C.1 List of Symbols

Here is a list of symbols used in the main manuscript and in this Web Appendix.

- \( w \): fixed motif length.
- \( L \): length of the observed nucleotide sequence \( S \).
- \( M \): known number of nucleotide types (typically =4 in practice).
- \( J \): number of motifs in the generative model (defined in Assumption 3.2).
- \( p_0 \): fixed motif frequency in the inference model (defined Section 2.1).
- \( S = (S_1, \ldots, S_L) \): observed sequence of nucleotides (defined Sec. 2.1).
- \( A = (A_1, \ldots, A_{L/w}) \): unknown vector of motif indicators (defined Sec. 2.1).
- \( \mathcal{X} = \{0, 1\}^{L/w} \): space of possible values for \( A \) (defined in Sec. 2.1).
- \( \theta_0 \): unknown length-\( M \) vector of background nucleotide frequencies (defined Sec. 2.1).
- \( \theta_{1:w} \): unknown matrix of position-specific nucleotide frequencies within the motif, where \( \theta_k \) has length \( M \) (defined Sec. 2.1).
- \( N(A^c); N(A^{(k)}); N(S) \): length-\( M \) nucleotide count vectors defined in (2.1).
- \( A_{[-i]} \): vector \( A \) with \( i \)th element removed; \( A_{[i,0]}, A_{[i,1]} \): vector \( A \) with \( i \)th element replaced by 0 or 1, respectively.
- \( \beta_0, \beta_1, \ldots, \beta_w \): fixed length-\( M \) vectors of constants (hyperparameters) used in the prior distribution of \( \theta_{0:w} \) (defined Sec. 2.1).
- \( p_1, \ldots, p_J \): as part of the generative model, the frequencies of the different “true” motifs (defined in Assumption 3.2).
- \( \theta_0^* \): as a part of the generative model, the true value of \( \theta_0 \) (defined in Assumption 3.2).
• $\theta^*_j: j \in \{1, \ldots, J\}$: as a part of the generative model, the multiple “true” values of the matrix $\theta_1$ (defined in Assumption 3.2).

• Gap($T$): the spectral gap of a transition matrix $T$ (defined in Section 2.3).

• $\pi(...)$: the likelihood, the prior, or the full, marginal, or conditional posterior distributions of the parameters, as distinguished by the arguments.

• $C(A); C(S)$: length-2 vectors of counts (defined in (5.3) and (5.4)).

• $\bar{X}$: space of possible values for $C(A)$ (defined in (5.5)).

• $\bar{\pi}(c|S)$: the marginal posterior distribution of $C(A)$, sometimes written with the dependence on $S$ suppressed (defined in (5.7)).

• $T$: the Markov transition matrix (2.6) associated with the Gibbs sampler; $\bar{T}$: the projection matrix (5.9) associated with the summary vector $C(A)$.

C.2 Proof of Lemma 3.1
For notational simplicity we give the proof for the case $M = 2$. With this choice, recall from (5.24) that the free parameters in $\theta_0$ are $\theta^*_k$, $1 \leq k \leq w$ for $k \in \{0, \ldots, w\}$, so we can write $\theta_0 \in [0, 1]^{w+1}$ and $\theta_1 \in [0, 1]^w$.

Let $\sum p_j$ be shorthand for $\sum_{j=1}^J p_j$. Define

$$\phi \triangleq \min \left\{ \frac{1}{1-p_1} \left(1 - \frac{1}{1-p_1} \frac{\sum_{j=1}^J p_j + \sum_{j=2}^J p_j}{1-p_1} \right) \right\}.$$  (C.1)

By Assumption 3.2 $\theta^*_{0,1} \in (0, 1)$, $p_j > 0$, and $\sum p_j < 1$, so

$$\phi \in \left(0, \min \{\theta^*_{0,1}, 1-\theta^*_{0,1}\}\right).$$  (C.2)

Using (3.4), define

$$\zeta \triangleq (\phi/4)^{\max\{2/\phi, 2/\phi\}} < \phi/4 < 1/4.$$  (C.3)

The constants $\phi, \zeta \in (0, 1)$ do not depend on $w$. Then, for any $w \in \{1, 2, \ldots\}$ and $j \in \{1, \ldots, J\}$ define

$$H^j_w \triangleq \left\{ \theta_1 \in [0, 1]^w : |\theta_{k,1} - \theta^*_{j,1}| \leq \zeta \; \forall k \in \{1, \ldots, w\} \right\}.$$  (C.4)

$$B^j_w \triangleq \left\{ \theta_1 \in [0, 1]^{w+1} : \theta_{1,w} \in H^1_w, \; \theta_{0,1} \in [\phi - \zeta, 1-\phi + \zeta] \right\}.$$  (C.5)

Since $\phi - \zeta > 0$, the interval $[\phi - \zeta, 1-\phi + \zeta]$ is bounded away from zero and one. By Assumption 3.3 for $w$ large enough and all $j, j' \in \{1, \ldots, J\}$ with $j \neq j'$ there is some $k \in \{1, \ldots, w\}$ such that $t^j_k \neq t^{j'}_k$. For this $k$ we have $\theta^*_k = 1 - \theta^*_j$, so $|\theta^*_k - \theta^*_{j'}| = 1 > 2\zeta$. So $B^j_w$ and $B^{j'}_w$ are disjoint.
Next we find a point $\theta^{(1)}_{0:w} \in B^1_w$ such that $\sup_{\partial B^1_w} \eta < \eta(\theta^{(1)}_{0:w})$. Then for any $j \neq 1$, $\exists \theta^{(j)}_{0:w} \in B^1_w$ with $\sup_{\partial B^1_w} \eta < \eta(\theta^{(j)}_{0:w})$ by symmetry, showing that (3.1) holds.

Also define
\[
h_w(\theta_{0:w}) \triangleq \sum_{s \in \{1,2\}^w} \left[ p_1 \prod_{k=1}^w \theta_{1,s_k} \right] \log \left[ p_0 \prod_{k=1}^w \theta_{0,s_k} \right]
\]
\[+ \sum_{s \in \{1,2\}^w} \left[ \sum_{j=2}^J p_j \prod_{k=1}^w \theta_{j,s_k}^* + (1 - \sum_{j=2}^J p_j) \prod_{k=1}^w \theta_{0,s_k}^* \right] \log \left[ (1 - p_0) \prod_{k=1}^w \theta_{0,s_k} \right]
\]
and note that
\[
\partial B^1_w = \text{cl}(B^1_w) \cap \text{cl}([0,1]^{w+1} \setminus B^1_w) \subset B^1_w
\]

since $B^1_w$ is closed. By (C.4)-(C.5),
\[
\partial B^1_w \subset \{ \theta_{0:w} : \theta_{0,1} \in \{ \phi - \zeta, 1 - \phi + \zeta \} \cup \{ \theta_{0,w} : \exists k : |\theta_{k,1} - \theta_{k,1}^*| = \zeta \} \}
\]

Lemma [C.1] below shows that $h_w(\theta_{0:w})$ is maximized at $(\hat{\theta}_0, \theta_{1:w}^*) \in B^1_w$ for some $\hat{\theta}_0$. We will show that
\[
\inf_{\theta_{0:w} \in \partial B^1_w} \left[ E \log f(s|\hat{\theta}_0, \theta_{1:w}^*) - E \log f(s|\theta_{0:w}) \right] > 0.
\]

Lemma [C.1] shows that $\exists b > 0$ such that for any $w$,
\[
\inf_{\theta_{0,w} \in \partial B^1_w} \left[ h_w(\hat{\theta}_0, \theta_{1:w}^*) - h_w(\theta_{0:w}) \right] > b > 0.
\]

For any constants $a_1, a_2, b_1, b_2$ we have that $a_1 - a_2 \geq b_1 - b_2 - |a_1 - b_1| - |a_2 - b_2|$. So for any $\theta_{0:w} \in \partial B^1_w$,
\[
E \log f(s|\hat{\theta}_0, \theta_{1:w}^*) - E \log f(s|\theta_{0:w})
\]
\[\geq h_w(\hat{\theta}_0, \theta_{1:w}^*) - h_w(\theta_{0:w}) - \left| E \log f(s|\hat{\theta}_0, \theta_{1:w}^*) - h_w(\hat{\theta}_0, \theta_{1:w}^*) \right|
\]
\[ - \left| E \log f(s|\theta_{0:w}) - h_w(\theta_{0:w}) \right|.
\]

Combining this with (C.7), (C.10), and Lemma [C.2] below, for $w$ large enough and any $\theta_{0:w} \in \partial B^1_w$
\[
E \log f(s|\hat{\theta}_0, \theta_{1:w}^*) - E \log f(s|\theta_{0:w}) > b - b/4 - b/4 \quad = b/2.
\]

So (C.9) holds for $w$ large enough, proving Lemma 3.1

Finally, we give the results used in the proof of Lemma 3.1
Lemma C.1. Under Assumptions 3.1-3.3, for any \( w \) the function \( h_w(\theta_{0:w}) \) defined in (C.6) is maximized at \( (\hat{\theta}_0, \theta_{1:w}^*) \) where

\[
\hat{\theta}_{0,1} = \frac{w (1 - \sum p_j) \theta_{0,1}^* + \sum_j p_j \sum_w \theta_{k,1}^*}{w(1 - p_1)} \in [\phi, 1 - \phi].
\] (C.11)

Also, using the definitions (C.5) and (C.7), Equation (C.10) holds for some \( b \) that does not depend on \( w \).

Proof. For \( s \in \{1, 2\}^w \) and \( m \in \{1, 2\} \) let \( \#\{s_k = m\} \) denote the number of indices \( k \in \{1, \ldots, w\} \) for which \( s_k = m \). Then

\[
\frac{\partial}{\partial \theta_{k,1}} h_w(\theta_{0:w}) = \sum_s \left[ p_1 \prod_{k'=1}^w \theta_{k',s_{k'}}^{1*} \right] \left[ \frac{1_{\{s_k=1\}}}{\theta_{k,1}} - \frac{1_{\{s_k=2\}}}{1 - \theta_{k,1}} \right] k \in \{1, \ldots, w\}
\]

\[
= p_1 \theta_{k,1}^{1*} \frac{1}{\theta_{k,1}} - p_1 (1 - \theta_{k,1}^{1*}) (1 - \theta_{k,1})
\] (C.12)

\[
\frac{\partial}{\partial \theta_{0,1}} h_w(\theta_{0:w}) = \sum_s \left[ \sum_{j=2}^J p_j \prod_{k=1}^w \theta_{s_k,j}^{j*} + (1 - \sum p_j) \prod_{k=1}^w \theta_{0,s_k}^* \right] \left[ \frac{\#\{s_k = 1\}}{\theta_{0,1}} - \frac{\#\{s_k = 2\}}{1 - \theta_{0,1}} \right]
\]

\[
= \frac{1}{\theta_{0,1}} \left( \sum_{j=2}^J p_j \sum_{k=1}^w \theta_{s_k,j}^{j*} + w(1 - \sum p_j) \theta_{0,1}^* \right) - \frac{1}{1 - \theta_{0,1}} \left( \sum_{j=2}^J p_j \sum_{k=1}^w (1 - \theta_{s_k,j}^{j*}) + w(1 - \sum p_j) (1 - \theta_{0,1}^*) \right).
\] (C.13)

Setting this equal to zero and solving for \( \theta_{0,1} \) and \( \theta_{k,1} \) shows that \( h_w(\theta_{0:w}) \) has a stationary point at \( (\hat{\theta}_0, \theta_{1:w}^*) \). Using (C.11), \( \hat{\theta}_{0,1} \in [\phi, 1 - \phi] \).

Note that \( \frac{\partial^2}{\partial \theta_{k,1} \partial \theta_{k',1}} h_w(\theta_{0:w}) = 0 \) for any \( k \neq k' \), that \( \frac{\partial^2}{\partial \theta_{k,1} \partial \theta_{0,1}} h_w(\theta_{0:w}) = 0 \) for any \( k \), and that

\[
\frac{\partial^2}{\partial \theta_{k,1}^2} h_w(\theta_{0:w}) = -p_1 \theta_{k,1}^{1*} \frac{1}{\theta_{k,1}^2} - p_1 (1 - \theta_{k,1}^{1*}) \frac{1}{(1 - \theta_{k,1})^2} \leq -p_1 \theta_{k,1}^{1*} - p_1 (1 - \theta_{k,1}^{1*}) = -p_1 \] (C.14)

\[
\frac{\partial^2}{\partial \theta_{0,1}^2} h_w(\theta_{0:w}) = -\frac{1}{\theta_{0,1}^2} \left( \sum_{j=2}^J p_j \sum_{k=1}^w \theta_{s_k,j}^{j*} + w(1 - \sum p_j) \theta_{0,1}^* \right) - \frac{1}{(1 - \theta_{0,1})^2} \left( \sum_{j=2}^J p_j \sum_{k=1}^w (1 - \theta_{s_k,j}^{j*}) + w(1 - \sum p_j) (1 - \theta_{0,1}^*) \right)
\]

\[
\leq -w(1 - p_1) \leq -(1 - p_1). \] (C.15)
So $h_w(\theta_{0:w})$ is maximized at $(\hat{\theta}_0, \theta_{1:w}^*)$.

To show the second part of Lemma C.1 recall (C.8). We first address $\theta_{0:w}$ such that $\theta_{0,1} = 1 - \phi + \zeta$. Using (C.13) we have $\frac{\partial}{\partial \theta_{0,1}} h_w(\theta_{0:w})\bigg|_{\theta_{0,1} = \hat{\theta}_{0,1}} = 0$. Applying (C.15), for any $\theta_{0:w}$ such that $\theta_{0,1} = 1 - \phi + \zeta$,

$$h_w(\theta_{0:w}) - h_w(\hat{\theta}_0, \theta_{1:w}) = \int_{\theta_{0,1}}^{1-\phi+\zeta} \frac{\partial}{\partial \theta_{0,1}} h_w(\theta_{0:w}) \bigg|_{\theta_{0,1} = z} \, dz$$

$$= \int_{\theta_{0,1}}^{1-\phi+\zeta} \int_{\theta_{0,1}}^z \frac{\partial^2}{\partial \theta_{0,1}^2} h_w(\theta_{0:w}) \bigg|_{\theta_{0,1} = w} \, dw \, dz$$

$$\leq -(1 - p_1)(1 - \phi + \zeta - \hat{\theta}_{0,1})^2 / 2 \leq -(1 - p_1)\zeta^2 / 2. \quad (C.16)$$

By (C.12), for any fixed value of $\theta_0$ the function $h_w(\theta_{0:w})$ is maximized at $(\theta_0, \theta_{1:w}^*)$. Combining with (C.16),

$$\inf_{\theta_{0:w}:\theta_{0,1} = 1-\phi+\zeta} \left[ h_w(\hat{\theta}_0, \theta_{1:w}^*) - h_w(\theta_{0:w}) \right] \geq \inf_{\theta_{0,w}:\theta_{0,1} = 1-\phi+\zeta} \left[ h_w(\hat{\theta}_0, \theta_{1:w}^*) - h_w(\theta_0, \theta_{1:w}^*) \right] \geq (1 - p_1)\zeta^2 / 2 \quad (C.17)$$

which is positive and does not depend on $w$.

Analogously, for $\theta_{0:w}$ such that $\theta_{0,1} = \phi - \zeta$ we have

$$\inf_{\theta_{0:w}:\theta_{0,1} = \phi-\zeta} \left[ h_w(\hat{\theta}_0, \theta_{1:w}^*) - h_w(\theta_{0:w}) \right] \geq (1 - p_1)\zeta^2 / 2. \quad (C.18)$$

Using the analogous argument to handle the case where $\exists k : |\theta_{k,1} - \theta_{k,1}^*| = \zeta$, and combining with (C.8), (C.17) and (C.18) yields (C.10). This proves Lemma C.1.

---

**Lemma C.2.** Under Assumptions 3.1-3.3 and using the definitions (C.5) and (C.6),

$$\sup_{\theta_{0,w} \in B_w} |E \log f(s|\theta_{0:w}) - h_w(\theta_{0:w})| \xrightarrow{w \to \infty} 0. \quad (C.19)$$

**Proof.** Using Assumption 3.3, $\prod_{k=1}^w \theta_{k,s_k}^* = 1$ if $s = t_{1:w}^1$ and $\prod_{k=1}^w \theta_{k,s_k}^* = 0$ for all other $s \in \{1, 2\}^w$. Combining with (2.8) and (3.3), the first term of $E \log f(s|\theta_{0:w}) = \sum_s g_P(s) \log f(s|\theta_{0:w})$ |
is

\[
\sum_s \left[ p_1 \prod_{k=1}^w \theta_{k,s_k}^{1*} \right] \log f(s|\theta_{0:w}) \tag{C.20}
\]

\[= p_1 \log \left[ p_0 \prod_{k=1}^w \theta_{k,t_k^1} + (1 - p_0) \prod_{k=1}^w \theta_{0,t_k^1} \right]. \]

We have that

\[
\log \left[ p_0 \prod_{k=1}^w \theta_{k,t_k^1} + (1 - p_0) \prod_{k=1}^w \theta_{0,t_k^1} \right] - \log \left[ p_0 \prod_{k=1}^w \theta_{k,t_k^1} \right] \geq 0. \tag{C.21}
\]

Also, using (C.3)-(C.5) and the fact that \( \theta_{1*,k,t_k^1} = 1 \) for all \( k \in \{1, \ldots, w\} \),

\[
\sup_{\theta_{0:w} \in B_w^1} \frac{(1 - p_0) \prod_{k=1}^w \theta_{0,t_k^1}}{p_0 \prod_{k=1}^w \theta_{k,t_k^1}} \leq \frac{(1 - p_0)(1 - \phi + \zeta) \prod_{k=1}^w \theta_{1*,k,t_k^1}}{p_0(1 - \zeta)^w} \xrightarrow{w \to \infty} 0
\]
since \( 1 - \phi + \zeta < 1 - \zeta \). So

\[
\sup_{\theta_{0:w} \in B_w^1} \left( \log \left[ p_0 \prod_{k=1}^w \theta_{k,t_k^1} + (1 - p_0) \prod_{k=1}^w \theta_{0,t_k^1} \right] - \log \left[ p_0 \prod_{k=1}^w \theta_{k,t_k^1} \right] \right)
\leq \log \left[ 1 + \frac{(1 - p_0)(1 - \phi + \zeta) \prod_{k=1}^w \theta_{1*,k,t_k^1}}{p_0(1 - \zeta)^w} \right] \xrightarrow{w \to \infty} 0.
\]

Combining with (C.21),

\[
\sup_{\theta_{0:w} \in B_w^1} \left| \log \left[ p_0 \prod_{k=1}^w \theta_{k,t_k^1} + (1 - p_0) \prod_{k=1}^w \theta_{0,t_k^1} \right] - \log \left[ p_0 \prod_{k=1}^w \theta_{k,t_k^1} \right] \right| \xrightarrow{w \to \infty} 0.
\]

So, using (C.20),

\[
\sup_{\theta_{0:w} \in B_w^1} \left| \sum_s \left[ p_1 \prod_{k=1}^w \theta_{k,s_k}^{1*} \right] \log f(s|\theta_{0:w}) - \sum_s \left[ p_1 \prod_{k=1}^w \theta_{k,s_k}^{1*} \right] \log \left[ p_0 \prod_{k=1}^w \theta_{k,s_k} \right] \right| \xrightarrow{w \to \infty} 0. \tag{C.22}
\]

Next we approximate the middle terms of \( \sum_s \log f(s|\theta_{0:w}) \). Using (2.8), (3.3), and Assumption 3.3 they are of the following form for \( j \in \{2, \ldots, J\} \).

\[
\sum_s \left[ p_j \prod_{k=1}^w \theta_{k,s_k}^{j*} \right] \log f(s|\theta_{0:w}) \tag{C.23}
\]

\[= p_j \log \left[ p_0 \prod_{k=1}^w \theta_{k,t_k^j} + (1 - p_0) \prod_{k=1}^w \theta_{0,t_k^j} \right]. \]
We have that
\[
\log \left[ p_0 \prod_{k=1}^w \theta_{k,t_k} + (1 - p_0) \prod_{k=1}^w \theta_{0,t_k} \right] - \log \left[ (1 - p_0) \prod_{k=1}^w \theta_{0,t_k} \right] \geq 0.
\] (C.24)

Let \( \#\{t_k^j = t_k^1\} \) indicate the number of indices \( k \in \{1, \ldots, w\} \) for which \( t_k^j = t_k^1 \). Using (C.4)-(C.5) and the fact that \( \theta_{k,t_k^j} = 0 \) for all \( k \) such that \( t_k^j \neq t_k^1 \), we have that
\[
\sup_{\theta_{0,w} \in B_{1,w}} \frac{p_0 \prod_{k=1}^w \theta_{k,t_k^j}}{(1 - p_0) \prod_{k=1}^w \theta_{0,t_k^j}} \leq \frac{p_0 \zeta^{w\alpha/2}}{(1 - p_0)(\phi - \zeta)^w}.
\]
Combining this with Assumption 3.3 and (C.3), for all \( w \) large enough
\[
\sup_{\theta_{0,w} \in B_{1,w}} \left( \log \left[ \frac{p_0 \prod_{k=1}^w \theta_{k,t_k^j} + (1 - p_0) \prod_{k=1}^w \theta_{0,t_k^j}}{(1 - p_0) \prod_{k=1}^w \theta_{0,t_k^j}} \right] - \log \left[ (1 - p_0) \prod_{k=1}^w \theta_{0,t_k^j} \right] \right) \leq \log \left( \frac{p_0(\phi/4)^w}{(1 - p_0)(\phi - \zeta)^w} + 1 \right) \rightarrow 0.
\] (C.25)

Using (C.24) and (C.25),
\[
\sup_{\theta_{0,w} \in B_{1,w}} \left( \log \left[ \frac{p_0 \prod_{k=1}^w \theta_{k,t_k^j} + (1 - p_0) \prod_{k=1}^w \theta_{0,t_k^j}}{(1 - p_0) \prod_{k=1}^w \theta_{0,t_k^j}} \right] - \log \left[ (1 - p_0) \prod_{k=1}^w \theta_{0,t_k^j} \right] \right) \rightarrow 0.
\]
Combining with (C.23), for \( j \in \{2, \ldots, J\} \)
\[
\sup_{\theta_{0,w} \in B_{1,w}} \sum_s \left[ \sum_{k=1}^w p_j \prod_{k=1}^w \theta_{k,s_k}^j \log f(s|\theta_{0,w}) - \sum_{k=1}^w \theta_{0,s_k}^j \log \left( 1 - p_0 \prod_{k=1}^w \theta_{0,s_k} \right) \right] \rightarrow 0.
\] (C.26)

Finally we address the last term of of \( \sum_s g_{\theta^*}(s) \log f(s|\theta_{0,w}) \). Using (2.8) and (3.3) it is
\[
\sum_s \left[ (1 - \sum_{k=1}^w \theta_{0,s_k}^j) \log f(s|\theta_{0,w}) \right] = \sum_s \left[ (1 - \sum_{k=1}^w \theta_{0,s_k}^j) \log \left( p_0 \prod_{k=1}^w \theta_{k,s_k} + (1 - p_0) \prod_{k=1}^w \theta_{0,s_k} \right) \right].
\] (C.27)
We will show that a subset of sequences $s$ can be omitted when considering (C.27). Denote by $F(x; n, q)$ the cumulative distribution function of a Binomial$(n, q)$ random variable, evaluated at $x \in \mathbb{R}$. For $s \in \{1, 2\}^w$ recall that $\# \{ s_k \neq t_k \}$ denotes the number of indices $k \in \{1, \ldots, w\}$ for which $s_k \neq t_k$. Define

$$D_w \triangleq \{ s : \# \{ s_k \neq t_k \} > w \phi / 4 \}. \quad (C.28)$$

Then

$$\sum_{s \in D_w} \left( \prod_{k=1}^{w} \theta_{0,s_k}^{s} \right) \geq \max \left\{ \sum_{s : \# \{ s_k \neq t_k \}} \left( \prod_{k=1}^{w} \theta_{0,s_k}^{s} \right), \sum_{s : \# \{ s_k \neq t_k, t_k \neq 2 \}} \left( \prod_{k=1}^{w} \theta_{0,s_k}^{s} \right) \right\}$$

$$= \max \left\{ \sum_{s : \# \{ s_k = 2, t_k = 1 \}} \left( \prod_{k=1}^{w} \theta_{0,s_k}^{s} \right), \sum_{s : \# \{ s_k = 1, t_k = 2 \}} \left( \prod_{k=1}^{w} \theta_{0,s_k}^{s} \right) \right\}$$

$$= \max \left\{ 1 - F(w \phi / 4; \# \{ t_k = 1 \}, 1 - \theta_{0,1}^{s}), 1 - F(w \phi / 4; \# \{ t_k = 2 \}, \theta_{0,1}^{s}) \right\}. \quad (C.29)$$

For fixed $x$, $F(x; n, q)$ is monotonic nonincreasing in $n$ and $q$. Using (C.2) and (C.29), since $\phi < \min \{ \theta_{0,1}, 1 - \theta_{0,1} \}$ and $w / 2 \leq \max\{ \# \{ t_k = 1 \}, \# \{ t_k = 2 \} \}$, we have the following.

$$\sum_{s \in D_w} \left( \prod_{k=1}^{w} \theta_{0,s_k}^{s} \right) \geq \max \left\{ 1 - F(w \phi / 4; \# \{ t_k = 1 \}, \phi), 1 - F(w \phi / 4; \# \{ t_k = 2 \}, \phi) \right\}$$

$$= 1 - F \left( w \phi / 4; \max \{ \# \{ t_k = 1 \}, \# \{ t_k = 2 \} \}, \phi \right)$$

$$\geq 1 - F \left( w \phi / 4; w / 2, \phi \right). \quad (C.30)$$

Using the normal approximation to the binomial distribution, the quantity $F(w \phi / 4; w / 2, \phi)$ decays exponentially in $w$. So by (C.30), the sum

$$\sum_{s \in D_w} \left( \prod_{k=1}^{w} \theta_{0,s_k}^{s} \right) = 1 - \sum_{s \in D_w} \left( \prod_{k=1}^{w} \theta_{0,s_k}^{s} \right) \quad (C.31)$$
decays exponentially in $w$. Using this fact and (C.5),

$$\sup_{\theta_{0:w} \in B^1_w} \left| \sum_{s \notin D_w} (1 - \sum p_j) \prod_{k=1}^w \theta^*_{0,s_k} \right| \log \left[ p_0 \prod_{k=1}^w \theta_{k,s_k} + (1 - p_0) \prod_{k=1}^w \theta_{0,s_k} \right]$$

$$\leq \sup_{\theta_{0:w} \in B^1_w} \left[ \sum_{s \notin D_w} (1 - \sum p_j) \prod_{k=1}^w \theta^*_{0,s_k} \right] \min_s \left[ (1 - p_0) \prod_{k=1}^w \theta_{0,s_k} \right]$$

$$\leq \left[ (1 - \sum p_j) \sum_{s \notin D_w} \prod_{k=1}^w \theta^*_{0,s_k} \right] \log \left[ (1 - p_0)(\phi - \zeta)^w \right]$$

$$\xrightarrow{w \to \infty} 0. \quad (C.32)$$

Using (C.3)-(C.5) and (C.28), for $\theta_{0:w} \in B^1_w$ and $s \in D_w$,

$$\frac{p_0 \prod_{k=1}^w \theta_{k,s_k}}{(1 - p_0) \prod_{k=1}^w \theta_{0,s_k}} \leq \frac{p_0 e^{\#(s \in t^1_w)}}{(1 - p_0)(\phi - \zeta)^w} \frac{p_0 \zeta^{w/4}}{(1 - p_0)(\phi - \zeta)^w} \leq \frac{p_0(\phi/4)^w}{(1 - p_0)(\phi - \zeta)^w} \xrightarrow{w \to \infty} 0$$

uniformly over $\theta_{0:w} \in B^1_w$ and $s \in D_w$, since $\phi/4 < \phi - \zeta$. So

$$\sum_{s \in D_w} \left[ (1 - \sum p_j) \prod_{k=1}^w \theta^*_{0,s_k} \right] \log \left[ p_0 \prod_{k=1}^w \theta_{k,s_k} + (1 - p_0) \prod_{k=1}^w \theta_{0,s_k} \right]$$

$$- \sum_{s \in D_w} \left[ (1 - \sum p_j) \prod_{k=1}^w \theta^*_{0,s_k} \right] \log \left[ (1 - p_0) \prod_{k=1}^w \theta_{0,s_k} \right] \xrightarrow{w \to \infty} 0 \quad (C.33)$$

uniformly over $\theta_{0:w} \in B^1_w$. Also, using an analogous argument to (C.32),

$$\sup_{\theta_{0:w} \in B^1_w} \left| \sum_{s \notin D_w} (1 - \sum p_j) \prod_{k=1}^w \theta^*_{0,s_k} \right| \log \left[ (1 - p_0) \prod_{k=1}^w \theta_{0,s_k} \right] \xrightarrow{w \to \infty} 0. \quad (C.34)$$

Combining (C.32)-(C.34),

$$\sup_{\theta_{0:w} \in B^1_w} \left| \sum_s \left[ (1 - \sum p_j) \prod_{k=1}^w \theta^*_{0,s_k} \right] \log f(s|\theta_{0:w}) \right.$$ 

$$- \sum_s \left[ (1 - \sum p_j) \prod_{k=1}^w \theta^*_{0,s_k} \right] \log \left[ (1 - p_0) \prod_{k=1}^w \theta_{0,s_k} \right] \xrightarrow{w \to \infty} 0. \quad (C.35)$$

Putting together the results (C.22), (C.26), and (C.35) for the various terms, we have that $\sum_s g_{\theta^*}(s) \log f(s|\theta_{0:w})$ converges to $h_w(\theta_{0:w})$, uniformly over $\theta_{0:w} \in B^1_w$. \qed
C.3 Proof of Theorem 3.3

For simplicity of notation we state the proof for the case \( M = 2 \) and \( \beta_{k,m} = 1 \) for all \( k, m \), although the proof is analogous for any other choices of these constants. Recall the definitions of \( C(\mathbf{A}), \bar{X}, \bar{\pi}, D_c, \) and \( \bar{T} \) from Equations (5.3), (5.5), and (5.7)-(5.9). In the case \( w = 1 \) and \( M = 2 \) the vector \( C(\mathbf{A}) \in \bar{X} \) only has two elements, \( n = C(\mathbf{A})_1 \) and \( r = C(\mathbf{A})_2 \). So we write \( \bar{\pi}(n, r) \), suppressing the dependence on \( \bar{\pi} \) on \( S \). Using (5.7), \( \bar{\pi}(n, r) = \sum_{\mathbf{A} : C(\mathbf{A}) = (n, r)} \bar{\pi}(\mathbf{A}|S) \).

Since \( D_{(n,r)} = \{ \mathbf{A} \in \bar{X} : C(\mathbf{A}) = (n, r) \} \), let \( |D_{(n,r)}| \) be the cardinality of \( D_{(n,r)} \) and note that \( |D_{(n,r)}| = \binom{N(S)}{n} \binom{N(S)}{r} \). Using (5.6) we have \( |A| = n + r, N(\mathbf{A}^{(1)})_1 = n, N(\mathbf{A}^{(1)})_2 = r, N(\mathbf{A}^c)_1 = N(S)_1 - n, \) and \( N(\mathbf{A}^c)_2 = N(S)_2 - r \). Then \( \bar{\pi} \) simplifies as follows, using (2.5):

\[
\bar{\pi}(n, r) \propto |D_{(n,r)}| p_0^{n+r} (1 - p_0)^{L-n-r} \frac{\Gamma(N(S)_1 - n + \beta_{0,1}) \Gamma(N(S)_2 - r + \beta_{0,2}) \Gamma(n + \beta_{1,1}) \Gamma(r + \beta_{1,2})}{\Gamma(L - n - r + |\beta_0|) \Gamma(n + r + |\beta_1|)}
\]

\[
= |D_{(n,r)}| p_0^{n+r} (1 - p_0)^{L-n-r} \frac{\Gamma(N(S)_1 - n + 1) \Gamma(N(S)_2 - r + 1) \Gamma(n + 1) \Gamma(r + 1)}{\Gamma(L - n - r + 2) \Gamma(n + r + 2)}
\]

\[
= \frac{N(S)_1!}{n!(N(S)_1 - n)!} \frac{N(S)_2!}{r!(N(S)_2 - r)!} p_0^{n+r} (1 - p_0)^{L-n-r} \times \frac{(N(S)_1 - n)!(N(S)_2 - r)!}{(L - n - r + 1)!} \frac{n!}{(n + r + 1)!}
\]

\[
\propto p_0^{n+r} (1 - p_0)^{L-n-r} \frac{1}{(L - n - r + 1)!(n + r + 1)!}.
\]

This is a function of \( (n + r) \) only; \( \bar{\pi}(n, r) \) is also unimodal in \( (n + r) \), shown as follows. The ratio

\[
\frac{\bar{\pi}(n+1, r)}{\bar{\pi}(n, r)} = \frac{\bar{\pi}(n, r + 1)}{\bar{\pi}(n, r)} = \frac{p_0}{1 - p_0} \left( \frac{L - n - r + 1}{n + r + 2} \right)
\]

is > 1 iff \( n + r < p_0L + 3p_0 - 2 \), showing that \( \bar{\pi}(n, r) \) is unimodal in \( (n + r) \).

Using (2.6) and (5.9), in each iteration of \( \bar{T} \) the quantity \( (n + r) \) can only be incremented or decremented by one. Using (C.37) we have that incrementing or decrementing \( (n + r) \) by one changes \( \bar{\pi}(n, r) \) by no more than a factor of

\[
d_2 \triangleq \max \left\{ \frac{L - n - r + 1}{1 - p_0}, \frac{n + r + 2}{p_0} \right\} = O(L).
\]

We will find a lower bound for the quantity \( d \) defined in (5.11), by defining a path \( \gamma_{c_1,c_2} \) in the graph of \( \bar{T} \) for every pair of states \( c_1, c_2 \in \bar{X} \). We will construct the paths in such a way that for any state \( c \in \gamma_{c_1,c_2} \) we have \( \bar{\pi}(c) \geq \min\{\bar{\pi}(c_1), \bar{\pi}(c_2)\} / d_2 \). Denote \( c_1 = (n_1, r_1) \) and \( c_2 = (n_2, r_2) \). If \( n_1 \leq n_2 \) and \( r_1 \leq r_2 \), then construct the path by first increasing the first coordinate \( n \) from \( n_1 \) to \( n_2 \), then by increasing the second coordinate \( r \) from \( r_1 \) to \( r_2 \).
Along this path, \( n + r \) increases at every step. Since \( \bar{\pi}(n, r) \) is a function only of \( n + r \) and is unimodal in \( n + r \), we have that for states \((n, r)\) along the path,

\[
\bar{\pi}(n, r) \geq \min\{\bar{\pi}(n_1, r_1), \bar{\pi}(n_2, r_2)\} \geq \min\{\bar{\pi}(n_1, r_1), \bar{\pi}(n_2, r_2)\}/d_2.
\]

The case where \( n_1 \geq n_2 \) and \( r_1 \geq r_2 \) is analogous, since we can construct a path in the opposite direction as above. Now consider the case where \( n_1 \leq n_2 \) and \( r_1 > r_2 \) (the case \( n_1 > n_2, r_1 \leq r_2 \) is equivalent). Starting at \((n_1, r_1)\), first decrement \( r \) by one, then increment \( n \) by one, and repeat until either \( r = r_2 \) or \( n = n_2 \). Notice that so far \( n + r \) has changed by at most one, so that \( \bar{\pi}(n, r) \) has changed by at most a factor of \( d_2 \). At this point, if \( r = r_2 \) then increase \( n \) until \( n = n_2 \), or if \( n = n_2 \) then decrease \( r \) until \( r = r_2 \). Any state \((n, r)\) along this path satisfies \( \bar{\pi}(n, r) \geq \min\{\bar{\pi}(n_1, r_1), \bar{\pi}(n_2, r_2)\}/d_2 \) as desired. Using (C.38), the quantity \( d \) defined in (5.11) satisfies \( d^{-1} = O(L) \). Combined with (5.13) and Proposition 5.2 this proves Theorem 3.3.

### C.4 Verifying the Assumptions of Theorem A.1

By (5.29) \( \Lambda \) is a Borel set, and \( \text{Int}(B_j) \) is a Borel set for \( j \in \{1, 2\} \) because it is open. So the spaces \( \Lambda_j \) for \( j \in \{1, 2\} \) are Borel subsets of the complete, separable metric space \( \mathbb{R}^{w+1} \) as required. Also, \( f(s|\theta_{0w}) \) is measurable jointly in \( s \) and \( \theta_{0w} \) since it is a continuous function of \( \theta_{0w} \) and since \( s \) takes a finite set of values. Of course, \( \Lambda_j \) might not be connected, in which case \( f(s|\theta_{0w}) \) being continuous simply means that it is continuous on each connected component of \( \Lambda_j \). Assumption 4 of Theorem A.1 is satisfied since \( \eta(\theta_{0w}) = E \log f(s|\theta_{0w}) \) is continuous. To show Assumption 2, observe that for all \( \theta_{0w} \in \Lambda_j \) where \( j \in \{1, 2\}, f(s|\theta_{0w}) > 0 \) for any \( s \in \{1, 2\}^w \), so \( G\{s \in \{1, 2\}^w : f(s|\theta_{0w}) > 0\} = 1 \) as desired.

To show Assumption 3 for \( \Lambda_1 \), take any compact \( F \subset \Lambda_1 \). We claim that there is some \( \zeta \in (0, \frac{1}{2}) \) such that

\[
F \subset ([\zeta, 1 - \zeta] \times [0, 1]^w) \cup ([0, 1] \times [\zeta, 1 - \zeta]^w) \setminus \text{Int}(B_2). \tag{C.39}
\]

Otherwise, there is some sequence \( \{\theta_{0w}^{(l)} : l \in \mathbb{N}\} \) such that \( \lim_{l \to \infty} \theta_{0w}^{(l)} = 0, 1 \) and \( \exists k \in \{1, \ldots, w\} \) such that \( \lim_{l \to \infty} \theta_{k,w}^{(l)} \in \{0, 1\} \). Since \( F \) is compact these points must have a limit point \( \tilde{\theta}_{0w} \in F \subset \Lambda_1 \). Then \( \tilde{\theta}_{0w} \in F \subset \Lambda_1 \). Then \( \tilde{\theta}_{0w} \in \{0, 1\} \) and \( \tilde{\theta}_{k,w} \in \{0, 1\} \), which is a contradiction.

By (C.39), for any \( \theta_{0w} \in F \) and any \( s \) we have \( f(s|\theta_{0w}) \geq \min\{p_0, 1 - p_0\} \zeta^w \). Then

\[
E \sup_{\theta_{0w} \in F} |\log f(s|\theta_{0w})| \leq \sup_{s \in \{1, 2\}^w, \theta_{0w} \in F} |\log f(s|\theta_{0w})| \leq -\log [\min\{p_0, 1 - p_0\} \zeta^w] < \infty.
\]

To show that Assumption 5 is satisfied for \( \Lambda_1 \), it is sufficient to consider values of \( r \in \mathbb{R} \) for which \( r < (\log \frac{1}{2})(\min s \theta_{0w}) \). Let \( \psi = \exp\{\frac{r}{\min \theta_{0w}}\} \), so that \( \psi \in (0, \frac{1}{2}) \). Then define \( D = \Lambda_1 \setminus D^c \) by letting \( D^c \) be the compact subset

\[
D^c = ([\psi, 1 - \psi] \times [0, 1]^w) \cup ([0, 1] \times [\psi, 1 - \psi]^w) \setminus \text{Int}(B_2) \quad \subset \Lambda_1.
\]
We will define a cover $D_1, \ldots, D_K$ of $D$ such that (A.1) holds. Define
\begin{align*}
D_{k00} &= \{ \theta_{0:w} \in [0,1]^{w+1} : \theta_{0,1} \in [0,\psi] \land \theta_{k,1} \in [0,\psi] \} \quad k \in \{1, \ldots, w\} \\
D_{k10} &= \{ \theta_{0:w} \in [0,1]^{w+1} : \theta_{0,1} \in (1-\psi, 1] \land \theta_{k,1} \in [0,\psi] \} \\
D_{k01} &= \{ \theta_{0:w} \in [0,1]^{w+1} : \theta_{0,1} \in [0,\psi] \land \theta_{k,1} \in (1-\psi, 1] \} \\
D_{k11} &= \{ \theta_{0:w} \in [0,1]^{w+1} : \theta_{0,1} \in (1-\psi, 1] \land \theta_{k,1} \in (1-\psi, 1] \}.
\end{align*}

For all $\theta_{0:w} \in D$ we have $\theta_{0,1} \in [0,\psi) \cup (1-\psi, 1]$ and $\exists k \in \{1, \ldots, w\} : \theta_{k,1} \in [0,\psi) \cup (1-\psi, 1]$. So
\[
D \subseteq \bigcup_{k=1}^{w} (D_{k00} \cup D_{k10} \cup D_{k01} \cup D_{k11}).
\]

Since $\log f(s|\theta_{0:w}) \leq 0$, for any $k \in \{1, \ldots, w\}$
\[
E \sup_{\theta_{0:w} \in D_{k00}} \log f(s|\theta_{0:w}) \leq g(t) \sup_{\theta_{0:w} \in D_{k00}} \log f(t|\theta_{0:w}) \quad \text{where } t = (1, \ldots, 1)
\leq g(t) \log [p_0\psi + (1-p_0)\psi] \leq \left[ \min_s g(s) \right] \log \psi = r.
\]

Also,
\[
E \sup_{\theta_{0:w} \in D_{k01}} \log f(s|\theta_{0:w}) \leq g(t) \sup_{\theta_{0:w} \in D_{k01}} \log f(t|\theta_{0:w}) \quad \text{where } t = (1, \ldots, 1, 2, 1, \ldots, 1) \quad k-1 \text{ ones}
\leq \left[ \min_s g(s) \right] \log [p_0\psi + (1-p_0)\psi] = r.
\]

Analogously, $E \sup_{\theta_{0:w} \in D_{k10}} \log f(s|\theta_{0:w}) \leq r$ and $E \sup_{\theta_{0:w} \in D_{k11}} \log f(s|\theta_{0:w}) \leq r$, showing that Assumption 5 holds for $A_1$. Since Assumptions 3 and 5 hold for $A_1$, they hold for $A_2$ by symmetry.

C.5 \textbf{Proof of Theorem 5.3}

Assume that there exist $\epsilon > 0$ and $B_1, B_2 \subset [0,1]^{w+1}$ separated by distance $\epsilon$ such that the ratios in (C.25) decrease exponentially in $L$, and take $F_1, F_2$ as in Proposition C.1 below. Letting $c_1$ be a maximizer of $\bar{\pi}(c|S)$ over $c \in F_1$, and $c_2$ be a maximizer of $\bar{\pi}(c|S)$ over $c \in F_2$ and using Proposition C.1 for all $L$ large enough
\[
\max\{\bar{\pi}(c_1|S), \bar{\pi}(c_2|S)\} \geq \frac{1}{2} (\bar{\pi}(c_1|S) + \bar{\pi}(c_2|S)) \geq \frac{\bar{\pi}(F_1|S)}{2|F_1|} + \frac{\bar{\pi}(F_2|S)}{2|F_2|} \geq \frac{1}{2|X|} (\bar{\pi}(F_1|S) + \bar{\pi}(F_2|S)) \geq \frac{1}{4|X|}.
\]

Combining with the fact that any path from $c_1$ to $c_2$ must include a state in $(F_1 \cup F_2)^c$,
\[
\begin{align*}
\max_{\gamma \in \Gamma_{e_1,e_2}} \min_{c \in c_{\gamma}} \frac{\bar{\pi}(c|S)}{\bar{\pi}(c_1|S) \bar{\pi}(c_2|S)} &\leq \max_{\gamma \in \Gamma_{e_1,e_2}} \min_{c \in c_{\gamma}} \frac{4|X| \bar{\pi}(c|S)}{\min\{\bar{\pi}(c_1|S), \bar{\pi}(c_2|S)\}} \\
&\leq \max_{c \in (F_1 \cup F_2)^c} \min\{\bar{\pi}(c_1|S), \bar{\pi}(c_2|S)\} \\
&\leq \frac{4|X|}{\min\{\bar{\pi}(c_1|S), \bar{\pi}(c_2|S)\}} \leq \frac{4|X|^2 \bar{\pi}(F_1 \cup F_2|S)}{\min\{\bar{\pi}(F_1|S), \bar{\pi}(F_2|S)\}}.
\end{align*}
\]
Since $|\tilde{X}|$ grows polynomially in $L$ (using (5.10)), and using Proposition C.1, the quantity $d$ decreases exponentially in $L$. □

**Proposition C.1.** If there exist $\epsilon > 0$ and two sets $B_1, B_2 \subset [0, 1]^{w+1}$ separated by Euclidean distance $\epsilon$ such that the ratios in (5.25) decrease exponentially in $L$, then there are two sets $F_1, F_2 \subset \tilde{X}$ such that:

1. For any $c_1 \in F_1$ and $c_2 \in F_2$, any path from $c_1$ to $c_2$ must include a state $c \notin (F_1 \cup F_2)$.

2. The quantities
\[
\frac{\bar{\pi}((F_1 \cup F_2)^c|S)}{\bar{\pi}(F_1|S)} \quad \text{and} \quad \frac{\bar{\pi}((F_1 \cup F_2)^c|S)}{\bar{\pi}(F_2|S)}
\]
(C.41)
decrease exponentially in $L$.

Before proving Proposition C.1 we need a few preliminary results. The notation $\sim$ means independently distributed as.

**Lemma C.3.** For any measure $\nu(dz)$ and nonnegative functions $a(z)$ and $b(z)$ on a space $z \in Z$,
\[
\int a(z)\nu(dz) \geq \inf_{z \in Z} \frac{a(z)}{b(z)}.
\]
where the ratio inside the infimum is taken to be $\infty$ whenever $b(z) = 0$.

*Proof.* We have
\[
\int a(z)\nu(dz) \geq \frac{\int (\inf_w \frac{a(w)}{b(w)})b(z)\nu(dz)}{\int b(z)\nu(dz)} = \inf_w \frac{a(w)}{b(w)}. \qed
\]

**Lemma C.4.** Regarding the density of the Beta($a,b$) distribution, where $a, b \geq 1$:

1. The density is unimodal if $a + b > 2$ and constant on $[0,1]$ if $a + b = 2$.

2. A global maximum of the density occurs at
\[
x^* = \begin{cases} 
\frac{a-1}{a+b-2} & a + b > 2 \\
0 & a + b = 2.
\end{cases}
\]
3. For $X \sim \text{Beta}(a, b)$ and any $\zeta > 0$, $\Pr(X \in [x^* - \zeta, x^* + \zeta]) \geq \min\{\zeta, 1\}$.

Proof. The first two statements are well-known. To show the last, assume WLOG that $x^* \leq 1 - x^*$. We handle three cases separately: $\zeta \leq x^*$, $x^* \in (x^*, 1 - x^*]$, and $\zeta > 1 - x^*$. For $\zeta > 1 - x^*$, $\Pr(X \in [x^* - \zeta, x^* + \zeta]) = 1$ so the result holds trivially.

For $\zeta \leq x^*$, letting $f(x)$ indicate the Beta$(a, b)$ density and using Lemma C.3 and the fact that $f(x)$ is monotonically nondecreasing for $x < x^*$ and monotonically nonincreasing for $x > x^*$,

$$\frac{\Pr(X \in [x^* - \zeta, x^* + \zeta])}{\Pr(X \not\in [x^* - \zeta, x^* + \zeta])} = \frac{\int_{x^* - \zeta}^{x^*} f(x)dx + \int_{x^*}^{x^* + \zeta} f(x)dx}{\int_{0}^{x^* - \zeta} f(x)dx + \int_{x^* + \zeta}^{1} f(x)dx} \geq \frac{f(x^* - \zeta)\zeta + f(x^* + \zeta)\zeta}{f(x^* - \zeta)(x^* - \zeta) + f(x^* + \zeta)(1 - x^* - \zeta)} \geq \min\left\{\frac{\zeta}{x^* - \zeta}, \frac{\zeta}{1 - x^* - \zeta}\right\} \geq \frac{\zeta}{1 - \zeta}.$$

So $\Pr(X \in [x^* - \zeta, x^* + \zeta]) \geq \zeta$.

Finally we address $\zeta \in (x^*, 1 - x^*]$. Then

$$\frac{\Pr(X \in [x^* - \zeta, x^* + \zeta])}{\Pr(X \not\in [x^* - \zeta, x^* + \zeta])} = \frac{\int_{x^*}^{x^* + \zeta} f(x)dx}{\int_{x^*}^{1} f(x)dx} \geq \frac{f(x^* + \zeta)\zeta}{f(x^* + \zeta)(1 - x^* - \zeta)} \geq \frac{\zeta}{1 - \zeta}$$

as desired. \qed

Lemma C.5. For any $\zeta > 0$ and any $K \in \mathbb{N}$ the following holds for any $D_1, D_2 \subset [0, 1]^K$ that are separated by Euclidean distance $\geq \zeta$. Let $X_k \sim \text{Beta}(a_k, b_k)$ for $k \in \{1, \ldots, K\}$, where $a_k, b_k \geq 1$. Assume that the mode $x^*_k = (x^*_1, \ldots, x^*_K)$ of the probability density function $f(x)$ of $X = (X_1, \ldots, X_K)$ satisfies $x^*_k \in D_1$, where $x^*_k$ for $k \in \{1, \ldots, K\}$ are the modes of the univariate Beta densities as defined in Lemma C.4. Then $\Pr(X \in D_1 \cup D_2) \geq \frac{\zeta}{2\sqrt{K}} \left(\frac{\zeta}{2\sqrt{K}}\right)^{K+1}$.

Proof. Consider the pdf $f(x)$ along any line segment originating at $x^*$. This density is monotonically nonincreasing with distance from $x^*$. For any set $D \subset [0, 1]^K$ one can calculate the integral $\int_D f(x)dx$ by first transforming to spherical coordinates, where the origin of the coordinate system is taken to be $x^*$. In this coordinate system let $\phi$ denote the $(K - 1)$-dimensional vector of angular coordinates, and $\rho \geq 0$ denote the radius, i.e. the distance
from $x^*$. Let $h(\rho, \phi)$ be the (invertible) function that maps from the spherical coordinates to the Euclidean coordinates. The Jacobian of the transformation $h$ takes the form $\rho^K g(\phi)$ for some function $g$. So for any $D \subset [0, 1]^K$ we can write

$$\int_D f(x) dx = \int_{h^{-1}(D)} f(h(\rho, \phi)) \rho^K g(\phi) d\rho d\phi.$$ 

In particular (using Lemma [C.3]),

$$\frac{\Pr(X \notin D_1 \cup D_2)}{\Pr(X \in D_2)} = \frac{\int_{h^{-1}((D_1 \cup D_2)^c)} f(h(\rho, \phi)) \rho^K g(\phi) d\rho d\phi}{\int_{h^{-1}(D_2)} f(h(\rho, \phi)) \rho^K g(\phi) d\rho d\phi}$$

$$= \frac{\int \left[ \frac{1}{h(\rho, \phi) \in (D_1 \cup D_2)^c} f(h(\rho, \phi)) \rho^K d\rho \right] g(\phi) d\phi}{\int \left[ \frac{1}{h(\rho, \phi) \in D_2} f(h(\rho, \phi)) \rho^K d\rho \right] g(\phi) d\phi}$$

$$\geq \inf_{\phi} \frac{\int_{h(\rho, \phi) \in (D_1 \cup D_2)^c} f(h(\rho, \phi)) \rho^K d\rho}{\int_{h(\rho, \phi) \in D_2} f(h(\rho, \phi)) \rho^K d\rho}$$

where we consider the ratio inside the infimum to be $= \infty$ if the denominator is zero. Then

$$\frac{\Pr(X \notin D_1 \cup D_2)}{\Pr(X \in D_2)} \geq \inf_{\phi} \left( \frac{\zeta}{2\sqrt{K}} \right)^K \frac{\int_{\zeta/2}^{\sqrt{K}} 1_{h(\rho, \phi) \in (D_1 \cup D_2)^c} f(h(\rho, \phi)) \rho^K d\rho}{\int_{\zeta/2}^{\sqrt{K}} 1_{h(\rho, \phi) \in D_2} f(h(\rho, \phi)) \rho^K d\rho}$$

(C.42)

For any fixed $\phi$ for which $0 \neq \int_{\zeta/2}^{\sqrt{K}} 1_{h(\rho, \phi) \in D_2} f(h(\rho, \phi)) d\rho$, there is some $\tilde{\rho}$ such that $h(\tilde{\rho}, \phi) \in D_2$. Since $x^* = h(0, \phi) \in D_1$ and since $D_1$ and $D_2$ are separated by distance $\zeta$, there must be an interval $[\rho_1(\phi), \rho_2(\phi)] \subset [0, \tilde{\rho}]$ of width at least $\zeta$ such that any $\rho \in [0, \rho_1(\phi)]$ satisfies $h(\rho, \phi) \notin D_2$ and any $\rho \in (\rho_1(\phi), \rho_2(\phi))$ satisfies $h(\rho, \phi) \in (D_1 \cup D_2)^c$. Using (C.42) and since $f(h(\rho, \phi))$ is monotonically nonincreasing in $\rho$,

$$\frac{\Pr(X \notin D_1 \cup D_2)}{\Pr(X \in D_2)} \geq \left( \frac{\zeta}{2\sqrt{K}} \right)^K \inf_{\phi} \frac{\int_{\rho_1(\phi)}^{\rho_2(\phi)} f(h(\rho, \phi)) d\rho}{\int_{\rho_1(\phi)}^{\rho_2(\phi)} f(h(\rho, \phi)) d\rho}$$

$$\geq \left( \frac{\zeta}{2\sqrt{K}} \right)^K \frac{\rho_2(\phi)}{\rho_1(\phi)}$$

$$\geq \left( \frac{\zeta}{2\sqrt{K}} \right)^K \frac{\rho_2(\phi)}{\rho_1(\phi)}$$

$$\geq \left( \frac{\zeta}{2\sqrt{K}} \right)^{K+1}.$$
Lemma C.6. For \( k \in \{1, \ldots , K\} \) let \( X_k \) \( \sim \) Beta\((a_k, b_k)\) where \( a_k, b_k \geq 1 \). Then for any set \( D \subset [0, 1]^K \) with positive Lebesgue measure \( (\lambda(D) > 0) \) and any \( d_3 > 1 \),
\[
\inf_{a_1, b_1, \ldots , a_K, b_K \in [1, d_3]} Pr(X \in D) > 0
\]
where \( X = (X_1, \ldots , X_K) \).

Proof. Since \( \lambda(D) > 0 \), there is some \( \zeta \in (0, 1/2) \) such that the set \( \tilde{D} = D \cap [\zeta, 1 - \zeta]^K \) satisfies \( \lambda(\tilde{D}) > 0 \). Letting \( f(x) \) indicate the density of any Beta\((a, b)\) distribution where \( a, b \in [1, d_3] \), and using Lemma C.4
\[
\frac{\inf_{x \in [\zeta, 1 - \zeta]} f(x)}{\sup_x f(x)} = \frac{\min\{f(\zeta), f(1 - \zeta)\}}{f\left(\frac{a-1}{a+b-2}\right)} \\
\geq \frac{\zeta^{a+b-2}(a+b-2)^{a+b-2}}{(a-1)^{a-1}(b-1)^{b-1}} \\
\geq \zeta^{a+b-2} \geq \zeta^{2d_3-2}.
\]

Now letting \( f(x) \) indicate the function on \( x \in [0, 1]^K \) that is the product of Beta\((a_k, b_k)\) densities where \( a_k, b_k \in [1, d_3] \),
\[
\inf_{x \in [\zeta, 1 - \zeta]^K} f(x) \geq \zeta^{K(2d_3-2)}.
\]
So
\[
\frac{Pr(X \in D)}{Pr(X \in D^c)} \geq \frac{Pr(X \in \tilde{D})}{Pr(X \in \tilde{D}^c)} \geq \frac{\lambda(\tilde{D}) \inf_{x \in [\zeta, 1 - \zeta]^K} f(x)}{(1 - \lambda(\tilde{D})) \sup_x f(x)} \geq \frac{\lambda(\tilde{D}) \zeta^{K(2d_3-2)}}{(1 - \lambda(\tilde{D}))} \tag{C.43}
\]
which is strictly positive and does not depend on \( \{a_k, b_k\}_{k=1}^K \). \( \square \)

Lemma C.7. Let \( X_k \) \( \sim \) Beta\((a_k, b_k)\) for \( k \in \{1, \ldots , Q\} \) where \( Q \in \mathbb{N} \) and \( a_k, b_k \geq 1 \). Also let \( x_k^* \) be the global mode of the density of Beta\((a_k, b_k)\) as defined in Lemma C.4. Let \( B(x, \delta) \) indicate the ball of radius \( \delta > 0 \) centered at a point \( x \in [0, 1]^Q \). Then for any fixed \( \delta > 0, d_3 \geq 1 \), and \( K \in \{1, \ldots , Q\} \),
\[
\inf_{a_k, b_k \in [1, d_3]: k=1, \ldots , K} \inf_{a_k, b_k \geq 1: k=K+1, \ldots , Q} \inf_{x \in [0, 1]^Q: x_k=x_k^*, k=K+1, \ldots , Q} Pr(X \in B(x, \delta)) > 0.
\]
Proof. Take a hypercube $H(x, \delta)$ centered at $x$ and with some fixed side length $2\delta_1 \in (0, 1]$ for which $H(x, \delta) \subset B(x, \delta)$. Then

$$\inf_{a_k, b_k \in [1, d_3]; k = 1, \ldots, K} \inf_{a_k, b_k \geq 1; k = K+1, \ldots, Q} \inf_{x \in [0, 1]^Q; x_k = x^*_k, k = K+1, \ldots, Q} \Pr(X \in B(x, \delta))$$

$$\geq \inf_{a_k, b_k \in [1, d_3]; k = 1, \ldots, K} \inf_{a_k, b_k \geq 1; k = K+1, \ldots, Q} \inf_{x \in [0, 1]^Q; x_k = x^*_k, k = K+1, \ldots, Q} \Pr(X \in H(x, \delta))$$

$$= \left[ \prod_{k=1}^K \inf_{a_k, b_k \in [1, d_3]} \inf_{x_k \in [0, 1]} \Pr(X_k \in [x_k - \delta_1, x_k + \delta_1]) \right] \prod_{k=K+1}^Q \inf_{a_k, b_k \geq 1} \Pr(X_k \in [x^*_k - \delta_1, x^*_k + \delta_1]).$$

(C.44)

By Lemma C.4, the second product in this expression is bounded below by $\delta_1^{Q-K}$. To bound the first product in (C.44) we will use the explicit lower bound (C.43) given in the proof of Lemma C.6 applied to the single variable $X_k$ where $k \in \{1, \ldots, K\}$. Here we take the set $D = [x_k - \delta_1, x_k + \delta_1] \cap [0, 1]$. Let $\zeta = \frac{\delta_1}{2}$ so that $\tilde{D} = D \cap \left[\frac{\delta_1}{2}, 1 - \frac{\delta_1}{2}\right]$. Noticing that $\lambda(\tilde{D}) \geq \frac{\delta_1}{2}$, the bound (C.43) gives

$$\frac{\Pr(X_k \in D)}{\Pr(X_k \in D^c)} \geq \left(\frac{\delta_1}{2}\right)^{1+\frac{2d_3-2}{1-\frac{\delta_1}{2}}} \geq \left(\frac{\delta_1}{2}\right)^{\frac{2(2d_3-1)}{2d_3}}.$$

So $\Pr(X_k \in D) \geq \left(\frac{\delta_1}{2}\right)^{(2d_3-1)}$; applying this method for each $k = 1, \ldots, K$ we have that

$$\inf_{a_k, b_k \in [1, d_3]; k = 1, \ldots, K} \inf_{a_k, b_k \geq 1; k = K+1, \ldots, Q} \inf_{x \in [0, 1]^Q; x_k = x^*_k, k = K+1, \ldots, Q} \Pr(X \in B(x, \delta))$$

$$\geq \left(\frac{\delta_1}{2}\right)^{K(2d_3-1)} \delta_1^{Q-K} > 0.$$

Proof of Proposition C.1. Recall the definition (Sec. 2.1) of $\beta_k$; we will take $\beta_{k,m} = 1$ for $k \in \{0, \ldots, w\}$ and $m \in \{1, 2\}$ for simplicity of exposition, although the results do not depend on this choice. Then the prior for $\theta_{0:w}$ is uniform: $\pi(\theta_{0:w}) \propto 1_{\{\theta_{0:w} \in [0, 1]^{w+1}\}}$.

The quantities $N(A^{(k)})$ and $N(A^c)$ only depend on $A$ via $C(A)$, due to (5.6). Consider
the conditional distribution \( \pi(\theta_{0:w} | C(A), S) \), which can be written as follows, using (2.3):

\[
\pi(\theta_{0:w} | C(A), S) \propto \pi(\theta_{0:w}, C(A), S) \propto \pi(\theta_{0:w}) \pi(C(A)) \pi(S | C(A), \theta_{0:w})
\]

\[
\propto \left[ \prod_{k=1}^{w} \prod_{m=1}^{2} \theta_{k,m}^{N(A(k))} \right] \prod_{m=1}^{2} \theta_{0,m}^{N(A^c)}
\]

\[
\propto \left[ \prod_{k=1}^{w} \text{Beta}(\theta_{k,1}; N(A(k))_1 + 1, N(A(k))_2 + 1) \right] \times
\]

\[
\text{Beta}(\theta_{0,1}; N(A^c)_1 + 1, N(A^c)_2 + 1).
\]  

(C.45)

where \( \text{Beta}(x; a, b) \) indicates the Beta density with parameters \( a, b \), evaluated at \( x \). By Lemma C.4, \( \pi(\theta_{0:w} | C(A), S) \) is a density with global maximum at \( \tilde{\theta}_{0:w} \) where

\[
\tilde{\theta}_{k,1} = \begin{cases} 
\frac{N(A(k))_1}{N(A(k))} & |N(A(k))| > 0 \\
0 & \text{else} 
\end{cases} 
\quad k \in \{1, \ldots, w\}  
\]  

(C.46)

\[
\tilde{\theta}_{0,1} = \begin{cases} 
\frac{N(A^c)_1}{N(A^c)} & |N(A^c)| > 0 \\
0 & \text{else.} 
\end{cases}
\]

To complete the notation define \( \tilde{\theta}_{k,2} = 1 - \tilde{\theta}_{k,1} \) for \( k \in \{0, \ldots, w\} \).

By (C.45) and since \( |N(A^c)| = L - \sum_{k=1}^{w} |N(A(k))| \), we have that \( \pi(\theta_{0:w} | C(A), S) \) only depends on \( C(A) \) via \( \tilde{\theta}_{0:w} \) and \( |N(A^{(1)})| = |N(A^{(2)})| = \ldots = |N(A^{(w)})| \). So

\[
\pi\left(\theta_{0:w} | \tilde{\theta}_{0:w}, |N(A^{(1)})|, S\right)
\]

\[
\propto \left[ \prod_{k=1}^{w} \text{Beta}\left(\theta_{k,1}; \tilde{\theta}_{k,1}|N(A^{(1)})| + 1, \tilde{\theta}_{k,2}|N(A^{(1)})| + 1\right) \right] \times
\]

\[
\text{Beta}\left(\theta_{0,1}; \tilde{\theta}_{0,1}(L - w|N(A^{(1)})|) + 1, \tilde{\theta}_{0,2}(L - w|N(A^{(1)})|) + 1\right).
\]  

(C.47)

Using Lemma C.4 and regardless of the value of \( |N(A^{(1)})| \), \( \pi\left(\theta_{0:w} | \tilde{\theta}_{0:w}, |N(A^{(1)})|, S\right) \) has a global maximum at \( \tilde{\theta}_{0:w} \).

For our analysis the only relevant quantities regarding \( C(A) \in \bar{X} \) will be \( \tilde{\theta}_{0:w} \) and \( |N(A^{(1)})| \), so we define \( F_1, F_2 \subset \bar{X} \) more conveniently as sets of possible values of \( (\tilde{\theta}_{0:w}, |N(A^{(1)})|) \), i.e. values that arise from some state \( C(A) \in \bar{X} \). We will define \( F_1 \) to be a particular set for which there is some constant \( d_4 > 0 \) satisfying

\[
\min_{(\theta_{0:w}, |N(A^{(1)})|) \not\in F_1} \frac{\Pr\left(\theta_{0:w} \not\in B_1 \cup B_2 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S\right)}{\Pr\left(\theta_{0:w} \in B_1 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S\right)} \geq d_4.
\]  

(C.48)

So \( F_1 \subset \bar{X} \) is associated with \( B_1 \subset [0, 1]^{w+1} \) in the sense that it (informally speaking) contains all the values of \( (\tilde{\theta}_{0:w}, |N(A^{(1)})|) \) for which \( \Pr\left(\theta_{0:w} \in B_1 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S\right) \) is much larger...
one iteration of $\hat{\theta}_{0:w}$ we define $E_1$ be the set of all points $x \in [0,1]^{w+1}$ that are within distance $\epsilon/3$ of the set $B_1$, and let $E_2$ be the set of all points that are separated by distance $\epsilon/3$ of the set $B_2$. This is illustrated in Web Appendix Figure 1. Then $E_1$ and $E_2$ are separated by distance $\epsilon_1 \triangleq \epsilon/3$. Let $d_{5} \triangleq \frac{w+1}{\epsilon_{1}}$, since $B_1, B_2 \subset [0,1]^{w+1}$ are separated by distance $\epsilon$, we have that $\epsilon \leq \sqrt{w+1}$ and so
\[ d_{5} = \frac{w+1}{\epsilon/3} > \frac{w+1}{\sqrt{w+1}} > 1. \] (C.49)

Also define
\[ V \triangleq \left\{ (\tilde{\theta}_{0:w}, |N(A^{(1)})|) : \max\{|N(A^{(1)})|, |N(A^{c})|/w\} > d_{5} \right\} \] (C.50)
\[ \cap \left\{ (\tilde{\theta}_{0:w}, |N(A^{(1)})|) : \exists \theta_{0} \in [0,1] \text{ s.t. } (\theta_{0}, \tilde{\theta}_{1:w}) \in (E_1 \cup E_2)^{c} \text{ then } |N(A^{c})|/w > d_{5} \right\} \]
\[ \cap \left\{ (\tilde{\theta}_{0:w}, |N(A^{(1)})|) : \exists \theta_{1:w} \in [0,1]^{w} \text{ s.t. } (\theta_{0}, \theta_{1:w}) \in (E_1 \cup E_2)^{c} \text{ then } |N(A^{(1)})| > d_{5} \right\} \]
\[ F_{j} \triangleq \left\{ (\tilde{\theta}_{0:w}, |N(A^{(1)})|) \in V : \tilde{\theta}_{0:w} \in E_{j} \right\}, \quad j \in \{1, 2\}. \]

First we show that it is not possible to move from any state $(\tilde{\theta}^{1}_{0:w}, |N(A^{(1)})|^{1}) \in F_{1}$ to any state $(\tilde{\theta}^{2}_{0:w}, |N(A^{(1)})|^{2}) \in F_{2}$ in one iteration of $\tilde{T}$. Since $\tilde{\theta}^{1}_{0:w} \in E_{1}$ and $\tilde{\theta}^{2}_{0:w} \in E_{2}$ satisfy $|\tilde{\theta}^{1}_{0:w} - \tilde{\theta}^{2}_{0:w}| \geq \epsilon_{1}$, we have that $\exists k \in \{0, \ldots, w\}$ such that $|\tilde{\theta}_{k,1}^{1} - \tilde{\theta}_{k,1}^{2}| \geq \frac{\epsilon_{1}}{w+1} = \frac{1}{d_{5}}$. We handle the four cases: 1. where $|N(A^{(1)})|^{1} \leq d_{5}$; 2. where $|N(A^{(1)})|^{1}/w \leq d_{5}$; 3. where $|N(A^{(1)})|^{1} > d_{5}$, $|N(A^{c})|^{1}/w > d_{5}$ and $k > 0$; 4. where $|N(A^{(1)})|^{1} > d_{5}$, $|N(A^{c})|^{1}/w > d_{5}$ and $k = 0$. We assume that it is it is possible to move from $(\tilde{\theta}^{1}_{0:w}, |N(A^{(1)})|^{1})$ to $(\tilde{\theta}^{2}_{0:w}, |N(A^{(1)})|^{2})$ in one iteration of $\tilde{T}$, and find a contradiction. We use the fact that, by (2.6) and (5.9), in one iteration of $\tilde{T}$ the vector $N(A^{(k)})$ can only change by incrementing or decrementing a single element by one, and so $|N(A^{(k)})| = |N(A^{(1)})|$ can only increase or decrease by one. Also, the vector $N(A^{c})$ can only change by either incrementing its elements by a total of $w$, which increases $|N(A^{c})|$ by $w$, or decrementing its elements by a total of $w$, which decreases $|N(A^{c})|$ by $w$. 

Figure 1: An illustration of the proof.
First take the case where $|N(A^{(1)})|^1 > d_5$, $|N(A^{c})|^1/w > d_5$ and $k > 0$. By \((C.49)\), $|N(A^{(1)})|^1 > 1$, so $|N(A^{(1)})|^2 > 0$. By \((C.46)\),

$$|\tilde{\theta}_{k,1}^1 - \tilde{\theta}_{k,1}^2| = \left| \frac{N(A^{(k)})_1}{|N(A^{(k)})|^1} - \frac{N(A^{(k)})_2}{|N(A^{(k)})|^2} \right|.$$  \hfill \(\text{(C.51)}\)

Also, we claim that this is bounded above by $\frac{1}{|N(A^{(k)})|^1}$. In the case where $N(A^{(k)})_1^2 = N(A^{(k)})_1 + \delta$ and $\delta \in \{-1, 1\}$, we have $N(A^{(k)})_1^2 \geq 0$ so $N(A^{(k)})_1 \geq -\delta$ and thus

$$|\tilde{\theta}_{k,1}^1 - \tilde{\theta}_{k,1}^2| = \left| \frac{N(A^{(k)})_1^2}{|N(A^{(k)})|^1} - \frac{N(A^{(k)})_1 + \delta}{|N(A^{(k)})|^1 + \delta} \right| = \left( \frac{|N(A^{(k)})|^1 - N(A^{(k)})_1^1}{|N(A^{(k)})|^1 + \delta} \right) |\delta| \leq \frac{1}{|N(A^{(k)})|^1} = \frac{1}{|N(A^{(k)})|^1}.$$  \hfill \(\text{(C.52)}\)

In the case where $N(A^{(k)})_2^2 = N(A^{(k)})_2 + \delta$ and $\delta \in \{-1, 1\}$, by using the fact that $|\tilde{\theta}_{k,1}^1 - \tilde{\theta}_{k,1}^2| = |\tilde{\theta}_{k,2}^1 - \tilde{\theta}_{k,2}^2|$ and applying the above argument we still obtain the upper bound $\frac{1}{|N(A^{(k)})|^1}$.

Combining with \((C.51)\) we have

$$|\tilde{\theta}_{k,1}^1 - \tilde{\theta}_{k,1}^2| \leq \frac{1}{|N(A^{(k)})|^1} < \frac{1}{d_5}$$  \hfill \(\text{(C.53)}\)

which is a contradiction (by the definition of $k$).

Now take the case where $|N(A^{(1)})|^1 \leq d_5$. Then by \((C.50)\) we must have $|N(A^{c})|^1/w > d_5$. Also, $\tilde{\theta}_{0,w}^1 \in E_1$ and there is no $\theta_{1,w}^1$ such that $(\tilde{\theta}_{0,w}^1, \theta_{1,w}^1) \in (E_1 \cup E_2)^c$, so $(\tilde{\theta}_{0,w}^1, \theta_{1,w}^1) \in E_1$. Therefore the Euclidean distance between $(\tilde{\theta}_{0,w}^1, \theta_{1,w}^1) \in E_1$ and $(\tilde{\theta}_{0,w}^2, \theta_{1,w}^2) \in E_2$ is $\geq \epsilon_1$. This implies $|\tilde{\theta}_{0,w}^1 - \tilde{\theta}_{0,w}^2| \geq \epsilon_1 > \frac{1}{d_5}$. However, by \((C.49)\), $|N(A^{c})|^1 > d_5 w > w$, so $|N(A^{c})|^2 > 0$. Then by \((C.46)\),

$$|\tilde{\theta}_{0,w}^1 - \tilde{\theta}_{0,w}^2| = \left| \frac{N(A^{c})_1}{|N(A^{c})|^1} - \frac{N(A^{c})_2}{|N(A^{c})|^2} \right|.$$  \hfill \(\text{(C.54)}\)

Also, we claim that this is bounded above by $\frac{w}{|N(A^{c})|^1}$. In the case where $N(A^{c})_1^2 = N(A^{c})_1 + \delta$ and $N(A^{c})_2^2 = N(A^{c})_2 + w - \delta$ for $\delta \in \{0, \ldots, w\}$,

$$|\tilde{\theta}_{0,w}^1 - \tilde{\theta}_{0,w}^2| = \left| \frac{N(A^{c})_1^2}{|N(A^{c})|^1} - \frac{N(A^{c})_1 + \delta}{|N(A^{c})|^1 + w} \right| = \left| \frac{wN(A^{c})_1}{|N(A^{c})|^1 (|N(A^{c})|^1 + w)} \right| \leq \frac{w}{|N(A^{c})|^1}.$$  \hfill \(\text{(C.55)}\)

In the case where $N(A^{c})_1^2 = N(A^{c})_1 - \delta$ and $N(A^{c})_2^2 = N(A^{c})_2 - w + \delta$ for $\delta \in \{0, \ldots, w\}$,

$$|\tilde{\theta}_{0,w}^1 - \tilde{\theta}_{0,w}^2| = \left| \frac{N(A^{c})_1^2}{|N(A^{c})|^1} - \frac{N(A^{c})_1 - \delta}{|N(A^{c})|^1 - w} \right| = \left| \frac{-wN(A^{c})_1^1 + \delta|N(A^{c})|^1}{|N(A^{c})|^1 (|N(A^{c})|^1 - w)} \right|.$$  \hfill \(\text{(C.56)}\)
This is largest when \( \delta \in \{0, w\} \). Note that \( N(A^c)^1 \geq 0 \) and \( N(A^c)^2 \geq 0 \) so \( N(A^c)^1 \geq \delta \) and \( N(A^c)^2 \geq w - \delta \). Using \((C.53)\), when \( \delta = 0 \) we have \( N(A^c)^2 \geq w \) and

\[
|\hat{\theta}_{0,1}^1 - \hat{\theta}_{0,1}^2| = \frac{w N(A^c)^1}{\left|N(A^c)^1\right| - \left|N(A^c)^1 - w\right|} = \frac{w \left|N(A^c)^1 - N(A^c)^2\right|}{\left|N(A^c)^1\right| - \left|N(A^c)^1 - w\right|} \leq \frac{w}{\left|N(A^c)^1\right|}.
\]

When \( \delta = w \) we have \( N(A^c)^1 \geq w \) and (using \((C.53)\))

\[
|\hat{\theta}_{0,1}^1 - \hat{\theta}_{0,1}^2| = \frac{w \left|N(A^c)^1 - N(A^c)^1\right|}{\left|N(A^c)^1\right| - \left|N(A^c)^1 - w\right|} \leq \frac{w}{\left|N(A^c)^1\right|}.
\]

as claimed. So \( |\hat{\theta}_{0,1}^1 - \hat{\theta}_{0,1}^2| \leq \frac{w}{\left|N(A^c)^1\right|} < \frac{1}{d_5} \), which is a contradiction. The case where \( \left|N(A^1)\right| > d_5, \left|N(A^c)\right|/w > d_5 \) and \( k = 0 \), and the case where \( \left|N(A_c)\right| \leq d_5w \), lead to contradictions analogously to the two cases handled above. So it is not possible to move from \((\hat{\theta}_{0:w}^1, \left|N(A^1)\right|)\) to \((\hat{\theta}_{0:w}^2, \left|N(A^1)\right|)\) in one iteration of \( \bar{T} \).

Next we show \((C.48)\). By Lemma \((C.5)\), \((C.47)\), \((C.50)\), and \( B_2 \subset E_2 \), there is some \( d_6 > 0 \) that depends only on \( w \) such that

\[
\min_{\hat{\theta}_{0:w} \in E_2} \min_{\left|N(A^1)\right|} \frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)} \geq \min_{\hat{\theta}_{0:w} \in E_2} \min_{\left|N(A^1)\right|} \frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)} \geq d_6. \tag{C.54}
\]

Also, by Lemma \((C.5)\) and \( E_1 \setminus B_1 \subset (B_1 \cup B_2)^c \), there exists \( d_7 > 0 \) that depends only on \( w \) such that

\[
\min_{\hat{\theta}_{0:w} \in (E_1 \cup E_2)^c} \min_{\left|N(A^1)\right|} \frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)} \geq \min_{\hat{\theta}_{0:w} \in (E_1 \cup E_2)^c} \min_{\left|N(A^1)\right|} \frac{\Pr(\theta_{0:w} \notin B_1 \cup E_2 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)} \geq d_7. \tag{C.55}
\]

Additionally, by Lemma \((C.6)\)

\[
\exists d_8 > 0 \text{ such that } \min_{\hat{\theta}_{0:w} \in (E_1 \cup E_2)^c} \min_{\left|N(A^1)\right|} \frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)} \geq \min_{\hat{\theta}_{0:w} \in (E_1 \cup E_2)^c} \min_{\left|N(A^1)\right|} \frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \hat{\theta}_{0:w}, \left|N(A^1)\right|, S)} > d_8. \tag{C.56}
\]
Also, for any $\theta_{1:w}$ such that $(\tilde{\theta}_0, \theta_{1:w}) \in (E_1 \cup E_2)^c$, a ball of radius $\epsilon_1/2 = \epsilon/6$ centered at $(\tilde{\theta}_0, \theta_{1:w})$ is entirely contained in $(B_1 \cup B_2)^c$. By Lemma C.7, $\exists d_9 > 0$

$$\min_{\tilde{\theta}_0, w : (\tilde{\theta}_0, \theta_{1:w}) \in (E_1 \cup E_2)^c} \min_{w : |N(A^{(1)})| \leq d_5} \frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S)} \geq d_9.$$  \hspace{1cm} (C.57)

By the analogous argument, $\exists d_{10} > 0$

$$\min_{\tilde{\theta}_0, w : (\tilde{\theta}_0, \theta_{1:w}) \in (E_1 \cup E_2)^c} \min_{w : |N(A^{(1)})|/w \leq d_5} \frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S)} \geq d_{10}. \hspace{1cm} (C.58)$$

By (C.50),

$$(F_1 \cup F_2)^c = \left\{ (\tilde{\theta}_{0:w}, |N(A^{(1)})|) : \tilde{\theta}_{0:w} \in (E_1 \cup E_2)^c \vee \max\{|N(A^{(1)})|, |N(A^{(1)})|/w\} \leq d_5 \right\}$$

$$\cup \left\{ (\tilde{\theta}_{0:w}, |N(A^{(1)})|) : |N(A^{(1)})|/w \leq d_5 \land \exists \theta_0 \text{ s.t. } (\theta_0, \tilde{\theta}_{1:w}) \in (E_1 \cup E_2)^c \right\}$$

$$\cup \left\{ (\tilde{\theta}_{0:w}, |N(A^{(1)})|) : |N(A^{(1)})| \leq d_5 \land \exists \theta_{1:w} \text{ s.t. } (\theta_0, \theta_{1:w}) \in (E_1 \cup E_2)^c \right\}$$

and due to (C.55)-(C.58) we have

$$\min_{\tilde{\theta}_{0:w} \mid |N(A^{(1)})| \in (F_1 \cup F_2)^c} \frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S)} \geq \min\{d_7, d_8, d_9, d_{10}\} > 0.$$  \hspace{1cm} (C.59)

Combining this result with (C.54) yields (C.48).

Now we prove the second part of Proposition C.1. Using Lemma C.3 and (C.48),

$$\frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid S, (\tilde{\theta}_{0:w}, |N(A^{(1)})|) \in F_2)}{\Pr(\theta_{0:w} \in B_1 \mid S, (\tilde{\theta}_{0:w}, |N(A^{(1)})|) \in F_2)} = \frac{\sum_{\tilde{\theta}_{0:w} \mid |N(A^{(1)})| \in F_2} \Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S) \pi(|N(A^{(1)})|, \tilde{\theta}_{0:w} \mid S)}{\sum_{\tilde{\theta}_{0:w} \mid |N(A^{(1)})| \in F_2} \Pr(\theta_{0:w} \in B_1 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S) \pi(|N(A^{(1)})|, \tilde{\theta}_{0:w} \mid S)} \geq \min_{\tilde{\theta}_{0:w} \mid |N(A^{(1)})| \in F_2} \frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S)}{\Pr(\theta_{0:w} \in B_1 \mid \tilde{\theta}_{0:w}, |N(A^{(1)})|, S)} \geq d_4. \hspace{1cm} (C.59)$$

Analogously,

$$\frac{\Pr(\theta_{0:w} \notin B_1 \cup B_2 \mid S, (\tilde{\theta}_{0:w}, |N(A^{(1)})|) \notin F_1 \cup F_2)}{\Pr(\theta_{0:w} \in B_1 \mid S, (\tilde{\theta}_{0:w}, |N(A^{(1)})|) \notin F_1 \cup F_2)} \geq d_4. \hspace{1cm} (C.60)$$
Then by symmetry we have
\[
\frac{\Pr(\theta_{0w} \not\in B_1 \cup B_2 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_1 \cup F_2)}{\Pr(\theta_{0w} \in B_2 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_1 \cup F_2)} \geq d_4
\]
which combined with \((C.60)\) yields
\[
\Pr(\theta_{0w} \not\in B_1 \cup B_2 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_1 \cup F_2) \geq \frac{d_4}{2 + d_4} > 0. \tag{C.61}
\]
Again using Lemma \([C.3]\),
\[
\frac{\Pr(\theta_{0w} \not\in B_1 \cup B_2 \mid S)}{\Pr(\theta_{0w} \in B_1 \mid S)} \geq \min \left\{ \frac{\Pr(\theta_{0w} \not\in B_1 \cup B_2 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \in F_2)}{\Pr(\theta_{0w} \in B_1 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \in F_2)}, \frac{\Pr(\theta_{0w} \not\in B_1 \cup B_2 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_2)}{\Pr(\theta_{0w} \in B_1 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_2)} \right\}
\]
Using this fact and \((C.59)\) and since the ratios in \((5.25)\) are exponentially decreasing in \(L\),
\[
\frac{\Pr(\theta_{0w} \not\in B_1 \cup B_2 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_2)}{\Pr(\theta_{0w} \in B_1 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_2)} \tag{C.62}
\]
is also exponentially decreasing in \(L\). Also, using \((C.60)-(C.61)\),
\[
\frac{\Pr(\theta_{0w} \in B_1 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_2)}{\Pr(\theta_{0w} \not\in B_1 \cup B_2 \mid S, (\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_2)} \leq \frac{2 + d_4}{d_4}
\]
Combining with the fact that \((C.62)\) is exponentially decreasing in \(L\),
\[
\frac{\Pr((\tilde{\theta}_{0w}, |N(A^{(1)})|)) \in F_1 \mid S)}{\Pr(\tilde{\theta}_{0w}, |N(A^{(1)})|) \not\in F_1 \mid S)} \quad \text{is also exponentially decreasing in } L.
\]
By symmetry, \(\Pr((\tilde{\theta}_{0w}, |N(A^{(1)})|)) \in F_2 \mid S)} \) decreases exponentially in \(L\), proving Proposition \(C.1\). \(\square\)