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 3
B Proof Sketch 3
C Proof of Theorem 6 5
D Proof of Theorems 1 and 5 9
E Proof of Theorem 4 10
F Proof of Propositions B.1 and B.2 11
  F.1 Stochastic ordering 12
  F.2 Stochastic ordering and Markov chains on $S_u$ and $S_o$ 13
  F.3 Proof of Propositions B.1 and B.2 17
  F.4 Proof of supporting lemmas in Section F.2 19
G Proof of Lemma 1 and Propositions 1 - 3 29
H Proof of Theorem 2 and Theorem 3 30
  H.1 Proof of Theorem 2 30
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 $S \subset \{1, \cdots, K\}$: $X_{S,t} = (X_{k,t})_{k \in S}$.

- $X_{k,s:t}$: $X_{k,s:t} = (X_{k,r})_{s \leq r \leq t}$.

- $\mathbb{T}^A$: an arbitrary sequential procedure.

- $S^A_t$: The set of active streams at time $t$ given by procedure $\mathbb{T}^A$.

- $\mathcal{F}^A_t$: $\sigma$-field of information obtained up to time $t$ following $\mathbb{T}^A$.

- $W^A_{k,t}$: posterior probability $\mathbb{P}(\tau_k < t|\mathcal{F}^A_t)$ at the $k$-th stream following $\mathbb{T}^A$ at time $t$.

- $W^A_{S,t}$ for some set $S \subset \{1, \cdots, K\}$: $W^A_{S,t} = (W^A_{k,t})_{k \in S}$.

- $\mathbb{T}^*$: the proposed sequential procedure.

- $S^*_t, \mathcal{F}^*_t, W^*_t, W^*_{S,t}$ are defined similarly for procedure $\mathbb{T}^*$.

- $\mathbb{T}^{AP}_{t_0}$: the sequential procedure that takes the same steps as $\mathbb{T}^A$ up to time $t_0$ (meaning $S^*_{t_0} = S^A_t$ for $1 \leq t \leq t_0$) and updates by Algorithm 1 from time $t_0 + 1$ and onward.

- $S^{AP}_{t_0}, \mathcal{F}^{AP}_{t_0}, W^{AP}_{k,t}, W^{AP}_{S,t}$ are defined similarly for procedure $\mathbb{T}^{AP}_{t_0}$.

- $\overset{d}{=}$: equal in distribution.

- $\emptyset$: a vector with zero length.

- dim: length of a vector, where dim($\emptyset$) = 0.

- $Z \sim \mathcal{N}(0, 1)$: 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 \( T^A \in \mathcal{T}_\alpha \) and any \( t \), (5.2) holds for \( s = s_0 + 1 \), assuming that it holds for \( s \leq s_0 \). The induction step is proved by the following three steps.

1. Show that \( T^{AP}_{t_0 + 1} \) is ‘better’ than \( T^A \) conditional on \( \mathcal{F}_{t_0} \).
2. Show that \( T^{AP}_{t_0} \) is ‘better’ than \( T^{AP}_{t_0 + 1} \) conditional on \( \mathcal{F}_{t_0} \).
3. Show that \( T^{AP}_{t_0} \) is ‘better’ than \( T^A \) conditional on \( \mathcal{F}_{t_0} \) 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 \( \mathcal{F}_{t_0} \). Roughly, we prove the first step by replacing \( t_0 \) with \( t_0 + 1 \) in the induction assumption and taking conditional expectation given \( \mathcal{F}_{t_0} \), 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 \((\mathcal{S}_0, \preceq)\) as follows. Let

\[
\mathcal{S}_0 = \bigcup_{k=1}^{K} \{ \mathbf{v} = (v_1, \ldots, v_k) \in [0,1]^k : 0 \leq v_1 \leq \cdots v_k \leq 1 \} \cup \{ \emptyset \},
\]

where \( \emptyset \) represents a vector with zero length. For \( \mathbf{u} \in \mathcal{S}_0 \), let \( \text{dim}(\mathbf{u}) \) be the length of the vector \( \mathbf{u} \).

**Definition 1.** For \( \mathbf{u}, \mathbf{v} \in \mathcal{S}_0 \), we say \( \mathbf{u} \preceq \mathbf{v} \) if \( \text{dim}(\mathbf{u}) \geq \text{dim}(\mathbf{v}) \) and \( u_i \leq v_i \) for \( i = 1, \ldots, \text{dim}(\mathbf{v}) \). In addition, we say \( \mathbf{u} \preceq \emptyset \) for any \( \mathbf{u} \in \mathcal{S}_0 \).

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

**Proposition B.1.** Let \( \{x_t, s_t, 1 \leq t \leq t_0\} \) be any sequence in the support of the stochastic process \( \{(X_{k,t})_{k \in S^A_t}, S^A_t, 1 \leq t \leq t_0\} \) following a sequential procedure \( T^A \in \mathcal{T}_\alpha \). Then, there
exists a coupling of $\mathcal{S}_o$-valued random variables $(\hat{W}, \hat{W}')$ such that

$$\hat{W} = \left( \left( W_{k,t_0}^{AP_{t_0}^A} \right)_{k \in S_{t_0+1}} \right) \left\{ \left( X_{k,t} \right)_{k \in S_t^A} = x_t, S_t^A = s_t, 1 \leq t \leq t_0 \right\},$$

$$\hat{W}' = \left( \left( W_{k,t_0+1}^{AP_{t_0+1}^A} \right)_{k \in S_{t_0+1}} \right) \left\{ \left( X_{k,t} \right)_{k \in S_t^A} = x_t, S_t^A = s_t, 1 \leq t \leq t_0 \right\},$$

and $\hat{W} \leq \hat{W}'$ a.s., where $\leq$ denotes that random variables on both sides are identically distributed.

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

$$Y_s = \left[ \left( W_{k,t_0+s}^{AP_{t_0}^A} \right)_{k \in S_{t_0+s}} \right] \in \mathcal{S}_o.$$

Under model $\mathcal{M}_s$, 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 $y, y' \in \mathcal{S}_o$ such that $y \leq y'$, there exists a coupling $(\hat{Y}_s, \hat{Y}'_s), s = 0, 1, \ldots$, satisfying

1. $\{\hat{Y}_s : s \geq 0\}$ has the same distribution as the conditional process $\{Y_s : s \geq 0\}$ given $Y_0 = y$, and $\{\hat{Y}'_s : s \geq 0\}$ has the same distribution as the conditional process $\{Y_s : s \geq 0\}$ given $Y_0 = y'$.

2. $\hat{Y}_s \leq \hat{Y}'_s$, a.s. for all $s \geq 0$.

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

Roughly, Proposition B.1 shows that the sequential procedure $\mathbb{T}_{AP_{t_0}}$ tends to have a stochastically smaller detection statistic, in terms of the partial order $\leq$, than that of $\mathbb{T}_{AP_{t_0+1}}$ 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 $t_0, s \geq 0$ and any sequential detection procedure $\mathbb{T}_A^A \in \mathcal{T}_o$, let $\mathcal{F}_{t}^{A}$ be the information filtration and $S_t^A$ be the set of active streams.
at time $t$ given by $T^A$. Then,

$$
E \left[ |S^A_{t_0+s}| \bigg| F^A_{t_0} \right] \leq E \left[ |S^{AP}_{t_0+s}| \bigg| F^A_{t_0} \right] \text{ a.s.}
$$

Proof of Theorem 5.2. 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 $T^A \in \mathcal{T}_0$ 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 $T^A \in \mathcal{T}_0$ 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 $T^{AP}_{t_0}$ is defined as the sequential procedure that takes the same steps as $T^A$ up to time $t_0$ and updates by Algorithm 1 from time $t_0 + 1$ and onward, and the sequential procedure $T^{AP}_{t_0+1}$ is defined similarly.

Step 1: comparing $T^{AP}_{t_0+1}$ and $T^A$. 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

$$
E \left[ |S^A_{t_0+s_0+1}| \bigg| F^A_{t_0+1} \right] \leq E \left[ |S^{AP}_{t_0+s_0+1}| \bigg| F^A_{t_0+1} \right] \text{ a.s.}
$$

Taking conditional expectation $E \left[ |F^A_{t_0} \right]$ on both sides, we arrive at

$$
E \left[ |S^A_{t_0+s_0+1}| \bigg| F^A_{t_0} \right] \leq E \left[ |S^{AP}_{t_0+s_0+1}| \bigg| F^A_{t_0} \right] \text{ a.s.} \tag{C.2}
$$

Step 2: comparing $T^{AP}_{t_0+1}$ and $T^{AP}_{t_0}$. First, define a function $\phi_{t,s} : S_0 \rightarrow \mathbb{R}$,

$$
\phi_{t,s}(u) = E \left[ \dim \left( \left[ W^{AP}_{S^{AP}_{t_0+s},t} \bigg| S_{t+s} \right] = u \right) \right] = E \left[ \dim \left( \left[ W^{AP}_{S^{AP}_{t_0+s},t} \bigg| S_{t+s} \right] = u \right) \right] \tag{C.3}
$$

for $t, s \geq 0$. Here, for a set $S$, and time points $s$ and $t$, $W^{AP}_{S,t} = (W^{AP}_{S,k})_{k \in S}$, where $W^{AP}_{S,t} = \mathbb{P} \left( \tau_k < s \bigg| F^A_{t} \right)$.

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

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

Here, to see the superscript of the process in the above equation is $AP^{t_0}$, we used the fact that if we follow the procedure $T^{AP}_{t_0}$ and switch to the proposed procedure at time $t_0 + 1$,
then the overall sequential procedure is still $T^{AP_t_0}$.

Also from Proposition B.2, we can see that for any $u \leq u' \in S_0$, there exists a coupling $(\widehat{Y}_s, \widehat{Y}'_s)$ such that $\widehat{Y}_s$ has the same distribution as $[W^{AP}_{t_0+1}, s_{t_0+1}, t+1]$ given $[W^{AP}_{t_0, t_0+1}, t_0+1] = u$, $\widehat{Y}'_s$ has the same distribution as $[W^{AP}_{t_0+1}, s_{t_0+1}, t+1]$ given $[W^{AP}_{t_0, t_0+1}, t_0+1] = u'$, and $\widehat{Y}_s \leq \widehat{Y}'_s$ a.s. Thus,

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

According to the definition of the partial relationship `$\leq$', $\widehat{Y}_s \leq \widehat{Y}'_s$ implies $\dim(\widehat{Y}_s) \geq \dim(\widehat{Y}'_s)$. Combining this result with the above display, we conclude that $\phi_{t,s}(u) \geq \phi_{t,s}(u')$ for any $u \leq u' \in S_0$.

Next, we write $E\left[S^{AP}_{t_0+1} \mid F^A_{t_0}\right]$ and $E\left[S^{AP}_{t_0+1} \mid F^A_{t_0}\right]$ in terms of the conditional expectation involving the function $\phi_{t,s}$. We start with $E\left[S^{AP}_{t_0+1} \mid F^A_{t_0}\right]$. By the iterative law of conditional expectation and (C.4), we obtain

$$E\left[S^{AP}_{t_0+1} \mid F^A_{t_0}\right] = E\left[E\left[S^{AP}_{t_0+1} \mid \left(W^{AP}_{t_0, t_0+1} \mid F^A_{t_0}\right)\right]\right].$$

According to the definition of the information filtration $F^A_{t_0}$, we further write the above conditional expectation as

$$E\left[S^{AP}_{t_0+1} \mid F^A_{t_0}\right] = E\left[\phi_{t_0+1, s_0}\left([W^{AP}_{t_0, t_0+1} \mid F^A_{t_0}\right), k \in S^A_{t_0+1}, 1 \leq r \leq t_0\right].$$

Similarly, we have

$$E\left[S^{AP}_{t_0+1} \mid F^A_{t_0}\right] = E\left[\phi_{t_0+1, s_0}\left([W^{AP}_{t_0+1, t_0+1} \mid F^A_{t_0}\right), k \in S^A_{t_0+1}, 1 \leq r \leq t_0\right].$$

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

$$\widehat{W} = \left[W^{AP}_{t_0, t_0+1} \mid \{X_{S^A_r}, x_r, S^A_r = s_r, 1 \leq r \leq t_0\}\right].$$
\[ \hat{W}' \overset{d}{=} \left[ W^{\text{AP},t_0+1}_{t_0+1,t_0+1} \right] \{ X_{s_{t_0+1},r} = x_r, S_{t_0+1} = s_r, 1 \leq r \leq t_0 \}, \]

and
\[ \hat{W} \leq \hat{W}' \text{ a.s.,} \]

where ‘\( \overset{d}{=} \)’ means two random variables on both sides have the same distribution. Thus,
\[
\mathbb{E} \left[ \phi_{t_0+1,s_0} \left( W^{\text{AP},t_0}_{t_0+1,t_0+1} \right) \right] \{ X_{s_{t_0+1},r} = x_r, S_{t_0+1} = s_r, 1 \leq r \leq t_0 \} = \mathbb{E} \phi_{t_0+1,s_0} (\hat{W}) \quad (C.7)
\]

and
\[
\mathbb{E} \left[ \phi_{t_0+1,s_0} \left( W^{\text{AP},t_0+1}_{t_0+1,t_0+1} \right) \right] \{ X_{s_{t_0+1},r} = x_r, S_{t_0+1} = s_r, 1 \leq r \leq t_0 \} = \mathbb{E} \phi_{t_0+1,s_0} (\hat{W}'). \quad (C.8)
\]

On the other hand, note that we have shown \( \phi_{t_0+1,s_0}(u) \geq \phi_{t_0+1,s_0}(u') \) for any \( u \leq u' \in S_0 \) and \( \hat{W} \leq \hat{W}' \text{ a.s. by the coupling. Thus,} \)
\[ \phi_{t_0+1,s_0}(\hat{W}) \geq \phi_{t_0+1,s_0}(\hat{W}') \text{ a.s.} \]

Combining the above inequality with (C.7) and (C.8), we arrive at
\[
\mathbb{E} \left[ \phi_{t_0+1,s_0} \left( W^{\text{AP},t_0}_{t_0+1,t_0+1} \right) \right] \{ X_{s_{t_0+1},r} = x_r, S_{t_0+1} = s_r, 1 \leq r \leq t_0 \}
\geq
\mathbb{E} \left[ \phi_{t_0+1,s_0} \left( W^{\text{AP},t_0}_{t_0+1,t_0+1} \right) \right] \{ X_{s_{t_0+1},r} = x_r, S_{t_0+1} = s_r, 1 \leq r \leq t_0 \}
\]

for each sequence \( \{ x_r, s_r, 1 \leq r \leq t_0 \} \) that is in the support of the process \( \{ X_{s_{t_0+1},r}, S_{t_0+1}, 1 \leq r \leq t_0 \} \). Comparing the above inequality with (C.5) and (C.6), we conclude that
\[
\mathbb{E} \left[ S_{t_0+1,s_0+1} \left\| \mathcal{F}_{t_0} \right\| \right] \geq \mathbb{E} \left[ S_{t_0+1,s_0+1} \left\| \mathcal{F}_{t_0} \right\| \right] \text{ a.s.} \quad (C.9)
\]

Step 3: combining results from Steps 1 and 2. Combining (C.2) and (C.9), we obtain
\[
\mathbb{E} \left[ \left| S_{t_0+1,s_0+1} \right\| \mathcal{F}_{t_0} \right\| \right] \leq \mathbb{E} \left[ \left| S_{t_0+1,s_0+1} \right\| \mathcal{F}_{t_0} \right\| \right] \text{ a.s.,} \quad (C.10)
\]

which implies that (5.2) holds for arbitrary \( T^A \in \mathcal{T}_0, t_0, \) and \( s = s_0 + 1. \) This completes the induction. \( \square \)

Remark C.1. Proposition B.2 is used in Step 2 of the above proof, where we only use the property that \( \hat{Y}_s \leq \hat{Y}_s' \) is independent of \( t_0 \) and \( T^A. \) The independence between \( (\hat{Y}_s, \hat{Y}_s') \) and \( \mathcal{F}_{t_0} \) is an additional result that further characterizes the coupling process. We did not use
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 $T^A \in \mathcal{T}_\alpha$ be an arbitrary sequential procedure. Further let $T^{A_{t_0}}$ and $T^{A_{t_0+1}}$ be the switching procedures described above, with switching time $t_0$ and $t_0 + 1$, respectively, for some $t_0 \geq 0$. Then, $T^{A_{t_0}}, T^{A_{t_0+1}} \in \mathcal{T}_\alpha$ and under model $M_s$

$$E \left( U_t(T^A) \right) \leq E \left( U_t(T^{A_{t_0+1}}) \right) \leq E \left( U_t(T^{A_{t_0}}) \right) \leq E \left( U_t(T^*) \right),$$

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

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

Applying Theorem 3 but replacing $t_0$ by $t_0 + 1$, and taking expectation on both sides of the inequality, we obtain

$$E|S_{t_0+1+s}| \leq E|S_{t_0+1+s}|$$

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

$$E|S_t^A| \leq E|S_t^{A_{t_0+1}}|.$$ 

For $t < t_0 + 1$, as $T^A$ and $T^{A_{t_0+1}}$ share the same index set, we have

$$E|S_t^A| = E|S_t^{A_{t_0+1}}|.$$ 

Combining the above inequalities, we obtain

$$E|S_t^A| \leq E|S_t^{A_{t_0+1}}|$$

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

$$E\{U_t(T^A)\} = \sum_{s=1}^{t} E|S_s^A| \leq \sum_{s=1}^{t} E|S_s^{A_{t_0+1}}| = E\{U_t(T^{A_{t_0+1}})\}.$$
This proves the inequality for comparing procedures $T^A$ and $T^{AP_{t_0+1}}$. We then compare $T^{AP_{t_0+1}}$ and $T^{AP_{t_0}}$, based on the same arguments above except that we replace $T^A$ by $T^{AP_{t_0+1}}$, and replace $T^{AP_{t_0+1}}$ by $T^{AP_{t_0}}$. We obtain
\[ E\{U_t(T^{AP_{t_0+1}})\} \leq E\{U_t(T^{AP_{t_0}})\} \]
for all $t \geq 0$.

Finally, we compare $T^{AP_{t_0}}$ and $T^* = T^{AP_{t_0}}$ using a similar argument, which gives
\[ E\{U_t(T^{AP_{t_0}})\} \leq E\{U_t(T^*)\}. \]

\[ \square \]

**E Proof of Theorem 4**

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

We proceed to the analysis of $RL_t(T)$. By interchanging the order of double summation, we have
\[ RL_t(T) = \sum_{k=1}^{K} (T_k \wedge \tau_k \wedge t) = \sum_{k=1}^{K} \sum_{s=1}^{t} 1(s \leq T_k \wedge \tau_k) = \sum_{s=1}^{t} \sum_{k=1}^{K} 1(s \leq T_k \wedge \tau_k) = \sum_{s=1}^{t} \sum_{k \in S_s} \{1 - 1(\tau_k < s)\} \]
which leads to
\[ E\{RL_t(T)\} = \sum_{s=1}^{t} E\left[ \sum_{k \in S_s} \{1 - 1(\tau_k < s)\} \right] = \sum_{s=1}^{t} E\left[ E\left[ \sum_{k \in S_s} \{1 - 1(\tau_k < s)\} \right] | \mathcal{F}_s \right]. \]
Recall that $W_{k,s} = P(\tau_k < s | \mathcal{F}_s)$ and $S_s \in \mathcal{F}_s$. The above display yields
\[ E\{RL_t(T)\} = \sum_{s=1}^{t} E\left\{ \sum_{k \in S_s} (1 - W_{k,s}) \right\}. \]

From the above equation, we can see that in order to show $E\{RL_t(T^*)\} = \sup_{T \in \mathcal{T}_\alpha} E\{RL_t(T)\}$, it suffices to show $E\{\sum_{k \in S_t} (1 - W_{k,t})\}$ 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 $t_0, s \geq 0$ and any sequential detection procedure $T^A \in \mathcal{T}_\alpha$, let $\mathcal{F}_t^A$ be the information filtration and $S_t^A$ be the set of active
streams at time $t$ given by $T^A$. Then,

$$E \left[ \Psi([W^A_{S_{t_0+s},t_0+s}]) \right] \leq E \left[ \Psi([W^{AP}_{t_0+s},t_0+s]) \right] \text{ a.s.,}$$

where $\Psi : \mathcal{S}_o \rightarrow \mathbb{R}$ is defined as $\Psi(w) = \sum_{k=1}^{m} (1 - w_k)$ for $w = (w_1, \ldots, w_m) \in \mathcal{S}_o$.

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 $|S^A_t|$ with $\Psi (W^A_{S^A_{t},t})$ for $t$ taking different values in the proof of Theorem 6, we obtain the following inequality that is similar to (C.2)

$$E \left[ \Psi([W^A_{S_{t_0+s},t_0+s}]) \right] \leq E \left[ \Psi([W^{AP}_{t_0+s},t_0+s]) \right] \text{ a.s.} \quad (E.1)$$

for all $t_0$ and $s_0$. Next, we define a function $\tilde{\phi}_{t,s} : \mathcal{S}_o \rightarrow \mathbb{R}$,

$$\tilde{\phi}_{t,s}(u) = E \left[ \Psi([W^{AP}_{S_{AP_{t+s}},t}]) \right] \text{ a.s.} \quad (E.2)$$

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

$$E \left[ \Psi([W^{AP}_{S_{t_0+s},t_0+s}]) \right] \leq E \left[ \Psi([W^{AP}_{S_{t_0+s},t_0+s}]) \right] \text{ a.s.} \quad (E.3)$$

We point out that to arrive at the above inequality, the following property about $\Psi$ is used: $\Psi(w') \leq \Psi(w)$ for any $w, w' \in \mathcal{S}_o$ satisfying $w \preceq w'$.

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

$$E \left[ \Psi([W^A_{S_{t_0+s},t_0+s}]) \right] \leq E \left[ \Psi([W^{AP}_{S_{t_0+s},t_0+s}]) \right] \text{ a.s.,}$$

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 \((S, \preceq)\) is said to be a partially ordered space (or pospace) if \(\preceq\) is a partial order relation over the topological space \(S\) and the set \(\{(x, y) \in S^2 : x \preceq y\}\) is a closed subset of \(S^2\).

Definition 3 (Increasing functions over a partially ordered space). Let \((S_1, \preceq_{S_1})\) and \((S_2, \preceq_{S_2})\) be partially ordered polish spaces. A map \(g: S_1 \rightarrow S_2\) is said to be increasing if \(g(u) \preceq_{S_2} g(v)\) for all \(u, v \in S_1\).

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 \(P(X \geq x) \leq P(Y \geq x)\) for all real number \(x\). In this case, we write \(X \preceq_{st} Y\).

The following statements give some equivalent definitions for \(X \preceq_{st} Y\)

Fact 1. The following statements are equivalent.

1. \(X \preceq_{st} Y\).
2. For all increasing, bounded, and measurable functions \(g: \mathbb{R} \rightarrow \mathbb{R}, \mathbb{E}(g(X)) \leq \mathbb{E}(g(Y))\).
3. There exists a coupling \((\hat{X}, \hat{Y})\) such that \(\hat{X} \overset{d}{=} X, \hat{Y} \overset{d}{=} Y\), and \(\hat{X} \preceq \hat{Y}\) 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 \((S, \preceq)\) be a partially ordered polish space, and let \(X\) and \(Y\) be \(S\)-valued random variables. We say \(Y\) stochastically dominates \(X\), denoted by \(X \preceq_{st} Y\) if for all bounded, increasing, and measurable function \(g: S \rightarrow \mathbb{R}, \mathbb{E}(g(X)) \leq \mathbb{E}(g(Y))\).

Fact 2 (Strassen’s theorem for polish pospace, Theorem 2.4 in Lindvall (2002)). Let \((S, \preceq)\) be a polish partially ordered space, and let \(X\) and \(Y\) be \(S\)-valued random variables. Then, \(X \preceq_{st} Y\) if and only if there exists a coupling \((\hat{X}, \hat{Y})\) such that \(\hat{X} \overset{d}{=} X, \hat{Y} \overset{d}{=} Y\) and \(\hat{X} \preceq \hat{Y}\) a.s.
**Definition 6** (Stochastic dominance for Markov kernels). Let $K$ and $\tilde{K}$ be transition kernels for Markov chains over a partially ordered polish space $(\mathcal{S}, \preceq)$. The transition kernel $\tilde{K}$ is said to stochastically dominate $K$ if

$$x \preceq y \implies K(x, \cdot) \preceq_{st} \tilde{K}(y, \cdot).$$

In particular, if the above is satisfied for the same kernel $K = \tilde{K}$, 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, $(\mathcal{S}, \preceq)$, with transition kernels $K$ and $\tilde{K}$ where $\tilde{K}$ stochastically dominates $K$. Then, for all initial points $x_0 \preceq y_0$, there is a coupling $\{(\hat{X}_t, \hat{Y}_t)\}$ of $\{X_t\}$ starting at $x_0$ and $\{Y_t\}$ starting at $y_0$ such that

$$\hat{X}_t \preceq \hat{Y}_t \quad \forall t \text{ a.s.}$$

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

### F.2 Stochastic ordering and Markov chains on $\mathcal{S}_u$ and $\mathcal{S}_o$

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

$$\mathcal{S}_o = \bigcup_{k=1}^{K} \left\{ \mathbf{v} = (v_1, \ldots, v_k) \in [0,1]^k : 0 \leq v_1 \leq \cdots \leq v_k \leq 1 \right\} \cup \{\emptyset\}. $$

Here, we also define a space with unordered elements.

$$\mathcal{S}_u = \bigcup_{k=1}^{K}[0,1]^k \cup \{\emptyset\}. $$

We first present a lemma showing that the space $(\mathcal{S}_o, \preceq)$ is a polish partial order space.

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

\[ d(u, v) = \begin{cases} 
\max_{1 \leq m \leq \dim(u)} |u_m - v_m| & \text{if } \dim(u) = \dim(v) \geq 1 \\
0 & \text{if } u = v = \emptyset \\
2 & \text{if } \dim(u) \neq \dim(v)
\end{cases} \]

for \( u, v \in S_o \).

We define mappings \( I_o : S_o \to \{0, \cdots, K\} \) and \( H_o : S_o \to S_o \) as follows. For any \( u \in S_o \), define

\[
I_o(u) = \begin{cases} 
\sup \left\{ n : \sum_{i=1}^n u_i \leq \alpha n, n \in \{0, \ldots, \dim(u)\} \right\} & \text{if } \dim(u) \geq 1, u = (u_1, \ldots, u_{\dim(u)}) \\
0 & \text{if } \dim(u) = 0,
\end{cases}
\]

and

\[
H_o(u) = \begin{cases} 
(u_1, \cdots, u_{I_o(u)}) & \text{if } I_o(u) \geq 1, \\
\emptyset & \text{otherwise}.
\end{cases}
\]

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 \((W_{k,t})_{k \in S_t} = u\) and an index set \( S_t \) with \(|S_t| = \dim(u)\) in Algorithm 1, then the output \( S_{t+1} \) satisfies

\[ |S_{t+1}| = I_o(u) \text{ and } [(W_{k,t})_{k \in S_{t+1}}] = H_o([u]). \]

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

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

\[ \mathcal{T} \in \mathcal{T} \text{ and } \sum_{k=1}^K 1(T_k > t)W_{k,t} \leq \alpha \sum_{k=1}^K 1(T_k > t) \text{ for } t = 0, 1, 2, \cdots \]

The above expression is equivalent to

\[ S_{t+1} \text{ is } \mathcal{F}_t \text{ measurable }, S_{t+1} \subset S_t, \sum_{k \in S_{t+1}} W_{k,t} \leq \alpha \cdot |S_{t+1}| \]

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 $\mathbf{u} = (u_1, \ldots, u_m) \in \mathcal{S}_u$ with $\dim(\mathbf{u}) = m \geq 1$. Let $k_1, \ldots, k_l \in \{1, \ldots, m\}$ be distinct and satisfy
$$
\sum_{i=1}^l u_{k_i} \leq \alpha l.
$$
Then, $H_o([\mathbf{u}]) \leq [(u_{k_1}, \ldots, u_{k_l})]$. Moreover, if $H_o([\mathbf{u}]) = \emptyset$, then for any $S \subseteq \{1, \ldots, m\}$ with $|S| \geq 1$, $\sum_{i \in S} u_i > \alpha |S|$.

**Lemma F.5.** For any $\mathbf{u} \preceq \mathbf{v} \in \mathcal{S}_o$, $H_o(\mathbf{u}) \leq H_o(\mathbf{v})$. That is, the mapping $H_o(\mathbf{u})$ 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 $\mu$ 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 $Z_\delta$ be a random variable with the density function $\delta q + (1 - \delta)p$ and $L_\delta = L(Z_\delta)$. Then, for $0 \leq \delta_1 < \delta_2 \leq 1$, we have
$$
L_{\delta_1} \leq_{st} L_{\delta_2}.
$$

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 $V_{k,t} = \mathbb{P}(\tau_k < t | X_{k,1}, \ldots, X_{k,t})$. Then,
$$
V_{k,0} = 0 \text{ and } V_{k,t+1} = \frac{q(X_{k,t+1})/p(X_{k,t+1})}{(1-\theta)(1-V_{k,t})/(\theta + (1-\theta)V_{k,t}) + q(X_{k,t+1})/p(X_{k,t+1})}, \quad (F.2)
$$
Moreover, $\{V_{k,t}\}_{t=0,1,\ldots}$ are independent and identically distributed processes for different $k$.

**Lemma F.8.** Assume model $\mathcal{M}_s$ holds. Let $\delta_{k,t} = \mathbb{P}(\tau_k \leq t | X_{k,1}, \ldots, X_{k,t})$, then
$$
\delta_{k,t} = \theta + (1-\theta)V_{k,t},
$$
where $V_{k,t}$ is defined in (F.2).
Lemma F.9. Under model $\mathcal{M}_t$, the process $\{V_{t+1, t} \}_{t \geq 0}$ 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 $t \geq 1$ and $\mathbb{T}^A$, $[W_{s_{t+1}, t+1}^A]$ is conditionally independent of $\mathcal{F}_{t}^A$ given $[W_{s_{t+1}, t}^A]$. Moreover, the conditional density of $[W_{s_{t+1}, t+1}^A]$ at $v$ given $[W_{s_{t+1}, t}^A] = u \in \mathcal{S}_o$ is

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

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

Lemma F.11. For each $u \in \mathcal{S}_o$ with $\dim(u) = m \geq 1$, generate an $\mathcal{S}_o$-valued random variable $V$ as follows,

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

In addition, if $m = 0$, we let $V = \emptyset$. Then, $V \sim K_o(u, \cdot)$.

Lemma F.12. For $u, u' \in \mathcal{S}_o$ with $u \preceq u'$, we have $K_o(u, \cdot) \preceq_{st} K_o(u', \cdot)$.

Lemma F.13. For any $t, s \geq 0$ and $\mathbb{T}^A$, $[W_{s_{t+s}, t+s}^{AP}]$ is conditionally independent of $\mathcal{F}_{t+s}$ given $[W_{s_{t+s}, t+s}^{AP}]$. Moreover, the conditional density of $[W_{s_{t+s}, t+s}^{AP}]$ at $v$ given $[W_{s_{t+s}, t+s}^{AP}] = u \in \mathcal{S}_o$ is

$$K_o(u, v) := K_o(H_o(u), v) = \begin{cases} \sum_{l \in \Lambda_o(u)} \prod_{l=1}^{\Lambda_o(u)} K(H_o(u)_l, v_{\pi(l)}) & \text{if } \dim(v) = I_o(u) \geq 1, \\ 1 & \text{if } \dim(v) = I_o(u) = 0, \\ 0 & \text{otherwise,} \end{cases}$$

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

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 $[W_{s_{t+1}, t+1}^A]$ given $[W_{s_{t+1}, t}^A]$, where the index set $S_{t+1}^A$ is the same for the two random vectors.
In Lemma F.13, we consider the conditional distribution of \( W_{S_{t+1}^{AP}}^{AP} \) given \( W_{S_{t}^{AP},t+s}^{AP} \), where the two random vectors are associated with two different index sets \( S_{t+1}^{AP} \) and \( S_{t}^{AP} \). This difference reflects a key difference between the proposed one-step update rule and an arbitrary procedure.

Lemma F.14. For each \( u \in S_{o} \) and \( m = I_{o}(u) \), generate an \( S_{o} \)-valued random variable \( V \) as follows,

1. For each \( k \in \{1, \cdots, m\} \), generate \( Z_{k} \sim K(H_{o}(u)_{k}, \cdot) \) independently for different \( k \).
2. Let \( V = [(Z_{1}, ..., Z_{m})] \).

In addition, if \( m = 0 \), we let \( V = \emptyset \). Then, \( V \sim \mathbb{K}_{o}(u, \cdot) \).

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

Proposition B.1. Let \( \{x_{t}, s_{t}, 1 \leq t \leq t_{0}\} \) be any sequence in the support of the stochastic process \( \{(X_{k,t})_{k \in S_{t}^{A}}, S_{t}^{A}, 1 \leq t \leq t_{0}\} \) following a sequential procedure \( T^{A} \in T_{o} \). Then, there exists a coupling of \( S_{o} \)-valued random variables \( (\hat{W}, \hat{W}') \) such that

\[
\hat{W} \overset{d}{=} \left[ \left( W_{k,t_{0}}^{AP} \right)_{k \in S_{t_{0}}^{AP}} \right] \mid \{(X_{k,t})_{k \in S_{t}^{A}} = x_{t}, S_{t}^{A} = s_{t}, 1 \leq t \leq t_{0}\},
\]

\[
\hat{W}' \overset{d}{=} \left[ \left( W_{k,t_{0}+1}^{AP} \right)_{k \in S_{t_{0}+1}^{AP}} \right] \mid \{(X_{k,t})_{k \in S_{t}^{A}} = x_{t}, S_{t}^{A} = s_{t}, 1 \leq t \leq t_{0}\},
\]

and \( \hat{W} \leq \hat{W}' \) a.s., where \( \overset{d}{=} \) denotes that random variables on both sides are identically distributed.

Proof of Proposition B.1 First, given \( \{X_{S_{t}^{A}} = x_{t}, S_{t}^{A} = s_{t}, 1 \leq t \leq t_{0}\} \), \( W_{S_{t_{0}}^{A},t_{0}}^{A} \) is determined. To simplify the notation, we assume \( W_{S_{t_{0}}^{A},t_{0}}^{A} = W_{t_{0}} \in S_{u} \) given \( \{X_{S_{t}^{A}} = x_{t}, S_{t}^{A} = s_{t}, 1 \leq t \leq t_{0}\} \).

In addition, \( W_{S_{t_{0}+1}^{A},t_{0}}^{A} \) is determined by \( \{X_{S_{t}^{A}} = x_{t}, S_{t}^{A} = s_{t}, 1 \leq t \leq t_{0}\} \) and the sequential procedure \( T^{A} \). To simplify the notation, we assume \( W_{S_{t_{0}+1}^{A},t_{0}}^{A} = W_{t_{0}+1}^{*} \in S_{o} \) given \( \{X_{S_{t}^{A}} = x_{t}, S_{t}^{A} = s_{t}, 1 \leq t \leq t_{0}\} \). We clarify that \( W_{t_{0}+1}^{*} \) is a deterministic (and measurable) function of \( x_{t}, s_{t} \) for \( 1 \leq t \leq t_{0} \) (depending on the sequential procedure \( T^{A} \)).

According to Lemma F.10 (replacing \( t \) by \( t_{0} \)), the conditional distribution of \( W_{S_{t_{0}+1}^{A},t_{0}}^{A} \) given \( \{X_{S_{t}^{A}} = x_{t}, S_{t}^{A} = s_{t}, 1 \leq t \leq t_{0}\} \) is the same as the conditional distribution given \( W_{S_{t_{0}+1}^{A},t_{0}}^{A} = W_{t_{0}+1}^{*} \). Moreover, the conditional density is \( \mathbb{K}_{o}(W_{t_{0}+1}^{*}, \cdot) \).
We perform a similar analysis by replacing $\mathbb{T}^A$ by $\mathbb{T}^{AP_{t_0}}$ in the above analysis. We denote $\mathbb{W}_{t_0+1}^{AP_{t_0}} = \mathbb{W}_{t_0+1}^{AP_{t_0}}$ and obtain that the conditional density of $\mathbb{W}_{t_0+1}^{AP_{t_0}}$ given $\mathbb{W}_{S_{t_0+1}^{AP_{t_0}}}$ is $\mathbb{K}_0(\mathbb{W}_{t_0+1}, \cdot)$.

According to the above analysis and Strassen Theorem for pospace (Fact 2), to prove the proposition, it is sufficient to show $\mathbb{K}_0(\mathbb{W}_{t_0+1}, \cdot) \leq_{st} \mathbb{K}_0(\mathbb{W}_{t_0+1}, \cdot)$. By Lemma F.12 we have $\mathbb{K}_0(\mathbb{W}_{t_0+1}, \cdot) \leq_{st} \mathbb{K}_0(\mathbb{W}_{t_0+1}, \cdot)$ for any $\mathbb{W}_{t_0+1} = H_0([\mathbb{W}_{t_0}])$. There are two cases: 1) $\mathbb{W}_{t_0+1} = \emptyset$, and 2) $\mathbb{W}_{t_0+1} \neq \emptyset$. We analyze these cases separately. For the first case, $\mathbb{W}_{t_0+1} = \emptyset$ by definition of the partial order. For the second case, according to Lemma F.3 and Lemma F.4, we can see that $\mathbb{W}_{t_0+1} = H_0([\mathbb{W}_{t_0}]) \neq \emptyset$. Write $\mathbb{W}_{t_0} = (\mathbb{W}_{t_0,0}, \ldots, \mathbb{W}_{t_0,m})$ for some $m$, then $\mathbb{W}_{t_0+1}$ can be written as $\mathbb{W}_{t_0+1} = (\mathbb{W}_{t_0,k_1}, \ldots, \mathbb{W}_{t_0,k_l})$ for some distinct $k_1, \ldots, k_l \in \{1, \ldots, m\}$. According to Lemma F.3 for $\mathbb{T}^A$ to control LFNR at time $t_0+1$, $\mathbb{W}_{t_0+1}$ satisfies $\sum_{i=1}^l \mathbb{W}_{t_0,k_i} \leq \alpha t$. Thus, according to Lemma F.4, $\mathbb{W}_{t_0+1} = H_0([\mathbb{W}_{t_0}]) \leq \mathbb{W}_{t_0+1} = \mathbb{W}_{t_0+1}$.

**Proposition B.2.** Suppose that model $\mathcal{M}_s$ holds. Then for any $\mathbb{Y}, \mathbb{Y}' \in \mathcal{S}_0$ such that $\mathbb{Y} \leq \mathbb{Y}'$, there exists a coupling $(\hat{\mathbb{Y}}_s, \hat{\mathbb{Y}}'_s)$, $s = 0, 1, \ldots$, satisfying

1. $(\hat{\mathbb{Y}}_s : s \geq 0)$ has the same distribution as the conditional process $\{\mathbb{Y}_s : s \geq 0\}$ given $\mathbb{Y}_0 = \mathbb{Y}$, and $(\hat{\mathbb{Y}}'_s : s \geq 0)$ has the same distribution as the conditional process $\{\mathbb{Y}'_s : s \geq 0\}$ given $\mathbb{Y}_0 = \mathbb{Y}'$.

2. $\hat{\mathbb{Y}}_s \leq \hat{\mathbb{Y}}'_s$, a.s. for all $s \geq 0$.

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

**Proof of Proposition B.2.** Recall $\mathbb{Y}_s = \left( \mathbb{W}_{k_0,k_0+s}^{AP_{t_0}} \right)_{k \in S_{t_0+s}^{AP_{t_0}}}$. By letting $t = t_0$ in Lemma F.13, we obtain that $(\mathbb{Y}_s)_{s \geq 0}$ is a homogeneous Markov chain, whose transition kernel is $\mathbb{K}_0$, which is independent of the sequential procedure $\mathbb{T}^A$, $t_0$, and the information filtration $\mathcal{F}^A_{t_0}$. For the rest of the proof, according to Definition 6 and Fact 3, it is sufficient to show that $\mathbb{K}_0$ is stochastically monotone. That is, $\mathbb{K}_0(\mathbb{W}, \cdot) \leq_{st} \mathbb{K}_0(\mathbb{W}', \cdot)$ for any $\mathbb{W}, \mathbb{W}' \in \mathcal{S}_0$ with $\mathbb{W} \leq \mathbb{W}'$.

Thus, it is sufficient to show that for all $\mathbb{W} \leq \mathbb{W}'$ there exists a coupling $(\hat{\mathbb{V}}, \hat{\mathbb{V}}')$ such that $\hat{\mathbb{V}} \sim \mathbb{K}_0(\mathbb{W}, \cdot)$, $\hat{\mathbb{V}}' \sim \mathbb{K}_0(\mathbb{W}', \cdot)$ and $\hat{\mathbb{V}} \leq \hat{\mathbb{V}}'$ a.s. In what follows, we construct such a coupling.

For $\mathbb{W} \leq \mathbb{W}'$ with $\mathbb{W}, \mathbb{W}' \in \mathcal{S}_0$, we know that $H_0(\mathbb{W}) \leq H_0(\mathbb{W}')$ by Lemma F.5. By the definition of the partial order, this implies that $\dim(H_0(\mathbb{W}')) \leq \dim(H_0(\mathbb{W}))$ and $H_0(\mathbb{W})_k \leq H_0(\mathbb{W}')_k$ for each $1 \leq k \leq \dim(H_0(\mathbb{W}'))$. By Lemma F.9 this further implies

$$K(H_0(\mathbb{W}), \cdot) \leq_{st} K(H_0(\mathbb{W}'), \cdot)$$
for $k = 1, \ldots, \dim(H_u(u'))$. Thus, by Strassen’s Theorem for random variables (Fact 1), this implies that there exists a coupling $(\tilde{Z}_k, \tilde{Z}_k')$ such that

$$\tilde{Z}_k \sim K(H_u(u_k), \cdot), \tilde{Z}_k' \sim K(H_u(u'_k), \cdot), \text{ and } \tilde{Z}_k \leq \tilde{Z}_k' \text{ a.s.}$$

for $k = 1, \ldots, \dim(H_u(u'))$. In addition, we choose the coupling so that $(\tilde{Z}_k, \tilde{Z}_k')$ are independent for different $k$. For $\dim(H_u(u')) < k \leq \dim(H_u(u))$, we construct $\tilde{Z}_k \sim K(H_u(u_k), \cdot)$ so that $\tilde{Z}_k$’s are independent for different $k$. Let $\tilde{Z} = (\tilde{Z}_1, \cdots, \tilde{Z}_{\dim(H_u(u))})$ and $\tilde{Z}' = (\tilde{Z}_1', \cdots, \tilde{Z}'_{\dim(H_u(u'))})$.

For this coupling, it is easy to verify

$$\dim(\tilde{Z}) \geq \dim(\tilde{Z}') \text{ and } \tilde{Z}_k \leq \tilde{Z}_k' \text{ for } 1 \leq k \leq \dim(\tilde{Z}') \text{ a.s.}$$

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

$$\hat{V} \leq \hat{V}' \text{ a.s.}$$

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

$$\hat{V} \sim \mathbb{K}_o(u, \cdot) \text{ and } \hat{V}' \sim \mathbb{K}_o(u', \cdot).$$

Therefore,

$$\mathbb{K}_o(u, \cdot) \leq_{st} \mathbb{K}_o(u', \cdot).$$

\[\square\]

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

**Lemma F.1.** $(S_o, \preceq)$ is a partially ordered space. In addition, $S_o$ is a polish space equipped with the metric

$$d(u, v) = \begin{cases} \max_{1 \leq m \leq \dim(u)} |u_m - v_m| & \text{if } \dim(u) = \dim(v) \\ 0 & \text{if } u = v = \emptyset \\ 2 & \text{if } \dim(u) \neq \dim(v) \end{cases}$$

for $u, v \in S_o$.

**Proof of Lemma F.1.** First, $S_o$ is the union of polish spaces \{u = (u_1, \cdots, u_m) : 0 \leq u_1 \leq \cdots \leq u_m \leq 1\} and \{\emptyset\}. Thus, it is also a polish space. Second, it is straightforward to
verify that $d(u, v)$ is a metric defined over $S_o$.

Now, we verify that the partial order relationship $\preceq$ is closed over $S_o$. To see this, let $u, v \in S_o$ satisfying $u \npreceq v$. There are two cases: 1) $\dim(u) < \dim(v)$, or 2) $\dim(u) \geq \dim(v)$ and there exists $m \in \{1, \cdots, \dim(v)\}$ such that $u_m > v_m$. Let $B_d(u, \delta)$ and $B_d(v, \delta)$ be d-balls centering at $u$ and $v$ with $\delta$ 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 $u' \in B_d(u, \delta)$ and $v' \in B_d(v, \delta)$, we have $u' \npreceq v'$. That is, the partial order relationship $\preceq$ is closed over $S_o$.

**Lemma F.2.** If we input $(W_{k,t})_{k \in S_t} = u$ and an index set $S_t$ with $|S_t| = \dim(u)$ in Algorithm 7, then the output $S_{t+1}$ satisfies

$$|S_{t+1}| = I_o(u) \text{ and } [(W_{k,t})_{k \in S_{t+1}}] = H_o([u]).$$

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

**Lemma F.3.** $T = (T_1, \cdots, T_K) \in T_\alpha$ if and only if

$$T \in T \text{ and } \sum_{k=1}^{K} 1(T_k > t)W_{k,t} \leq \alpha \sum_{k=1}^{K} 1(T_k > t) \text{ for } t = 0, 1, 2, \cdots$$

The above expression is equivalent to

$$S_{t+1} \text{ is } F_t \text{ measurable }, \ S_{t+1} \subset S_t, \sum_{k \in S_{t+1}} W_{k,t} \leq \alpha \cdot |S_{t+1}|$$

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

**Proof of Lemma F.3.** By definition and the $F_t$ measurability of $S_{t+1}$,

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

Thus, $T \in T_\alpha$ if and only if

$$\frac{\sum_{k \in S_{t+1}} W_{k,t}}{|S_{t+1}| \vee 1} \leq \alpha \text{ a.s.},$$
which is equivalent to
\[ \sum_{k \in S_{t+1}} W_{k,t} \leq \alpha|S_{t+1}| \quad \text{a.s.,} \]
for every \( t \).

**Lemma F.4.** Let \( u = (u_1, \ldots, u_m) \in S_u \) with \( \dim(u) = m \geq 1 \). Let \( k_1, \ldots, k_l \in \{1, \ldots, m\} \) be distinct and satisfy
\[ \sum_{i=1}^{l} u_{k_i} \leq \alpha l. \]
Then, \( H_o([u]) \leq [(u_{k_1}, \ldots, u_{k_l})] \). Moreover, if \( H_o([u]) = \emptyset \), then for any \( S \subset \{1, \ldots, m\} \) with \( |S| \geq 1 \), \( \sum_{i \in S} u_i > \alpha|S| \).

**Proof of Lemma F.4.** We first prove the ‘Moreover’ part of the lemma by contradiction. If on the contrary \( H_o([u]) = \emptyset \) and there exists a non-empty set \( S \subset \{1, \ldots, m\} \) such that \( \sum_{i \in S} u_i \leq \alpha|S| \), then there exists \( i \in S \) such that \( u_i \leq \alpha \). This further implies \( [u]_1 \leq u_i \leq \alpha \) and \( I_o([u]) \geq 1 \), which contracts with the assumption \( H_o([u]) = \emptyset \).

We proceed to the proof of the rest of the lemma. We first prove that \( l \) in the lemma satisfies \( l \leq I_o([u]) \). To see this, recall that \( ([u]_1, \ldots, [u]_m) \) is the order statistic of \( (u_1, \ldots, u_m) \).

Thus,
\[ \sum_{i=1}^{l} [u]_i \leq \sum_{i=1}^{l} u_{k_i} \leq \alpha l. \tag{F.3} \]
Recall \( I_o([u]) = \text{sup}\{n : \sum_{i=1}^{n} [u]_i \leq \alpha n, n \in \{0, \ldots, m\}\} \). Thus, (F.3) implies \( l \leq I_o([u]) \).

Next, we prove that \( H_o([u]) \leq [(u_{k_1}, \ldots, u_{k_l})] \). Without loss of generality, assume \( u_{k_1}, \ldots, u_{k_l} \) are ordered. That is, \( u_{k_1} \leq \cdots \leq u_{k_l} \) and \( [(u_{k_1}, \ldots, u_{k_l})] = (u_{k_1}, \ldots, u_{k_l}) \). Then, according to the definition of the order statistic \([u]_i \), we have \([u]_i \leq u_{k_i} \) for \( i = 1, \ldots, l \).

Recall \( H_o([u]) = ([u]_1, \ldots, [u]_{I_o(u)}) \). This implies \( H_o([u]) \leq [(u_{k_1}, \ldots, u_{k_l})] \).

**Lemma F.5.** For any \( u \preceq v \in S_o \), \( H_o(u) \preceq H_o(v) \). That is, the mapping \( H_o(u) \) is increasing in \( u \).

**Proof of Lemma F.5.** If \( v = \emptyset \), then \( H_o(v) = \emptyset \) and \( H_o(u) \preceq \emptyset = H_o(v) \) by the definition of the partial order. In the rest of the proof we assume \( \dim(v) \geq 1 \) and \( v = (v_1, \ldots, v_{\dim(v)}) \). As we assumed \( u \preceq v \), this implies \( \dim(u) \geq \dim(v) \geq 1 \). We further denote \( u = (u_1, \ldots, u_{\dim(u)}) \).

We first show that if \( \sum_{i=1}^{L+1} v_i \leq \alpha(L + 1) \) for some \( L \), then \( \sum_{i=1}^{L} v_i \leq \alpha L \). That is, \( (\sum_{i=1}^{L} v_i)/L \) is increasing in \( L \). To see this, consider two cases. If \( v_{L+1} \leq \alpha \), then \( v_1 \leq \cdots \leq v_L \leq \alpha \) and thus \( \sum_{i=1}^{L} v_i \leq \alpha L \). If \( v_{L+1} > \alpha \), then \( \sum_{i=1}^{L} v_i \leq \sum_{i=1}^{L+1} v_i - \alpha \leq \alpha L \). This result implies that \( \sum_{i=1}^{L} v_i \leq \alpha L \) for all \( 1 \leq L \leq I_o(v) \).
Now we show that $I_o(u) \geq I_o(v)$ by contradiction. If on the contrary $I_o(u) < I_o(v)$, then $I_o(u) + 1 \leq I_o(v) \leq \dim(v)$ and

$$\sum_{i=1}^{I_o(u)+1} u_i \leq \sum_{i=1}^{I_o(u)+1} v_i \leq \alpha(I_o(u) + 1).$$

This contradicts with the definition of $I_o(u)$. Therefore, $I_o(u) \geq I_o(v)$.

We proceed to showing $H_o(u) \leq H_o(v)$. By the definition of $H_o$, we have $H_o(u) = (u_1, \cdots, u_{I_o(u)})$ and $H_o(v) = (v_1, \cdots, v_{I_o(v)})$. Since we assume $u \leq v$, we have $u_i \leq v_i$ for all $i = 1, \cdots, I_o(v)$. This shows that $H_o(u) \leq H_o(v)$.

Lemma F.6. Let $p(x)$ and $q(x)$ be two density functions with respect to some baseline measure $\mu$ 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 $Z_\delta$ be a random variable with the density function $\delta q + (1 - \delta)p$ and $L_\delta = L(Z_\delta)$. Then, for $0 \leq \delta_1 < \delta_2 \leq 1$, we have

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

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

$$\mathbb{E}g(L_{\delta_2}) - \mathbb{E}g(L_{\delta_1}) = \mathbb{E}Z_{\sim q}(L(Z)) - \mathbb{E}Z_{\sim \delta_1 q + (1-\delta_1)p}g(L(Z))$$

$$= \delta_2 \mathbb{E}Z_{\sim q}g(L(Z)) + (1 - \delta_2)\mathbb{E}Z_{\sim p}g(L(Z))$$

$$- \{\delta_1 \mathbb{E}Z_{\sim q}g(L(Z)) + (1 - \delta_1)\mathbb{E}Z_{\sim p}g(L(Z))\}$$

$$= (\delta_2 - \delta_1) \{\mathbb{E}Z_{\sim q}(L(Z)) - \mathbb{E}Z_{\sim p}(L(Z))\}.$$

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

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

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

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

The last inequality in the above display is due to the fact that $(Y - 1)(g(Y) - g(1)) \geq 0$
for all increasing function $g$. We remark that it is also a special case of Harris inequality [Harris 1960].

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

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

(F.2)

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

**Proof of Lemma F.7.** First, it is easy to see that $\{V_{k,s}\}_{s \geq 0}$ 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 $(X_{k,r})_{s \leq r \leq t}.$ First, $\mathbb{P}(\tau_1 < 0|X_{1,1,0}) = \mathbb{P}(\tau_1 < 0) = 0 = V_0$. Thus, it is sufficient to verify the update rule for $V_{1,t}$. A direct calculation gives

$$\mathbb{P}(\tau_1 \leq t - 1|X_{1,1,t}) = \frac{\sum_{s=0}^{t-1} \mathbb{P}(\tau_1 = s) \prod_{r=1}^{s} p(X_{1,r}) \prod_{r=s+1}^{t} q(X_{1,r})}{\sum_{s=0}^{t-1} \mathbb{P}(\tau_1 = s) \prod_{r=1}^{s} p(X_{1,r}) \prod_{r=s+1}^{t} q(X_{1,r}) + \mathbb{P}(\tau_1 \geq t) \prod_{r=1}^{t} p(X_{1,r})}$$

$$= \frac{\sum_{s=0}^{t-1} \theta(1 - \theta)^s L_{1,(s+1):t}}{\sum_{s=0}^{t-1} \theta(1 - \theta)^s L_{1,(s+1):t} + (1 - \theta)^t}$$

$$= \frac{Q_{1,t}}{Q_{1,t} + (1 - \theta)^t}$$

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,

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

Note that

$$Q_{1,t+1} = \sum_{s=0}^{t} \theta(1 - \theta)^s L_{1,(s+1):t+1} = q(X_{1,t+1})/p(X_{1,t+1}) \{ \theta(1 - \theta)^t + Q_{1,t} \}.$$
Thus,
\[
\mathbb{P}(\tau_1 \leq t|X_{1,1:t+1}) = \frac{Q_{1,t+1}}{Q_{1,t+1} + (1 - \theta)^{t+1}}
\]
\[
= \frac{q(X_{1,t+1})/p(X_{1,t+1}) \{ \theta(1 - \theta)^t + Q_{1,t} \}}{q(X_{1,t+1})/p(X_{1,t+1}) \{ \theta(1 - \theta)^t + Q_{1,t} \} + (1 - \theta)^{t+1}}
\]
\[
= \frac{Q_{1,t+1}/p(X_{1,t+1})}{q(X_{1,t+1})/p(X_{1,t+1}) + (1 - \theta)/\{ \theta + (1 - \theta)^{-t}Q_{1,t} \}}
\]
\[
= \frac{q(X_{1,t+1})/p(X_{1,t+1}) + (1 - \theta)/\{ \theta + \mathbb{P}(\tau_1 \leq t-1|X_{1,1:t}) \}}{1 - \mathbb{P}(\tau_1 \leq t-1|X_{1,1:t})}.
\]

We complete the proof by simplifying the above result.

**Lemma F.8.** Assume model \( \mathcal{M}_s \) holds. Let \( \delta_{k,t} = \mathbb{P}(\tau_k \leq t|X_{k,1}, \ldots, X_{k,t}) \), 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
\[
\delta_{1,t} = \frac{\sum_{s=0}^{t-1} \mathbb{P}(\tau_1 = s) \prod_{r=1}^{s} p(X_{1,r}) \prod_{r=s+1}^{t} q(X_{1,r}) + \mathbb{P}(\tau_1 = t) \prod_{r=1}^{t} p(X_{1,r})}{\sum_{s=0}^{t-1} \mathbb{P}(\tau_1 = s) \prod_{r=1}^{s} p(X_{1,r}) \prod_{r=s+1}^{t} q(X_{1,r}) + \mathbb{P}(\tau_1 \geq t) \prod_{r=1}^{t} p(X_{1,r})}
\]
\[
= \frac{\sum_{s=0}^{t-1} \theta(1 - \theta)^s L_{1,(s+1):t} + \theta(1 - \theta)^t}{\sum_{s=0}^{t-1} \theta(1 - \theta)^s L_{1,(s+1):t} + (1 - \theta)^t}
\]
\[
= \frac{Q_{1,t} + \theta(1 - \theta)^t}{Q_{1,t} + (1 - \theta)^t}
\]
\[
= V_{1,t} + \theta(1 - V_{1,t})
\]
\[
= \theta + (1 - \theta)V_{1,t}.
\]

**Lemma F.9.** Under model \( \mathcal{M}_s \), the process \( \{V_{1,t}\}_{t \geq 0} \) 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 \( V_{1,0}, \ldots, V_{1,t} \). According to the change point model \( \mathcal{M}_s \), we know that \( X_{1,t+1} \) is conditionally independent
of \( V_{1,0}, \ldots, V_{1,t} \) given the event \( \{ \tau_1 \leq t \} \). That is, given \( V_{1,0}, \ldots, V_{1,t} \), the conditional density function of \( X_{1,t+1} \) is \( \delta_{1,t} q(x) + (1 - \delta_{1,t}) p(x) \), which depends on \( X_{1,1}, \ldots, X_{1,t} \) 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 \( V_{1,0}, \ldots, V_{1,t} \) only depends on \( V_{1,t} \). According to the iteration \( \text{(F.2)} \), this implies that the process \( \{ V_{1,t} \}_{t \geq 0} \) is a Markov process. Note that \( \delta_{1,t} \) and the iteration \( \text{(F.2)} \) 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 \( \text{(F.2)} \) 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 \( V_{1,t} \) given \( V_{1,t} = x \). In other words, \( V(x) \) has the density function \( K(x, \cdot) \).

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

Because \( \hat{L} \leq \hat{L}' \) and \( x \leq x' \),

\[
\hat{V} = \frac{\hat{L}}{\hat{L} + (1 - \theta)(1 - x)/(\theta + (1 - \theta)x)} \leq \frac{\hat{L}'}{\hat{L}' + (1 - \theta)(1 - x')/(\theta + (1 - \theta)x')} = \hat{V}' \quad \text{a.s.}
\]

That is, \( \hat{V} \leq \hat{V}' \) a.s., and \( (\hat{V}, \hat{V}') \) is a coupling of \( (V(x), V(x')) \). Thus, \( V(x) \leq_{st} V(x') \) and so is \( K(x, \cdot) \leq_{st} K(x', \cdot) \). 

\( \square \)
Lemma F.10. For any $t \geq 1$ and $\mathbb{T}, [W^{A}_{S_{t+1},t+1}]$ is conditionally independent of $F^{A}_{t}$ given $[W^{A}_{S_{t+1}}]$. Moreover, the conditional density of $[W^{A}_{S_{t+1},t+1}]$ at $v$ given $[W^{A}_{S_{t+1}}] = u \in S_{o}$ is

$$\mathbb{K}_{a}(u, v) := \begin{cases} 
\sum_{\pi \in P_{dim(u)}} \prod_{l=1}^{dim(u)} K(u_{l}, v_{\pi(l)}) & \text{if } dim(u) = dim(v) \geq 1, \\
1 & \text{if } dim(u) = dim(v) = 0, \\
0 & \text{otherwise},
\end{cases}$$

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

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

We start with deriving the conditional density of $W^{A}_{S_{t+1},t+1}$ at $v \in S_{o}$ given $x_{S_{t+1}} = x_{1}, S_{1} = s_{1}, \cdots, X_{S_{t+1}} = x_{t}, S_{t} = s_{t}, S_{t+1} = s_{t+1}$ and $W^{A}_{S_{t+1}} = u$ for some $x_{1}, \cdots, x_{t}$ and $s_{1}, \cdots, s_{t+1}$, and $u \in S_{o}$. Clearly, the conditional density is 0 when $dim(u) \neq dim(v)$, and is arbitrary when $dim(u) = dim(v)$ (the density of the random variable being conditional on is zero). Thus, we will focus on the case where $dim(u) = dim(v) = dim(v) = |s_{t+1}| = m$ for some $m \in \{1, \cdots, K\}$, and we will write $u = (u_{1}, \cdots, u_{m})$ and $v = (v_{1}, \cdots, v_{m})$.

Note that given $S_{t+1} = s_{t+1}, W^{A}_{S_{t+1},t+1} = u$, $W^{A}_{S_{t+1},t+1}$’s are independent for different $k \in s_{t+1}$. Moreover, given $S_{t+1} = s_{t+1}, W^{A}_{S_{t+1},t+1} = u$, $W^{A}_{S_{t+1},t+1}$ is the same as $V_{k,t+1}$ (defined in (F.2)) for $k \in s_{t+1}$, and is independent of $X_{S_{t+1}} = x_{1}, S_{t+1} = s_{1}, \cdots, X_{S_{t}} = x_{t}$ and $S_{t} = s_{t}$. Thus, $W^{A}_{S_{t+1},t+1}$ is conditionally independent of $F_{t}^{A}$ given $S_{t+1} = s_{t+1}, W^{A}_{S_{t+1},t+1} = u$, and its conditional density (by Lemma F.9) is

$$\prod_{l=1}^{m} K(u_{l}, v_{l}),$$

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

$$\sum_{\pi \in P_{m}} \prod_{l=1}^{m} K(u_{l}, v_{\pi(l)}) = \sum_{\pi \in P_{m}} \prod_{l=1}^{m} K([u]_{l}, v_{\pi(l)}) = \mathbb{K}_{a}([u], v),$$

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

Lemma F.11. For each $u \in S_{o}$ with $dim(u) = m \geq 1$, generate an $S_{o}$-valued random
variable $V$ as follows,

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

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

In addition, if $m = 0$, we let $V = \emptyset$. Then, $V \sim \mathcal{K}_a(u, \cdot)$.

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

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

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

$$\sum_{\pi \in \mathcal{P}_m} \prod_{i=1}^{m} K(u_i, z_{\pi(i)}) = \mathcal{K}_a(u, z).$$

Lemma F.12. For $u, u' \in S_o$ with $u \preceq u'$, we have $\mathcal{K}_a(u, \cdot) \preceq_{st} \mathcal{K}_a(u', \cdot)$.

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

Let $Z = (Z_1, \ldots, Z_m) \sim \mathcal{K}_a(u, \cdot)$, $Z' = (Z'_1, \ldots, Z'_m)$, $V = [Z]$ and $V' = [Z']$. Then, $V \preceq V'$ a.s. On the other hand, by Lemma F.11, we have

$$V \sim \mathcal{K}_a(u, \cdot) \text{ and } V' \sim \mathcal{K}_a(u', \cdot),$$

and $V \preceq V'$ a.s. By Fact 2, the existence of such a coupling implies $\mathcal{K}_a(u, \cdot) \preceq_{st} \mathcal{K}_a(u', \cdot)$. □

Lemma F.13. For any $t, s \geq 0$ and $\mathbb{T}^A$, $[W_{t+s+1}^{\text{AP}}, \ S_{t+s+1}^{\text{AP}, t+s+1}]$ is conditionally independent of $\mathcal{F}_{t+s}^{\text{AP}, t}$ given $[W_{s+t+s}^{\text{AP}, t}, S_{s+t+s}^{\text{AP}, t+s}]$. Moreover, the conditional density of $[W_{s+t+s}^{\text{AP}, t}, S_{s+t+s}^{\text{AP}, t+s}]$ at $v$ given
\[
[W_{\mathcal{AP}_{t+s}}^{t+s}] = u \in \mathcal{S}_o \text{ is}
\]

\[
\mathbb{K}_o(u, v) := \mathbb{K}_o(H_o(u), v) = \begin{cases} 
\sum_{\pi \in \mathcal{P}_m} \prod_{t=1}^{L_o(u)} K(H_o(u)_t, v_{\pi(t)}) & \text{if } \dim(v) = I_o(u) \geq 1 \\
1 & \text{if } \dim(v) = I_o(u) = 0 \\
0 & \text{otherwise,}
\end{cases}
\]

where \(\mathcal{P}_m\) denotes the set of all permutations over \(\{1, \cdots, m\}\).

**Proof of Lemma F.13** Apply Lemma F.10 by replacing \(T^A\) by \(T^{AP}\), and \(t\) by \(t+s\), we obtain that \([W_{\mathcal{AP}_{t+s}}^{t+s}]\) is conditionally independent of \(\mathcal{F}_{t+s}\) given \([W_{\mathcal{AP}_{t+s+1}}^{t+s+1}]\). On the other hand, according to the one-step update rule in Algorithm 1 and Lemma F.2, we can see that \([W_{\mathcal{AP}_{t+s}}^{t+s}] = H_o([W_{\mathcal{AP}_{t+s+1}}^{t+s+1}]).\) Therefore, we further obtain that \([W_{\mathcal{AP}_{t+s+1}}^{t+s+1}]\) is conditionally independent of \(\mathcal{F}_{t+s}\) given \([W_{\mathcal{AP}_{t+s}}^{t+s}]\).

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

\[
\mathbb{K}_o(H_o(u), v) = \sum_{\pi \in \mathcal{P}_m} \prod_{t=1}^{L_o(u)} K(H_o(u)_t, v_{\pi(t)}) = \mathbb{K}_o(u, v).
\]

This completes the proof of the lemma.

**Lemma F.14.** For each \(u \in \mathcal{S}_o\) and \(m = I_o(u)\), generate an \(\mathcal{S}_o\)-valued random variable \(V\) as follows,

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

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

**Proof of Lemma F.14** The lemma is a direct application of Lemma F.11 and \(\mathbb{K}_o(u, v) = \mathbb{K}_o(H_o(u), v)\). □
G Proof of Lemma 1 and Propositions 1 - 3

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

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

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

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

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

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

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

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

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

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

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

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

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,
\[
\frac{\sum_{k \in S_{t+1}} W_{k,t}}{|S_{t+1}| \vee 1} = R_n \leq \alpha.
\]
Therefore, \( E\left( \frac{\sum_{k \in S_{t+1}} 1(1_{k < t})}{|S_{t+1}| \vee 1} | F_t \right) \leq \alpha. \)

\[\square\]

**Proposition 2.** Let \( T^* \) be defined in Algorithm 2. Then, \( T^* \in T_\alpha. \)

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

**Proposition 3.** Given LFNR level \( \alpha \) and information filtration \( F_t \), 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, \( |S_{t+1}| = I_0([W_{S,t}]) \) and \( [W_{S_{t+1},t}] = H_0([W_{S,t}]) \). There are two cases: 1) \( |S_{t+1}| = 0 \), and 2) \( |S_{t+1}| = n \geq 1 \). For the first case, \( [W_{S_{t+1},t}] = \emptyset \). Note that \( E\left( \frac{\sum_{k \in S_{t+1}} 1(1_{k < t})}{|S| \vee 1} | F_t \right) = \frac{\sum_{k \in S} W_{k,t}}{|S| \vee 1} \). By the ‘Moreover’ part of Lemma F.4, we can see that the only set \( S \) satisfying \( E\left( \frac{\sum_{k \in S} 1(1_{k < t})}{|S| \vee 1} | F_t \right) \leq \alpha \) is \( S = \emptyset \). That is \( |S| = 0 \). Thus, \( |S_{t+1}| \geq |S| \).

For the second case where \( |S_{t+1}| = n \geq 1 \) and any set \( |S| \) satisfying \( E\left( \frac{\sum_{k \in S} 1(1_{k < t})}{|S| \vee 1} | F_t \right) \leq \alpha \), we use Lemma F.4 again and obtain that \( [W_{S_{t+1},t}] = H_0([W_{S,t}]) \leq [W_{S,t}] \). This implies \( |S_{t+1}| = \dim([W_{S_{t+1},t}]) \geq \dim([W_{S,t}]) = |S| \). \[\square\]

**H Proof of Theorem 2 and Theorem 3**

**H.1 Proof of Theorem 2**

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

1. \( \lim_{K \to \infty} \hat{\lambda}_{K,t} = \lambda_t \) a.s., where \( \hat{\lambda}_{K,t} = \max \{ W_{k,t} : k \in S^*_{K,t+1} \} \) is the threshold used by \( T^*_K \).
2. \( \lim_{K \to \infty} \text{LFNR}_{t+1}(T^*_K) = \mathbb{E} \left( V_t \mid V_s \leq \lambda_s, 0 \leq s \leq t \right) \), a.s. Moreover,
\[
\mathbb{E} \left( V_t \mid V_s \leq \lambda_s, 0 \leq s \leq t \right) = \begin{cases} 
1 - (1 - \theta)^t, & t < \frac{\log(1 - \alpha)}{\log(1 - \theta)}, \\
\alpha, & t \geq \frac{\log(1 - \alpha)}{\log(1 - \theta)}. 
\end{cases}
\]

3. \( \lim_{K \to \infty} K^{-1}|S_{K,t+1}^*| = \mathbb{P} \left( V_1 \leq \lambda_1, \ldots, V_t \leq \lambda_t \right) \) 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 \( t \geq 1 \), \( (V_1, \ldots, V_t) \) has a continuous and strictly positive joint density function over \( (0, 1)^t \) (with respect to the Lebesgue measure).

2. For any \( (v_1, \ldots, v_t) \in (0, 1)^t \), \( \mathbb{P}(V_1 \leq v_1, \ldots, V_t \leq v_t) > 0. \)

3. For any \( (v_1, \ldots, v_t) \in (0, 1)^t \), the conditional distribution of \( V_{t+1} \) given \( V_1 \leq v_1, \ldots, V_t \leq v_t \) 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 \( (V_1, \ldots, V_{t_0}) \), 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 \( \Lambda_0(Q) = 1 \) for every \( Q \). Then, for \( t \geq 1 \), define
\[
D_t(\lambda, Q) = Q( V_t \leq \lambda, V_{t-1} \leq \Lambda_{t-1}(Q)),
\]
\[
N_t(\lambda, Q) = \mathbb{E}_Q [V_t \mathbb{I} \{ V_t \leq \lambda, V_{t-1} \leq \Lambda_{t-1}(Q) \}],
\]
\[
G_t(\lambda, Q) = \frac{N_t(\lambda, Q)}{D_t(\lambda, Q)} = \mathbb{E}_Q [V_t \mid V_t \leq \lambda, V_{t-1} \leq \Lambda_{t-1}(Q)],
\]
and
\[
\Lambda_t(Q) = \sup \{ \lambda : G_t(\lambda, Q) \leq \alpha \text{ and } \lambda \in [0, 1] \}.
\]

In the above equations, we use notation \( \mathbf{V}_t = (V_1, \ldots, V_t) \) and \( \Lambda_t(Q) = (\Lambda_1(Q), \ldots, \Lambda_t(Q)). \) In addition, \( \{ \mathbf{V}_t \leq \Lambda_t(Q) \} \) denotes the event \( \{ V_1 \leq \Lambda_1(Q), \ldots, V_t \leq \Lambda_t(Q) \} \).

The next lemma, whose proof is given in Section H.3, provides results about the above mappings. For two probability measures \( Q \) and \( Q' \) for a \( t_0 \)-dimensional random vector \( \mathbf{V}_t \), their sup-norm is defined as \( \| Q - Q' \|_\infty = \sup_{v \in \mathbb{R}^{t_0}} |Q(\mathbf{V}_t \leq v) - Q'(\mathbf{V}_t \leq v)| \). Then, we say a mapping \( f(Q') \) is sup-norm continuous at \( Q' = Q \) if \( \lim_{\epsilon \to 0} \sup_{Q' : \| Q' - Q \|_\infty < \epsilon} |f(Q') - f(Q)| = 0. \)
Lemma H.2. For each $1 \leq t \leq t_0$, we have the following results.

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

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

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

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

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

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

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-Chervonenkis class and thus, $\lim_{K \to \infty} \sup_{C \in \mathcal{C}} |\mathbb{P}_K(V_{t_0} \in C) - \mathbb{P}^*(V_{t_0} \in C)| = 0$ a.s. (see, e.g., Shorack and Wellner (2009)). In other words,

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

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

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

That is, $\lim_{K \to \infty} \hat{\lambda}_{K,t} = \lambda_t$ 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

$$
J_t(Q) = \mathbb{E}_Q(V_t \mathbb{1}_{\{V_t \leq \Lambda_t(Q)\}}) \text{ and } H_t(Q) = Q(V_t \leq \Lambda_t(Q)).
$$

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

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

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

and

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

Note that

\[ H_t(\mathbb{P}_K) = K^{-1}|S_{t+1}^*| \text{ and } J_t(\mathbb{P}_K) = \mathbb{E}(\text{FNP}_{t+1}(T)|\mathcal{F}_t). \quad (H.4) \]

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

\[ G_t(\lambda, \mathbb{P}^*) = \mathbb{E}[V_t|V_t \leq \lambda, V_{t-1} \leq \Lambda_{t-1}(\mathbb{P}^*)] = \mathbb{E}[V_t|V_t \leq \lambda]. \]

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

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

This completes the induction. As a result, for \( 1 \leq t \leq L \), \( \mathbb{E}[V_t|V_t \leq \lambda_t, V_{t-1} \leq \Lambda_{t-1}] = G_t(1, \mathbb{P}^*) = 1 - (1-\theta)^t \).

We proceed to the proof of (4.2) for \( t > L \). Note that \( N_t(\lambda, \mathbb{P}^*) \) and \( D_t(\lambda, \mathbb{P}^*) \) are continuous in \( \lambda \in (0, 1) \) (note that \( V_t \) 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

\[ G_t(1, \mathbb{P}^*) \geq \alpha. \quad (H.5) \]

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

\[ G_{|L|+1}(1, \mathbb{P}^*) = \mathbb{E}(V_{|L|+1}) = 1 - (1-\theta)^{|L|+1} > \alpha. \]

This proves the base case for the induction.
Assume that for $1 \leq s \leq t - 1$, $G_s(1, \mathbb{P}^*) > \alpha$. Then,

$$G_t(1, \mathbb{P}^*) = \mathbb{E}[V_t | V_{t-1} \leq \lambda_{t-1}] = \mathbb{E}[\mathbb{E}(V_t | X_{1:t-1}) | V_{t-1} \leq \lambda_{t-1}],$$

(\text{H.6})

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

$$\mathbb{E}(V_t | X_{1:t-1}) = \mathbb{E}[\mathbb{P}(\tau_1 < t | X_{1:t-1}) | X_{1:t-1}] = \mathbb{P}(\tau_1 < t | X_{1:t-1}) = \delta_{1:t-1} = \theta + (1 - \theta)V_{t-1},$$

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

$$G_t(1, \mathbb{P}^*) = \mathbb{E}[\theta + (1 - \theta)V_{t-1} | V_{t-1} \leq \lambda_{t-1}] = \theta + (1 - \theta)\mathbb{E}[V_{t-1} | V_{t-1} \leq \lambda_{t-1}].$$

By induction assumption, we have

$$\mathbb{E}[V_{t-1} | V_{t-1} \leq \lambda_{t-1}] = \alpha.$$

The above two equations give

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

This completes our proof.

\textbf{Remark H.1.} A key observation in the above proof is that $\hat{\lambda}_{K,t} = \Lambda_t(\mathbb{P}_K)$ 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 $\{V_{k,t}\}_{1 \leq t \leq t_0}$. Thus, to show that $\hat{\lambda}_{K,t}$ converges to $\lambda_t$ (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 $K \rightarrow \infty$. In the proof, the above heuristics are justified through Vapnik-\v{C}ervonenkis (VC) theory. In particular, as a standard result in VC theory, the empirical measure converges to the underlying measure uniformly over the set $C = \{(-\infty, x] : x \in \mathbb{R}^{t_0}\}$. That is, $\mathbb{P}_K$ converges to $\mathbb{P}^*$ in $\|\cdot\|_\infty$ norm almost surely. The supporting lemma (Lemma \textbf{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 \cite{shorack2009empirical} 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 \( \tau_0 \sim \text{Geom}(\theta) \), and Assumption A2 holds. Let

\[
W_t = \mathbb{P}\left( \tau_0 < \left| X_{k,s}, 1 \leq k \leq K, 1 \leq s \leq t \right| \right),
\]

and

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

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

1. \( \lim_{K \to \infty} (T - \tau_0) = 1 \) a.s.,
2. \( \lim_{K \to \infty} \text{LFNR}_{t+1}(T^*_K) = 0 \) a.s.,
3. \( \lim_{K \to \infty} K^{-1}|S^*_K| \geq 1(\tau_0 > t) \) 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 \( T^* \), if \( W_{1,t} \leq \alpha \), then \( \sum_{k \in S_t} W_{t,k} \leq \alpha|S_t| \), 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 \( S \neq \emptyset \), \( \sum_{k \in S_t} W_{k,t+1} = W_{k,t+1}|S| > \alpha|S| \), and thus \( S_{t+1} = \emptyset \). Thus, \( T^* = (T, \cdots, T) \). In other words, \( S_t = \{1, \cdots, K\} \) for \( t \leq T \) and \( S_t = \emptyset \) for \( t > T \).

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

\[
\tilde{W}_{k,t} = \frac{\sum_{s=0}^{t-1} \theta(1-\theta)^s \prod_{r=s+1}^{t} \prod_{k=1}^{K} q(X_{k,r})/p(X_{k,r})}{\sum_{s=0}^{t-1} \theta(1-\theta)^s \prod_{r=s+1}^{t} \prod_{k=1}^{K} q(X_{k,r})/p(X_{k,r}) + (1-\theta)^t}.
\]

Let \( l_{k,s,t} = \sum_{r=s+1}^{t} \log(q(X_{k,r})/p(X_{k,r})) \).
For each \( u \in \mathbb{Z}_+ \cup \{0\} \), let \( A_u = \{ \tau_0 = u \} \). By the strong law of large numbers, under Assumption A2,
\[
P \left( \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} l_{k,s,t} = \mathbb{E}(l_{1,s,t}|\tau_0 = u) | A_u \right) = 1 \tag{H.8}
\]
for each \( s, t, u \in \mathbb{Z}_+ \cup \{0\} \) with \( s < t \). In particular,
\[
\mathbb{E}(l_{1,s,t}|\tau_0 = u) = \begin{cases} 
-(t-s)\mathbb{E}_{Z_1 \sim p} \log(p(Z_1)/q(Z_1)) & \text{if } t \leq u \\
\mathbb{E}_{Z_2 \sim q} \log(q(Z_2)/p(Z_2)) & \text{if } t = u + 1 \text{ and } s = u.
\end{cases}
\]
Thus, for each \( s < t \leq u \) we have
\[
P \left( \lim_{K \to \infty} \sum_{k=1}^{K} l_{k,s,t} = -\infty | A_u \right) = 1, \tag{H.9}
\]
and for \( t = u + 1 = s + 1 \),
\[
P \left( \lim_{K \to \infty} \sum_{k=1}^{K} l_{k,s,t} = \infty | A_u \right) = 1.
\]
According to (H.7), (H.8) and (H.9), we have that for each \( t \leq u \)
\[
P \left( \lim_{K \to \infty} \tilde{W}_{k,t} = 0 | A_u \right) = 1.
\]
Moreover, for \( t \geq u + 1 \),
\[
P \left( \lim_{K \to \infty} \tilde{W}_{k,t} = 1 | A_u \right) = 1.
\]
Combining the above two equations for different \( u \in \mathbb{Z}_+ \cup \{0\} \), we arrive at
\[
P \left( \lim_{K \to \infty} \tilde{W}_{1,t} = 1(t \geq \tau_0 + 1) \right) = 1.
\]
In other words,
\[
\lim_{K \to \infty} \tilde{W}_{1,t} = 1(t \geq \tau_0 + 1) \text{ a.s.}
\]
Now we turn to the analysis of \( W_{k,t} \) and \( S_t \) for the proposed procedure. Let \( \omega \) be a sample path with \( \lim_{K \to \infty} \tilde{W}_{k,t}(\omega) = 1(t \geq \tau_0(\omega) + 1) \) for all \( t = 1, 2, \cdots \). Then, there exists \( K_0(\omega) \) large enough such that \( \tilde{W}_{1,t}(\omega) < \alpha \) for \( t \leq \tau_0(\omega) \) and \( \tilde{W}_{1,\tau_0(\omega)+1}(\omega) > \alpha \) for all \( K \geq K_0(\omega) \). Then, we have \( T(\omega) = \inf\{t : \tilde{W}_{1,t}(\omega) > \alpha\} = \tau_0(\omega) + 1 \). Note that the set of such sample
path $\omega$ has a probability of one. Thus,
\[ \lim_{K \to \infty} (T - \tau_0) = 1 \] and
\[ \lim_{K \to \infty} W_{k,t} = 0 \text{ for } t \leq \tau_0 \text{ a.s.} \]
This proves the first statement of the theorem. For the second statement, we have
\[ \lim_{K \to \infty} \mathbb{E}(FNP_{t+1}(\mathbb{T}^*))|\mathcal{F}_t) = \lim_{K \to \infty} \frac{\sum_{k=1}^{K} 1(T > t)W_{k,t}}{\left(\sum_{k=1}^{K} 1(T > t)\right)} = \lim_{K \to \infty} W_{k,t} 1(T > t) = 0 \text{ a.s.} \]
For the third statement, we have
\[ \lim_{K \to \infty} K^{-1}|S_{t+1}| = \lim_{K \to \infty} 1(T > t) = 1(\tau_0 \geq t) \text{ a.s.} \]

\[ \Box \]

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

Lemma H.1. Under model $M_s$ and Assumption A1, we have the following results.

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

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

3. For any $(v_1, \cdots, v_t) \in (0,1)^t$, the conditional distribution of $V_{t+1}$ given $V_1 \leq v_1, \cdots, V_t \leq v_t$ 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 $M_s$, $X_{1,1}$ 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
\[ f_{V_1}(v) = \frac{c}{(1 - v)^2} \left\{ (1 - \theta)f_1 \left( \frac{cv}{1 - v} \right) + \theta f_2 \left( \frac{cv}{1 - v} \right) \right\}, \quad (H.10) \]
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 $V_{t+1} = \frac{q(X_{t+1})/p(X_{t+1})}{(1-\theta)(1-V_1) / \theta (1-\theta) V_1 + q(X_{t+1}) / p(X_{t+1})}$. 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

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

where we define $c_t = \frac{(1-\theta)(1-v_t)}{\theta(1-v_t)} > 0$ and $\theta_t = \mathbb{P}(\tau_1 \leq t | V_1 = v_1, \cdots, V_t = v_t) = v_t(1-\theta) + \theta \in (0,1)$. It is easy to see that both $c_t$ and $\theta_t$ 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.

\[\square\]

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

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

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

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

**Proof of Lemma [H.2]** For $t = 0, 1, \cdots$ and $\lambda < \lambda'$, let $\tilde{V}$ be a random variable following the same distribution as $V_t|V_{t-1} \leq \Lambda_{t-1}(Q)$. Then, by the definition of conditional expectation, we have

$$G_t(\lambda', Q) - G_t(\lambda, Q) = Z^{-1} \left[ \mathbb{E}_Q \left( \tilde{V} \mathbb{1} \{ \tilde{V} \leq \lambda' \} \right) Q \left( \tilde{V} \leq \lambda \right) - \mathbb{E}_Q \left( \tilde{V} \mathbb{1} \{ \tilde{V} \leq \lambda' \} \right) Q \left( \tilde{V} \leq \lambda' \right) \right]$$

$$= Z^{-1} \left[ \mathbb{E}_Q \left( \tilde{V} \mathbb{1} \{ \lambda < \tilde{V} \leq \lambda' \} \right) Q \left( \tilde{V} \leq \lambda \right) - \mathbb{E}_Q \left( \tilde{V} \mathbb{1} \{ \tilde{V} \leq \lambda' \} \right) Q \left( \lambda < \tilde{V} \leq \lambda' \right) \right]$$

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

\[ G_t(\lambda', \mathbb{Q}) - G_t(\lambda, \mathbb{Q}) \\
= Z^{-1} \left[ \mathbb{E}_Q \left( \tilde{V} \mathbb{1} \left\{ \lambda < \tilde{V}' \leq \lambda', \tilde{V} \leq \lambda \right\} \right) - \mathbb{E}_Q \left( \tilde{V} \mathbb{1} \left\{ \lambda < \tilde{V}' \leq \lambda', \tilde{V} \leq \lambda \right\} \right) \right] \]  \hspace{1cm} (H.11)

\[ = Z^{-1} \mathbb{E}_Q \left[ \left( \tilde{V}' - \tilde{V} \right) \mathbb{1} \left\{ \lambda < \tilde{V}' \leq \lambda', \tilde{V} \leq \lambda \right\} \right] , \]

Because \( \left( \tilde{V}' - \tilde{V} \right) \mathbb{1} \left\{ \lambda < \tilde{V}' \leq \lambda', \tilde{V} \leq \lambda \right\} \geq 0, G_t(\lambda', \mathbb{Q}) - G_t(\lambda, \mathbb{Q}) \geq 0 \) 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 \( t = 1, 2, \ldots, t_0 \):

\[ G_t(\lambda, \mathbb{P}^*) \text{ is strictly increasing in } \lambda; \]  \hspace{1cm} (H.12)

\[ D_t(\lambda, \mathbb{Q}), N_t(\lambda, \mathbb{Q}), \text{ and } G_t(\lambda, \mathbb{Q}) \text{ are sup-norm continuous at } \mathbb{Q} = \mathbb{P}^*; \]  \hspace{1cm} (H.13)

\[ \Lambda_t(\mathbb{Q}) \text{ is sup-norm continuous at } \mathbb{Q} = \mathbb{P}^*. \]  \hspace{1cm} (H.14)

We start with the base case that \( t = 1 \). In this case, the conditional distribution \( V_1|V_0 \leq \Lambda_0(\mathbb{Q}) \) 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, \( \mathbb{P}^* \left( \left( \tilde{V}' - \tilde{V} \right) \mathbb{1} \left\{ \lambda < \tilde{V}' \leq \lambda', \tilde{V} \leq \lambda \right\} \geq 0 \right) > 0 \) for \( \tilde{V} \) and \( \tilde{V}' \) are identically distributed as \( V_1 \). According to (H.11), \( G_1(\lambda', \mathbb{P}^*) - G_1(\lambda, \mathbb{P}^*) \geq 0 \). That is, \( G_1(\lambda, \mathbb{P}^*) \) is strictly increasing in \( \lambda \). 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 \( t = 1, 2, \ldots, s - 1 \). We proceed to prove these equations for \( t = s \). First, note that \( V_t|V_{t-1} \leq \Lambda_{t-1}(\mathbb{P}^*) \) 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, \( \Lambda_1(\mathbb{Q}), \ldots, \Lambda_{s-1}(\mathbb{Q}) \) is sup-norm continuous in \( \mathbb{Q} \) at \( \mathbb{Q} = \mathbb{P}^* \). This implies that \( (\lambda, \Lambda_{s-1}(\mathbb{Q})) \), a vector-valued mapping, is also sup-norm continuous in \( \mathbb{Q} \) at \( \mathbb{Q} = \mathbb{P}^* \). On the other hand, \( (\lambda, \Lambda_{s-1}(\mathbb{P}^*)) \in (0, 1]^s \) by induction assumptions, and \( V_t \) has a continuous joint probability cumulative function at \( (\lambda, \Lambda_{s-1}(\mathbb{P}^*)) \) (by Lemma [H.1]). Combining these results, we can see that \( \mathbb{P}^*(V_s \leq \lambda, V_{s-1} \leq \Lambda_{s-1}(\mathbb{Q})) \) is sup-norm continuous at \( \mathbb{Q} = \mathbb{P}^* \).
Now we analyze the mapping \( D_s(\lambda, Q) = Q(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)) \).

\[
\begin{align*}
|D_s(\lambda, Q) - D_s(\lambda, \mathbb{P}^*)| &= |Q(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)) - \mathbb{P}^*(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*))| \\
&\leq |Q(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)) - \mathbb{P}^*(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q))| \\
&\quad + |\mathbb{P}^*(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)) - \mathbb{P}^*(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*))| \\
&\leq \|Q - \mathbb{P}^*\|_{\infty} \\
&\quad + \lim_{|Q - \mathbb{P}^*|_{\infty} \to 0} |\mathbb{P}^*(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)) - \mathbb{P}^*(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*))|.
\end{align*}
\]

Therefore,

\[
\limsup_{\|Q - \mathbb{P}^*\|_{\infty} \to 0} |D_s(\lambda, Q) - D_s(\lambda, \mathbb{P}^*)| = \lim_{\|Q - \mathbb{P}^*\|_{\infty} \to 0} \|Q - \mathbb{P}^*\|_{\infty} \\
+ \lim_{|Q - \mathbb{P}^*|_{\infty} \to 0} |\mathbb{P}^*(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)) - \mathbb{P}^*(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*))|.
\]

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

We proceed to the analysis of \( N_s(\lambda, Q) \). We have

\[
N_s(\lambda, Q) = \mathbb{E}_Q[V_s 1\{V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)\}]
= \mathbb{E}_Q\left[\int_0^1 1\{r < V_s\}dr 1\{V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)\}\right]
= \int_0^1 Q(r < V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)) dr
= Q(V_s \leq \lambda, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q))
\quad - \int_0^\lambda Q(V_s \leq r, \mathbf{V}_{s-1} \leq \Lambda_{s-1}(Q)) \, dr
\quad = D_s(\lambda, Q) - \int_0^\lambda D_s(r, Q) \, dr.
\]

We have already shown that the first term \( D_s(\lambda, Q) \) 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,

$$
\left| \int_0^\lambda D_s(r, Q)dr - \int_0^\lambda D_s(r, \mathbb{P}^*)dr \right|
\leq \int_0^\lambda \left| Q \left(V_s \leq r, V_{s-1} \leq \Lambda_{s-1}(Q)\right) - \mathbb{P}^* \left(V_s \leq r, V_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*)\right) \right| dr
+ \int_0^\lambda \left| \mathbb{P}^* \left(V_s \leq r, V_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*)\right) - \mathbb{P}^* \left(V_s \leq r, V_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*)\right) \right| dr
\leq \|Q - \mathbb{P}^*\|_\infty
+ \int_0^\lambda \left| \mathbb{P}^* \left(V_s \leq r, V_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*)\right) - \mathbb{P}^* \left(V_s \leq r, V_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*)\right) \right| dr.
\tag{H.16}
$$

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

$$
\sup_{|Q - \mathbb{P}^*| \leq \delta} \left| \mathbb{P}^* \left(V_s \leq r, V_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*)\right) - \mathbb{P}^* \left(V_s \leq r, V_{s-1} \leq \Lambda_{s-1}(\mathbb{P}^*)\right) \right| 
\leq \sup_{|v_s - v_s'| \leq \varepsilon, v_s, v_s' \in [0, 1]^s} \left| \mathbb{P}^* \left(V_s \leq v_s\right) - \mathbb{P}^* \left(V_s \leq v_s'\right) \right|.
\tag{H.17}
$$

By Lemma [H.1], $V_s$ has a continuous density function. Thus, its cumulative distribution function, $\mathbb{P}^* \left(V_s \leq v_s\right)$, 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 $\epsilon_1$ small enough, there is $\epsilon > 0$, such that

$$
\sup_{|v_s - v_s'| \leq \varepsilon, v_s, v_s' \in [0, 1]^s} \left| \mathbb{P}^* \left(V_s \leq v_s\right) - \mathbb{P}^* \left(V_s \leq v_s'\right) \right| \leq \varepsilon_1.
$$

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 $\|Q - \mathbb{P}^*\|_\infty \leq \delta$,

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

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

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

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

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

\[
G_s(\lambda', Q) \\
\leq G_s(\Lambda_s(P^*) - \varepsilon, Q) \\
\leq G_s(\Lambda_s(P^*) - \varepsilon, P^*) + |G_s(\Lambda_s(P^*) - \varepsilon, Q) - G_s(\Lambda_s(P^*) - \varepsilon, P^*)| \\
\leq G_s(\Lambda_s(P^*) - \varepsilon, P^*) + \varepsilon_1 \\
\leq G_s(\Lambda_s(P^*), P^*) - \varepsilon_1 \\
\leq \alpha - \varepsilon_1.
\]

This implies \( 1 - \varepsilon = \Lambda_s(P^*) - \varepsilon \leq \Lambda_s(Q) \leq 1 \) for all \( \|Q - P^*\|_x \leq \delta \).

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

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

\[
G_s(\lambda', Q) \\
\geq G_s(\Lambda_s(P^*) + \varepsilon, Q) \\
\geq \alpha + 2\varepsilon_1 - |G_s(\Lambda_s(P^*) + \varepsilon, Q) - G_s(\Lambda_s(P^*) + \varepsilon, P^*)| \\
\geq \alpha + \varepsilon_1.
\]

This implies that for \( \lambda' > \Lambda_s(P^*) + \varepsilon \) and \( \|Q - P^*\|_x \leq \delta \), \( G_s(\lambda', Q) > \alpha \). Thus, \( \Lambda_s(Q) \leq \Lambda_s(P^*) + \varepsilon \) for \( \|Q - P^*\|_x \leq \delta \). Combining the upper bound and lower bound of \( \Lambda_s(Q) \), we arrive at

\[
|\Lambda_s(Q) - \Lambda_s(P^*)| \leq \varepsilon
\]

for \( \|Q - P^*\|_x \leq \delta \).

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 $\lambda$ and $\lim_{\lambda \to 0^+} G_t(\lambda, \mathbb{P}^*) = 0 < \alpha$.

I Calculations for Example 3

We start with calculating $\mathbb{P}(\tau_k = 0| X_{k,1} = x_{k,1}, \ldots, X_{k,t} = x_{k,t})$. 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 $k = 1, \ldots, 4$. As a result, we have

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

for $t \geq t_k + 1$.

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

$$\mathbb{P}(\tau_k \leq t - 1| X_{k,1} = x_{k,1}, \ldots, X_{k,t} = x_{k,t}) \leq \mathbb{P}(\tau_k \leq t - 1| X_{k,1} = 1, \ldots, X_{k,t} = 1) \quad (I.1)$$

Indeed, direct calculation gives

$$\mathbb{P}(\tau_k \leq t - 1| X_{k,1} = x_{k,1}, \ldots, X_{k,t} = x_{k,t}) = \frac{\mathbb{P}(\tau_k = 0) - (0.51)\sum_{t=1}^t x_{k,t}(0.49)^t - \sum_{t=1}^t x_{k,t}P(\tau_k = t_k)(0.5)^t}{\mathbb{P}(\tau_k = 0)(0.51)\sum_{t=1}^t x_{k,t}(0.49)^t - \sum_{t=1}^t x_{k,t}P(\tau_k = t_k)(0.5)^t} \quad (I.2)$$

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

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

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

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 $\tilde{W}_{k,t} = 1$ for $t \geq 4$, and for $0 \leq t \leq 3$, the a.s. range of $\tilde{W}_{k,t}$s are given below (numbers are
rounded to the third decimal place).

<table>
<thead>
<tr>
<th>$\hat{W}_{k,t} \in$</th>
<th>$t = 1$</th>
<th>$t = 2$</th>
<th>$t = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k = 1$</td>
<td>[0.098, 0.102]</td>
<td>[0.096, 0.104]</td>
<td>[0.095, 0.105]</td>
</tr>
<tr>
<td>$k = 2$</td>
<td>[0.395, 0.405]</td>
<td>{1}</td>
<td>{1}</td>
</tr>
<tr>
<td>$k = 3$</td>
<td>[0.425, 0.435]</td>
<td>{1}</td>
<td>{1}</td>
</tr>
<tr>
<td>$k = 4$</td>
<td>[0.545, 0.555]</td>
<td>[0.540, 0.560]</td>
<td>[0.535, 0.565]</td>
</tr>
</tbody>
</table>

With these numbers, the following inequalities can be verified.

$$\hat{W}_{1,1} < \alpha < \hat{W}_{2,1} < \hat{W}_{3,1} < \hat{W}_{4,1},$$

$$\frac{1}{3}(\hat{W}_{1,1} + \hat{W}_{2,1} + \hat{W}_{3,1}) \leq 0.314 < \alpha = 0.34,$$

$$\frac{1}{3}(\hat{W}_{1,1} + \hat{W}_{2,1} + \hat{W}_{4,1}) \geq 0.346 > \alpha$$

$$\frac{1}{2}(\hat{W}_{1,1} + \hat{W}_{4,1}) \leq 0.329 < \alpha.$$

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

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

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

$$\hat{W}_{1,2} < \alpha < \hat{W}_{4,2} < \hat{W}_{2,2} = \hat{W}_{3,2},$$

$$\frac{1}{2}(\hat{W}_{1,2} + \hat{W}_{2,2}) = \frac{1}{2}(\hat{W}_{1,2} + \hat{W}_{3,2}) \geq 0.548 > \alpha$$

$$\frac{1}{2}(\hat{W}_{1,2} + \hat{W}_{4,2}) \leq 0.332 < \alpha.$$

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

$$S_3 \in \{\{1, 4\}, \{1\}, \emptyset\}.$$

Similarly, for $S_4$, we have

$$\hat{W}_{1,3} < \alpha < \hat{W}_{4,3} < \hat{W}_{2,3} = \hat{W}_{3,3},$$

$$\frac{1}{2}(\hat{W}_{1,2} + \hat{W}_{2,2}) = \frac{1}{2}(\hat{W}_{1,2} + \hat{W}_{3,2}) \geq 0.547 > \alpha$$

$$\frac{1}{2}(\hat{W}_{1,2} + \hat{W}_{4,2}) \leq 0.336 < \alpha.$$
This implies that $\mathbb{E}[\text{FNP}_4(T)|\mathcal{F}_3] \leq \alpha$ is equivalent to that $S_4 \subset S_3$ and

$$S_4 \in \{\{1, 4\}, \{1\}, \emptyset\}.$$ 

Finally, since $\tilde{W}_{k,t} = 1$ for all $t \geq 4$ and $k = 1, \ldots, 4$, we obtain $S_t = \emptyset$ for $t \geq 5$.

Enumerating all the index sets satisfying the constraint, we obtain that $\sup_{T \in \mathcal{T}_n} \mathbb{E}(U_2(T)) = 7$ and the maximum achieved if and only if $S_1 = \{1, 2, 3, 4\}$ and $S_2 = \{1, 2, 3\}$. In addition, $\sup_{T \in \mathcal{T}_n} \mathbb{E}(U_4(T)) = 10$ 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**


