda_s(\mathbb{Q})$.

Note that in this case, $G_s(\Lambda_s(\mathbb{P}^*), \mathbb{P}^*) = \alpha$. According to the definition of $\Lambda_s(\mathbb{P}^*)$, for any $\varepsilon > 0$, then there exists $\varepsilon_1 > 0$ such that $G_s(\lambda', \mathbb{P}^*) > \alpha + 2\varepsilon_1$ for all $\lambda' \geqslant \Lambda_s(\mathbb{P}^*) + \varepsilon$. On the other hand, according to the sup-norm continuity of $G_s(\Lambda_s(\mathbb{P}^*) + \varepsilon, \mathbb{Q})$ at $\mathbb{Q} = \mathbb{P}^*$, there exists δ such that $|G_s(\Lambda_s(\mathbb{P}^*) + \varepsilon, \mathbb{Q}) - G_s(\Lambda_s(\mathbb{P}^*) + \varepsilon, \mathbb{P}^*)| \leqslant \varepsilon_1$ for all $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$. Then, for all $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$ and $\lambda' > \Lambda_s(\mathbb{P}^*) + \varepsilon$, we have

$$
\begin{aligned}
& G_s(\lambda', \mathbb{Q}) \\
\geqslant & G_s(\Lambda_s(\mathbb{P}^*) + \varepsilon, \mathbb{Q}) \\
\geqslant & \alpha + 2\varepsilon_1 - |G_s(\Lambda_s(\mathbb{P}^*) + \varepsilon, \mathbb{Q}) - G_s(\Lambda_s(\mathbb{P}^*) + \varepsilon, \mathbb{P}^*)| \\
\geqslant & \alpha + \varepsilon_1.
\end{aligned}
$$

This implies that for $\lambda' > \Lambda_s(\mathbb{P}^*) + \varepsilon$ and $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$, $G_s(\lambda', \mathbb{Q}) > \alpha$. Thus, $\Lambda_s(\mathbb{Q}) \leqslant \Lambda_s(\mathbb{P}^*) + \varepsilon$ for $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$. Combining the upper bound and lower bound of $\Lambda_s(\mathbb{Q})$, we arrive at

$$
|\Lambda_s(\mathbb{Q}) - \Lambda_s(\mathbb{P}^*)| \leqslant \varepsilon
$$

for $\|\mathbb{Q} - \mathbb{P}^*\|_\infty \leqslant \delta$.

This completes the proof of (H.14).

Finally, we show $\Lambda_t(\mathbb{P}^*) > 0$. This is true because $G_t(\lambda, \mathbb{P}^*)$ is continuous and strictly increasing in λ and $\lim_{\lambda\to 0+} G_t(\lambda, \mathbb{P}^*) = 0 < \alpha$.

$\square$

I Calculations for Example 3

We start with calculating $\mathbb{P}\left(\tau_k = 0 | X_{k,1} = x_{k,1}, \cdots, X_{k,t} = x_{k,t}\right)$. Let $t_1 = t_4 = 3$ and $t_2 = t_3 = 1$. Under the model specified in the example, we have $\tau_k = 0$ or $\tau_k = t_k$ a.s. for $k = 1, \cdots, 4$. As a result, we have

$$\mathbb{P}(\tau_k \leqslant t-1 | X_{k,1} = x_{k,1}, \cdots, X_{k,t} = x_{k,t}) = 1$$

for $t \geqslant t_k + 1$.

To simplify the calculation for the other cases, we first prove the following auxiliary result: under the model specified in this example, for any $x_{k,1}, \cdots, x_{k,t} \in \{0, 1\}$ and $0 \leqslant t \leqslant t_k$,

$$\begin{aligned}
&\mathbb{P}\left(\tau_k \leqslant t-1 | X_{k,1} = 0, \cdots, X_{k,t} = 0\right) \\
\leqslant &\mathbb{P}\left(\tau_k \leqslant t-1 | X_{k,1} = x_{k,1}, \cdots, X_{k,t} = x_{k,t}\right) \\
\leqslant &\mathbb{P}\left(\tau_k \leqslant t-1 | X_{k,1} = 1, \cdots, X_{k,t} = 1\right).
\end{aligned} \tag{I.1}$$

Indeed, direct calculation gives

$$\begin{aligned}
&\mathbb{P}\left(\tau_k \leqslant t-1 | X_{k,1} = x_{k,1}, \cdots, X_{k,t} = x_{k,t}\right) \\
= &\frac{\mathbb{P}(\tau_k = 0)(0.51)^{\sum_{s=1}^t x_{k,t}}(0.49)^{t-\sum_{s=1}^t x_{k,t}}}{\mathbb{P}(\tau_k = 0)(0.51)^{\sum_{s=1}^t x_{k,t}}(0.49)^{t-\sum_{s=1}^t x_{k,t}} + \mathbb{P}(\tau_k = t_k)(0.5)^t}.
\end{aligned} \tag{I.2}$$

The above display is monotonically increasing in $\sum_{s=1}^t x_{k,t}$. Thus, (I.1) is proved.

Let $\widetilde{W}_{k,t} := \mathbb{P}\left(\tau_k \leqslant t-1 | X_{k,1} = x_{k,1}, \cdots, X_{k,t} = x_{k,t}\right)$. Using (I.1) and (I.2), we obtain that for $0 \leqslant t \leqslant t_k$,

$$\widetilde{W}_{k,t} \in \left[\frac{\mathbb{P}(\tau_k = 0)(0.49)^t}{\mathbb{P}(\tau_k = 0)(0.49)^t + \mathbb{P}(\tau_k = t_k)(0.5)^t}, \frac{\mathbb{P}(\tau_k = 0)(0.51)^t}{\mathbb{P}(\tau_k = 0)(0.51)^t + \mathbb{P}(\tau_k = t_k)(0.5)^t}\right].$$

Plugging $\mathbb{P}(\tau_k = 0)$ and $\mathbb{P}(\tau_k = t_k) = 1 - \mathbb{P}(\tau_k = 0)$ into the above equations, we obtain that $\widetilde{W}_{k,t} = 1$ for $t \geqslant 4$, and for $0 \leqslant t \leqslant 3$, the a.s. range of $\widetilde{W}_{k,t}$s are given below (numbers are rounded to the third decimal place).

$\widetilde{W}_{k,t} \in$	$t=1$	$t=2$	$t=3$
$k=1$	$[0.098, 0.102]$	$[0.096, 0.104]$	$[0.095, 0.105]$
$k=2$	$[0.395, 0.405]$	$\{1\}$	$\{1\}$
$k=3$	$[0.425, 0.435]$	$\{1\}$	$\{1\}$
$k=4$	$[0.545, 0.555]$	$[0.540, 0.560]$	$[0.535, 0.565]$

With these numbers, the following inequalities can be verified.

$$\begin{aligned}\widetilde{W}_{1,1} < \alpha < \widetilde{W}_{2,1} < \widetilde{W}_{3,1} < \widetilde{W}_{4,1},\\ \frac{1}{3}(\widetilde{W}_{1,1} + \widetilde{W}_{2,1} + \widetilde{W}_{3,1}) \leqslant 0.314 < \alpha = 0.34,\\ \frac{1}{3}(\widetilde{W}_{1,1} + \widetilde{W}_{2,1} + \widetilde{W}_{4,1}) \geqslant 0.346 > \alpha\\ \frac{1}{2}(\widetilde{W}_{1,1} + \widetilde{W}_{4,1}) \leqslant 0.329 < \alpha.\end{aligned}$$

The above inequalities implies that $\mathbb{E}[\mathrm{FNP}_2(\mathbb{T})|\mathcal{F}_1] \leqslant \alpha$ is equivalent to

$$S_2 \in \big\{\{1,2,3\}, \{1,2\}, \{1,3\}, \{1,4\}, \{1\}, \varnothing\big\}.$$

Now we consider S_3. We can verify the following inequalities.

$$\begin{aligned}\widetilde{W}_{1,2} < \alpha < \widetilde{W}_{4,2} < \widetilde{W}_{2,2} = \widetilde{W}_{3,2},\\ \frac{1}{2}(\widetilde{W}_{1,2} + \widetilde{W}_{2,2}) = \frac{1}{2}(\widetilde{W}_{1,2} + \widetilde{W}_{3,2}) \geqslant 0.548 > \alpha\\ \frac{1}{2}(\widetilde{W}_{1,2} + \widetilde{W}_{4,2}) \leqslant 0.332 < \alpha.\end{aligned}$$

The above inequalities implies that $\mathbb{E}[\mathrm{FNP}_3(\mathbb{T})|\mathcal{F}_2] \leqslant \alpha$ is equivalent to that $S_3 \subset S_2$ and

$$S_3 \in \big\{\{1,4\}, \{1\}, \varnothing\big\}.$$

Similarly, for S_4, we have

$$\begin{aligned}\widetilde{W}_{1,3} < \alpha < \widetilde{W}_{4,3} < \widetilde{W}_{2,3} = \widetilde{W}_{3,3},\\ \frac{1}{2}(\widetilde{W}_{1,2} + \widetilde{W}_{2,2}) = \frac{1}{2}(\widetilde{W}_{1,2} + \widetilde{W}_{3,2}) \geqslant 0.547 > \alpha\\ \frac{1}{2}(\widetilde{W}_{1,2} + \widetilde{W}_{4,}) \leqslant 0.336 < \alpha.\end{aligned}$$

This implies that $\mathbb{E}[\mathrm{FNP}_4(\mathbb{T})|\mathcal{F}_3] \leqslant \alpha$ is equivalent to that $S_4 \subset S_3$ and

$$S_4 \in \big\{\{1,4\},\{1\},\varnothing\big\}.$$

Finally, since $\widetilde{W}_{k,t} = 1$ for all $t \geqslant 4$ and $k = 1,\cdots,4$, we obtain $S_t = \varnothing$ for $t \geqslant 5$.

Enumerating all the index sets satisfying the constraint, we obtain that $\sup_{\mathbb{T}\in\mathcal{T}_\alpha} \mathbb{E}(\mathrm{U}_2(\mathbb{T})) = 7$ and the maximum achieved if and only if $S_1 = \{1,2,3,4\}$ and $S_2 = \{1,2,3\}$. In addition, $\sup_{\mathbb{T}\in\mathcal{T}_\alpha} \mathbb{E}(\mathrm{U}_4(\mathbb{T})) = 10$ and the maximum is achieved if and only if $S_1 = \{1,2,3,4\}$, $S_2 = \{1,4\}$, $S_3 = \{1,4\}$ and $S_4 = \{1,4\}$. However, these two maxima cannot be achieved at the same time as they require different choices of S_2.

References

Harris, T. E. (1960). A lower bound for the critical probability in a certain percolation process, *Mathematical Proceedings of the Cambridge Philosophical Society* **56**: 13–20.

Kamae, T., Krengel, U. and O'Brien, G. L. (1977). Stochastic inequalities on partially ordered spaces, *The Annals of Probability* **5**: 899–912.

Lindvall, T. (1999). On Strassen's theorem on stochastic domination, *Electronic Communications in Probability* **4**: 51–59.

Lindvall, T. (2002). *Lectures on the coupling method*, Dover Publications, New York, NY.

Shorack, G. R. and Wellner, J. A. (2009). *Empirical processes with applications to statistics*, SIAM, Philadelphia, PA.

Strassen, V. (1965). The existence of probability measures with given marginals, *The Annals of Mathematical Statistics* **36**: 423–439.

Thorisson, H. (2000). *Coupling, Stationarity, and Regeneration*, Springer, New York, NY.