The Annals of Statistics
2011, Vol. 39, No. 1, 439–473
DOI: 10.1214/10-AOS822

EFFECTS OF STATISTICAL DEPENDENCE ON MULTIPLE TESTING UNDER A HIDDEN MARKOV MODEL

BY ZHIYI CHI[1]

University of Connecticut

The performance of multiple hypothesis testing is known to be affected by the statistical dependence among random variables involved. The mechanisms responsible for this, however, are not well understood. We study the effects of the dependence structure of a finite state hidden Markov model (HMM) on the likelihood ratios critical for optimal multiple testing on the hidden states. Various convergence results are obtained for the likelihood ratios as the observations of the HMM form an increasing long chain. Analytic expansions of the first and second order derivatives are obtained for the case of binary states, explicitly showing the effects of the parameters of the HMM on the likelihood ratios.

1. Introduction. Statistical dependence in data poses a challenge to multiple hypothesis testing. Under the framework of the false discovery rate (FDR), many efforts have been made to establish the control of FDR under dependence [5, 14, 25, 27, 29]. Meanwhile, many empirical and analytical works have described the effects of dependence on the outputs of multiple tests [12, 16, 22, 23]. However, in what way the dependence impacts multiple testing is not well understood.

A useful model that incorporates tractable dependence in multiple testing is the hidden Markov model (HMM) [27]. In the model, the nulls are organized as H_t, where the index t takes integer values. Each H_t is associated with a random variable that determines whether the null is true or false. The random variables form a Markov chain but are hidden and unobservable. Instead, the observations X_t each is a many-to-one transform of the hidden variable corresponding to H_t. In the context of multiple testing, it will be useful to treat the hidden variable as consisting of two parts, η_t and Z_t. On the one hand, η_t encodes the "true identity," or state of the signal associated with H_t and in general can take two or more possible values. On the other, Z_t acts as the noise that blurs or distorts the signal. Then X_t can be thought of as the result of a deterministic interaction between η_t and Z_t.

To understand the role of dependence in the multiple tests on the nulls, the "oracle" approach assumes the parameters in the HMM are known and explores what amounts to an optimal testing procedure. The advantage of this approach is

Received April 2009; revised March 2010.

[1] Supported in part by NSF Grant DMS-07-06048 and NIH MH 68028.
AMS 2000 subject classifications. 62H15, 62M02.
Key words and phrases. HMM, multiple hypothesis testing, FDR, contraction, nonlinear filtering.

that it can reveal effects purely due to dependence, without confounding with effects due to specific parameter estimation methods. Suppose the observations are $X_{-m}, \ldots, X_n$. With the parameters being known, for each null H_t, the conditional likelihood $\mathsf{Pr}\{H_t \text{ is true} \mid X_{-m}, \ldots, X_n\}$ can be computed. The importance of the conditional likelihood for multiple testing has been shown in various contexts [6, 13, 21, 26, 27]. For the HMM, [27] shows that under a certain loss function, an optimal procedure is to reject H_t if and only if the corresponding conditional likelihood is small enough. The loss function is a linear combination of the numbers of Types I and II errors and is related to the FDR. The importance of the conditional likelihood can also be argued directly based on the FDR criterion, and in fact without particular assumption on dependence; see the Appendix.

In view of the role of the conditional likelihood, our aim is to investigate how it is affected by the parameters of the HMM. The parameters can be divided into two types, respectively, characterizing the dependence among η_t and the "strength" of useful signals. In addition, the conditional likelihood also depends on how η_t and Z_t interact. The next example illustrates what role may be expected for these factors.

EXAMPLE 1.1. Suppose the states η_t are equal to $\mathbf{1}\{H_t \text{ is false}\}$ and form a stationary Markov chain with transition probabilities $q_{ij} = \mathsf{Pr}\{\eta_t = j \mid \eta_{t-1} = i\} > 0$; moreover, conditional on $\eta = (\eta_t)$, X_t are independent $\sim N(\varepsilon\eta_t, 1)$ with $\varepsilon > 0$. Write $X_t = Z_t + \varepsilon\eta_t$. Then (Z_t, η_t) form a hidden Markov chain, with Z_t i.i.d. $\sim N(0, 1)$. The strength of the signals is measured by ε, the interaction between the noise Z_t and η_t is additive, such that $X_t = \varphi(Z_t, \varepsilon\eta_t)$ with $\varphi(z, \vartheta) = z + \vartheta$.

In many cases, the observations form a long chain $X_{-m}, \ldots, X_n$, with $m, n \gg 1$, so the effect of the parameters can be studied through the properties of

$$\mathsf{Pr}\{\eta_t = 0 \mid X\} = \lim_{m,n\to\infty} \mathsf{Pr}\{\eta_t = 0 \mid X_{-m}, \ldots, X_n\}$$

for each t, where $X = (X_t, t \in \mathbb{Z})$. Since $\mathsf{Pr}\{\eta_t = 0 \mid X_{-m}, \ldots, X_n\}$ form a martingale for any increasing sequence of m and n, the (almost sure) existence of the limit is guaranteed. However, this says nothing about how the limit depends on ε and q_{ij}. To get some insight, consider instead the likelihood ratios

$$\frac{\mathsf{Pr}\{\eta_t = 1 \mid X\}}{\mathsf{Pr}\{\eta_t = 0 \mid X\}} = \frac{1}{\mathsf{Pr}\{\eta_t = 0 \mid X\}} - 1,$$

which turn out to be a little more convenient to study. Regarding them as functions of ε, we next consider their Taylor expansions. In principle, the likelihood ratios can be expanded around any value of ε. Since large values of $|\varepsilon|$ correspond to strong signals whose detection is easy, we shall expand around $\varepsilon = 0$ to get insight into the case where the strength of signal ranges from being weak to moderate.

Without loss of generality, consider the likelihood ratio for η_0. Since η is stationary, $\Pr\{\eta_{t-1} = j \mid \eta_t = i\} = q_{ij}$. By the Bayes rule and Markov property,

$$\begin{aligned}
&\Pr\{\eta_0 = a \mid X_{-m}, \dots, X_n\} \\
&\quad \propto P(a) \sum_{\substack{\sigma_{-m},\dots,\sigma_n \\ \sigma_0 = a}} \exp\left\{-\frac{1}{2}\sum_{t=-m}^{n}(Z_t + \varepsilon\eta_t - \varepsilon\sigma_t)^2\right\} \prod_{t=0}^{n-1} q_{\sigma_t\sigma_{t+1}} \prod_{t=0}^{m-1} q_{\sigma_{-t}\sigma_{-t-1}}
\end{aligned}$$

for $a = 0, 1$, where $P(a) = \Pr\{\eta_0 = a\}$. Then, formally, one can get

$$\begin{aligned}
\frac{d}{d\varepsilon}\left[\ln\frac{\Pr\{\eta_0 = 1 \mid X\}}{\Pr\{\eta_0 = 0 \mid X\}}\right]_{\varepsilon=0} &= \lim_{m,n\to\infty}\frac{d}{d\varepsilon}\left[\ln\frac{\Pr\{\eta_0 = 1 \mid X_{-m},\dots,X_n\}}{\Pr\{\eta_0 = 0 \mid X_{-m},\dots,X_n\}}\right]_{\varepsilon=0} \\
&= \sum_{t=-\infty}^{\infty} Z_t[\Pr\{\eta_t = 1 \mid \eta_0 = 1\} - \Pr\{\eta_t = 1 \mid \eta_0 = 0\}] \\
&= \sum_{t=-\infty}^{\infty} r^{|t|} Z_t,
\end{aligned}$$

where $r \neq 1$ is one of the two eigenvalues of the matrix (q_{ij}), the other being 1.

We shall refer to the above conditional likelihood ratio as the full likelihood ratio (FLR), as it is based on the entire X. On the other hand, if the information of the dependence (i.e., q_{ij}) is not available, but the values of all other parameters are known, including $P(a)$, then the likelihood ratio would have to be evaluated as

$$\frac{\Pr\{\eta_0 = 1 \mid X_0\}}{\Pr\{\eta_0 = 0 \mid X_0\}} = \frac{P(1)}{P(0)}\frac{f(X_0 - \varepsilon)}{f(X_0)} = \frac{P(1)}{P(0)}\exp\left\{\varepsilon Z_0 + \frac{\varepsilon^2}{2}(2\eta_0 - 1)\right\},$$

where $f(x)$ is the density of $N(0, 1)$. We shall refer to this conditional likelihood ratio as the local likelihood ratio (LLR). It then can be seen that

$$\ln\frac{\text{FLR}}{\text{LLR}} = \varepsilon\sum_{t\neq 0} r^{|t|} Z_t + o(\varepsilon).$$

Thus, at the first order, the dependence in η merely adds noise but no "net effect," regardless of the actual values of η. If there is any state-dependent effect, it should be reflected in a higher order term of ε. To see if this is the case, take the second order derivative in ε. Again, the calculation can be done formally. To evaluate the state-dependent net effect, proceed with

$$\begin{aligned}
\mathsf{E}[(\ln \text{FLR})''_{\varepsilon=0} \mid \eta_0] &= \lim_{m,n\to\infty}\mathsf{E}\left[\frac{d^2}{d\varepsilon^2}\left[\ln\frac{\Pr\{\eta_0 = 1 \mid X_{-m},\dots,X_n\}}{\Pr\{\eta_0 = 0 \mid X_{-m},\dots,X_n\}}\right]_{\varepsilon=0} \,\Big|\, \eta_0\right] \\
&= (2\eta_0 - 1)\sum_t r^{|2t|},
\end{aligned}$$

giving

$$\mathsf{E}\left[\left(\ln\frac{\mathrm{FLR}}{\mathrm{LLR}}\right)''_{\varepsilon=0} \,\Big|\, \eta_0\right] = (2\eta_0 - 1)\sum_{t\neq 0} r^{2|t|}.$$

It follows that, comparing to $\ln \mathrm{LLR}$, if $\eta_0 = 1$, on average $\ln \mathrm{FLR}$ is larger, making H_0 more likely to be (correctly) rejected, whereas if $\eta_0 = 0$, it is smaller, making H_0 less likely to be (falsely) rejected.

So far the expansions are expressed in terms of the unobservable Z_t. One question is whether similar expansions in terms of the observable X_t can be obtained. As will be seen in Section 3.2, this is possible after we get more information on higher order derivatives.

From the expansions, the effect of the dependence in η on the likelihood ratio is apparent. In both the first and second order derivatives, the effect is determined by r. In particular, when $r = 0$, η_t are i.i.d. and FLR is equal to LLR. Consistent with this, the derivatives of the difference between the two ratios become 0.

As the example, the rest of the paper develops Taylor's expansion in terms of ε for the FLRs $\mathsf{Pr}\{H_t \text{ is false} \mid X\}/\mathsf{Pr}\{H_t \text{ is true} \mid X\}$ to study the effects of dependence structure of HMM. The differentiation involved in the expansion should be interpreted as follows. During the differentiation, both the signal η and noise Z are fixed. As the strength ε of the signal varies, the observed values X_t become functions of ε. The likelihood ratio is affected by ε in two ways: not only the value of X_t is changed, but also the parametric form of the density function of X_t. Both have to be taken into account in the derivatives.

Several issues need to be addressed. First, we have only considered a stationary process of the signals η. In applications, it is useful to consider nonstationary η that has time-dependent transition probabilities. Moreover, it is useful to consider various types of interactions between η_t and Z_t besides the additive one.

Second, in Example 1.1, each η_t is binary, indicating whether a null is true or false. For more generality, one can assume a finite state Markov chain, such that a subset of the states are associated with true nulls and the rest with false nulls. Even for a binary process, it can be useful to reformulate it as a multistate Markov chain. For example, let η be a second order binary Markov chain, that is, $\mathsf{Pr}\{\eta_t \mid \eta_s, s < t\} = \mathsf{Pr}\{\eta_t \mid \eta_{t-1}, \eta_{t-2}\}$. Then one can define a first order Markov chain $\tilde{\eta}$ by $\tilde{\eta}_t = (\eta_{t-1}, \eta_t)$. If $\eta_t = \mathbf{1}\{H_t \text{ is false}\}$, then in terms of $\tilde{\eta}$, $(0, 0)$ and $(1, 0)$ are states associated with true nulls, and $(0, 1)$ and $(1, 1)$ are states associated with false nulls.

Third, in Example 1.1, limit operation, differentiation, and expectation are freely interchanged for $\mathsf{Pr}\{\eta_t \mid X_{-m}, \dots, X_n\}$ for fixed t. This has to be justified. Note that the likelihood bears similarity to $\mathsf{Pr}\{\eta_n \mid X_0, \dots, X_n\}$, a quantity extensively studied in the literature on nonlinear filtering and related issues [1–3, 7–11, 15, 17–20, 28]. As in most of the cited works, in this paper, convergence results are established using geometric contraction. On the other hand, in those works, the goal is to establish weak convergence of the conditional probability for η_n under the assumption of stationary transition probabilities. As seen in Example 1.1, the convergence of the conditional probability for η_t follows from the martingale convergence. So instead, the goal here is to establish convergence for the derivatives of the conditional likelihood with arbitrary transition probabilities.

The rest of the paper proceeds as follows. In Section 2, a HMM is set up in the context of multiple testing and then various convergence results on the likelihood ratio are stated. In Section 3, the likelihood ratio for a first order HMM with binary states is considered in more detail, which allows more explicit expressions for the first and second derivatives of the likelihood ratio. Several examples are given, with Example 1.1 being a special case. Theoretical details are provided in Section 4.

2. Main results.

2.1. *A HMM setup.* Let $\eta = \{\eta_t, t \in \mathbb{Z}\}$ be a finite state process, such that the state space $\mathcal{H}$ is partitioned into $\mathcal{H}_0$ and $\mathcal{H}_1$, with states in $\mathcal{H}_0$ being associated with true nulls, while those in $\mathcal{H}_1$ associated with false nulls. The noise process is $Z = \{Z_t, t \in \mathbb{Z}\}$, with each Z_t taking values in a Euclidean space $\mathcal{Z}$. To model the interaction between η_t and Z_t, let $\{\varphi(z, \vartheta), \vartheta \in \Theta\}$ be a family of mappings $\mathcal{Z} \to \mathcal{X}$ indexed by ϑ, where Θ is an open set in $\mathbb{R}^d$ and $\mathcal{X}$ a Euclidean space. Then, let

$$\theta_a : \mathbb{R}^p \to \Theta, \qquad a \in \mathcal{H},$$

be a family of functions, such that each $\varepsilon \in \mathbb{R}^p$ specifies a scenario where the observations are

$$X_t = X_t(\varepsilon) = \varphi(Z_t, \theta_{\eta_t}(\varepsilon)). \tag{2.1}$$

Intuitively, $\varphi(Z_t, \vartheta)$ determines how Z_t interacts with a possible manifestation of η_t to generate an observation X_t; the manifestation of η_t is $\theta_{\eta_t}(\varepsilon)$, with ε being the tuning parameter that determines how strongly η_t manifests itself. The dimension p of ε may be greater than 1 to take into account different aspects of the tuning. We will assume that (η, Z) is defined on the canonical space $\mathcal{H}^{\mathbb{Z}} \times \mathcal{Z}^{\mathbb{Z}}$ equipped with the product Borel σ-algebra.

For function $h : \mathbb{R}^s \to \mathbb{R}$ and s-tuple of nonnegative integers $\nu = (\nu_1, \dots, \nu_s)$, denote the νth derivative of h and its order, respectively, by

$$h^{(\nu)}(x) = \frac{\partial^{|\nu|} h(x)}{\partial x_1^{\nu_1} \cdots \partial x_s^{\nu_s}}, \qquad |\nu| = \nu_1 + \cdots + \nu_s.$$

Denote $h^{(\nu)} := h$ if $\nu = 0 := (0, \dots, 0)$. For $q \in \mathbb{N}$, denote $h \in C^{(q)}$ if $h^{(\nu)}$ exists and is continuous for every $|\nu| \le q$. If $i = (i_1, \dots, i_s)$ and $\nu = (\nu_1, \dots, \nu_s)$, denote $i \le \nu$ if $i_k \le \nu_k$ for every $k = 1, \dots, s$ and denote $i < \nu$ if $i \le \nu$ and $i \ne \nu$.

Assumptions. The following assumptions will be needed for different occasions:

1. Z is independent of η and Z_t are i.i.d. such that for each $\vartheta \in \Theta$ and $t \in \mathbb{Z}$, $\varphi(Z_t, \vartheta)$ has a density $f(x, \vartheta)$.
2. η is a Markov chain and there are $\kappa \geq 1$, $\phi_* > 0$, such that for all $a, b \in \mathcal{H}$ and $s, t \in \mathbb{Z}$ with $|s - t| \geq \kappa$, $\Pr\{\eta_t = b \mid \eta_s = a\} \geq \phi_*$.
3. For each $z \in \mathcal{Z}$ and $a, b \in \mathcal{H}$, $0 < f(\varphi(z, \theta_a(\varepsilon)), \theta_b(\varepsilon)) < \infty$ and is continuous in ε.
4. There is $q \in \mathbb{N}$, such that for each $z \in \mathcal{Z}$ and $a, b \in \mathcal{H}$, $f(\varphi(z, \theta_a(\varepsilon)), \theta_b(\varepsilon))$ as a function in ε belongs to $C^{(q)}$ and all its partial derivatives of order $\leq q$ are continuous in (z, ε). Furthermore, for $r > 0$, there is $c = c(r) > 2$, such that
$$\Pr\{M_q(Z_0, r) \geq u\} = O((\log u)^{-c}), \qquad u \to \infty,$$
where, letting $\ell_{z,ab}(\varepsilon) = \ln f(\varphi(z, \theta_a(\varepsilon)), \theta_b(\varepsilon))$, $M_0(z, r) = 1$ and for $k > 0$,
$$M_k(z, r) = \sup\{|\ell_{z,ab}^{(\nu)}(\varepsilon)| : 1 \leq |\nu| \leq k, |\varepsilon| \leq r, a, b \in \mathcal{H}\}.$$
5. For any $r > 0$, $E[M_q(Z_0, r)^k] < \infty$, where $k = q^2(q + 1)/2$.

Henceforth, for $s, t \in \mathbb{Z}$ and $a, b \in \mathcal{H}$, denote
$$P_t(a) = \Pr\{\eta_t = a\}, \qquad P_{st}(a, b) = \Pr\{\eta_t = b \mid \eta_s = a\}.$$

Remarks. 1. Some examples of φ are given Section 3.3.

2. η need not be stationary or have time-homogeneous transitions.

3. Assumption 3 implies that no value of X_t can decisively identify or rule out any elements in $\mathcal{H}$ as possible values for η_t.

4. In Example 1.1, since $\ell_{z,ab}(\varepsilon) = -\frac{1}{2}[z + \varepsilon(a - b)]^2 - \ln\sqrt{2\pi}$ and $Z_t \sim N(0, 1)$, Assumption 5 is satisfied. The assumption is stronger than Assumption 4. To get results on almost sure convergence, Assumption 4 suffices. However, to get results on expectations, Assumption 5 will be used.

5. Assumption 2 can be relaxed as follows: there are $\phi_* > 0$ and $\cdots < s_k < t_k < s_{k+1} < t_{k+1} < \cdots$, with $s_k \to \pm\infty$ as $k \to \pm\infty$, such that $P_{s_k,t_k}(a, b) \geq \phi_*$ and for $n \gg 1$, $\#\{k : -n \leq s_k \leq 0\}/n$ and $\#\{k : 0 \leq s_k \leq n\}/n$ are bounded away from 0. The analysis under the relaxed assumption follows the same line as the rest of the paper but is more technical. We will not pursue it here.

2.2. *Derivatives of full likelihood ratios*. Given ε and $m, n \in \mathbb{N}$, if the observations consist of $X_s(\varepsilon) = \varphi_s(Z_s, \theta_{\eta_s}(\varepsilon))$ with $s = -m, \ldots, n$, the likelihood ratio for false null vs. true null at t is
$$\rho_{t,mn}(\varepsilon) = \frac{\Pr\{\eta_t \in \mathcal{H}_1 \mid X_{-m}(\varepsilon), \ldots, X_n(\varepsilon)\}}{\Pr\{\eta_t \in \mathcal{H}_0 \mid X_{-m}(\varepsilon), \ldots, X_n(\varepsilon)\}}.$$

Let $\sigma = (\sigma_t)$ be an independent copy of η that is independent of Z as well. Denote by E_σ the expectation with respect to σ. By Bayes formula,

$$\rho_{t,mn}(\varepsilon) = \frac{\sum_{a\in\mathcal{H}_1} P_t(a)\mathsf{E}_\sigma[\prod_{s=-m}^n \psi_s(\varepsilon,\sigma_s) \mid \sigma_t = a]}{\sum_{a\in\mathcal{H}_0} P_t(a)\mathsf{E}_\sigma[\prod_{s=-m}^n \psi_s(\varepsilon,\sigma_s) \mid \sigma_t = a]}, \tag{2.2}$$

where for $c \in \mathcal{H}$,

$$\psi_t(\varepsilon, c) = f(X_t(\varepsilon), \theta_c(\varepsilon)) = f(\varphi(Z_t, \theta_{\eta_t}(\varepsilon)), \theta_c(\varepsilon)). \tag{2.3}$$

As discussed in the Introduction,

$$\rho_t(\varepsilon) = \lim_{m,n\to\infty} \rho_{t,mn}(\varepsilon) = \frac{\Pr\{\eta_t \in \mathcal{H}_1 \mid X_s(\varepsilon), s\in\mathbb{Z}\}}{\Pr\{\eta_t \in \mathcal{H}_0 \mid X_s(\varepsilon), s\in\mathbb{Z}\}}$$

exists almost surely due to martingale convergence and plays an important role in optimal multiple testing procedures.

THEOREM 2.1. *Suppose Assumptions 1–4 hold.*

1. *Almost surely, $\rho_{t,mn} \in C^{(q)}$ for $t = -m+\kappa, \dots, n-\kappa$.*
2. *Almost surely, $\rho_t(\varepsilon)$ is strictly positive for all t and ε.*
3. *There is a deterministic function $r_{t,\nu}(\varepsilon_0) \in (0,1)$ in $\varepsilon_0 > 0$ for each $t \in \mathbb{Z}$ and ν with $|\nu| \le q$, such that almost surely, as $m, n \to \infty$, $\rho_{t,mn}^{(\nu)}(\varepsilon)$ converges, with*

$$\sup_{|\varepsilon|\le\varepsilon_0} \left|\rho_{t,mn}^{(\nu)}(\varepsilon) - \lim_{m,n\to\infty} \rho_{t,mn}^{(\nu)}(\varepsilon)\right| = o(r_{t,\nu}^{m\wedge n}(\varepsilon_0)),$$

for all $t \in \mathbb{Z}$, ν with $|\nu| \le q$ and $\varepsilon_0 > 0$.

Due to the uniform convergence of $\rho_{t,mn}^{(\nu)}$ on every compact set,

$$\rho_t \in C^{(q)}, \qquad \rho_t^{(\nu)}(\varepsilon) = \lim_{m,n\to\infty} \rho_{t,mn}^{(\nu)}(\varepsilon), \qquad t\in\mathbb{Z}, |\nu| \le q \tag{2.4}$$

(cf. [24], Theorem 7.17). Then, as $\rho_t(\varepsilon)$ are strictly positive, the interchange between limit operation and differentiation for the logarithms of $\rho_{t,mn}(\varepsilon)$ in Example 1.1 is justified.

Since $\mathbb{Z}$ is countable, in order to establish Theorem 2.1, it suffices to show it holds for each fixed $t \in \mathbb{Z}$. Without loss of generality, we shall focus on $t = 0$. For ease of notation, henceforth denote $\rho_{mn} = \rho_{0,mn}$.

By the conditional independence of $(\sigma_t, t < 0)$ and $(\sigma_t, t > 0)$ given σ_0,

$$\mathsf{E}_\sigma\left[\prod_{s=-m}^n \psi_s(\varepsilon,\sigma_s) \,\middle|\, \sigma_0\right] = \psi_0(\varepsilon,\sigma_0)\mathsf{E}_\sigma\left[\prod_{s=1}^n \psi_s(\varepsilon,\sigma_s) \,\middle|\, \sigma_0\right] \times \mathsf{E}_\sigma\left[\prod_{s=1}^m \psi_{-s}(\varepsilon,\sigma_{-s}) \,\middle|\, \sigma_0\right].$$

Fix an arbitrary $\iota \in \mathcal{H}$. Define

$$\Lambda_{\pm n,a} = \Lambda_{\pm n,a}(\varepsilon) = \frac{\mathsf{E}_\sigma[\prod_{s=1}^n \psi_s(\varepsilon, \sigma_{\pm s}) \mid \sigma_0 = a]}{\mathsf{E}_\sigma[\prod_{s=1}^n \psi_s(\varepsilon, \sigma_{\pm s}) \mid \sigma_0 = \iota]}. \tag{2.5}$$

Then (2.2) for $t = 0$ can be written as

$$\rho_{mn}(\varepsilon) = \frac{\sum_{a \in \mathcal{H}_1} \psi_0(\varepsilon, a) P_0(a) \Lambda_{-m,a} \Lambda_{n,a}}{\sum_{a \in \mathcal{H}_0} \psi_0(\varepsilon, a) P_0(a) \Lambda_{-m,a} \Lambda_{n,a}}. \tag{2.6}$$

Although $\Lambda_{\pm n,a}$ depends on ι, $\rho_{mn}(\varepsilon)$ is independent of ι. For brevity, ι is omitted in the notation.

From (2.6), it is seen that Theorem 2.1 follows from the next two assertions on uniform geometric contraction of functions and their derivatives on any compact interval of ε.

THEOREM 2.2. *Let Assumptions 1–3 hold. Almost surely, as $n \to \infty$, for all $a \in \mathcal{H}$, $\Lambda_{\pm n,a}(\varepsilon)$ converge uniformly on every compact set of ε. The limits*

$$\mathsf{L}_a(\varepsilon) = \lim_{n \to \infty} \Lambda_{n,a}(\varepsilon), \qquad \bar{\mathsf{L}}_a(\varepsilon) = \lim_{n \to \infty} \Lambda_{-n,a}(\varepsilon) \tag{2.7}$$

are strictly positive and continuous, and there is a deterministic increasing function $r(\varepsilon_0) \in (0, 1)$ in $\varepsilon_0 > 0$, such that almost surely, as $n \to \infty$,

$$\sup_{|\varepsilon| \le \varepsilon_0} |\Lambda_{n,a}(\varepsilon) - \mathsf{L}_a(\varepsilon)| = o(r(\varepsilon_0)^n) \qquad \forall \varepsilon_0 > 0,$$

and likewise for $\Lambda_{-n,a}$ and $\bar{\mathsf{L}}_a(\varepsilon)$.

THEOREM 2.3. *Let Assumptions 1–4 hold. Almost surely, as $n \to \infty$, for each nonzero ν with $|\nu| \le q$ and $a \in \mathcal{H}$, $\Lambda^{(\nu)}_{\pm n,a}$ converge uniformly on every compact set of ε. Let*

$$\mathsf{L}_{\nu,a}(\varepsilon) = \lim_{n \to \infty} \Lambda^{(\nu)}_{n,a}(\varepsilon), \qquad \bar{\mathsf{L}}_{\nu,a}(\varepsilon) = \lim_{n \to \infty} \Lambda^{(\nu)}_{-n,a}(\varepsilon).$$

There is an increasing deterministic function $r_\nu(\varepsilon_0) \in (0, 1)$ in $\varepsilon_0 > 0$, such that almost surely, as $n \to \infty$,

$$\max_a \sup_{|\varepsilon| \le \varepsilon_0} \left|\Lambda^{(\nu)}_{n,a}(\varepsilon) - \mathsf{L}_{\nu,a}(\varepsilon)\right| = o(r_\nu^n(\varepsilon_0)) \qquad \forall \varepsilon_0 > 0,$$

and likewise for $\Lambda_{-n,a}$ and $\bar{\mathsf{L}}_{\nu,a}(\varepsilon)$.

Basically, the two theorems say that $\mathsf{L}_a(\varepsilon)$ and $\bar{\mathsf{L}}_a(\varepsilon)$ are q times differentiable, and for ν with $|\nu| \le q$, $\mathsf{L}_a^{(\nu)}(\varepsilon) = \mathsf{L}_{\nu,a}(\varepsilon)$, $\bar{\mathsf{L}}_a^{(\nu)}(\varepsilon) = \bar{\mathsf{L}}_{\nu,a}(\varepsilon)$, that is, $(\lim \Lambda_{\pm n,a})^{(\nu)} = \lim \Lambda^{(\nu)}_{\pm n,a}$. As a result, $\rho(\varepsilon)$ is q times differentiable, with

$$\rho^{(\nu)}(\varepsilon) = \left[\frac{\sum_{a \in \mathcal{H}_1} \psi_0(\varepsilon, a) P_0(a) \mathsf{L}_a(\varepsilon) \bar{\mathsf{L}}_a(\varepsilon)}{\sum_{a \in \mathcal{H}_0} \psi_0(\varepsilon, a) P_0(a) \mathsf{L}_a(\varepsilon) \bar{\mathsf{L}}_a(\varepsilon)}\right]^{(\nu)}. \tag{2.8}$$

Note that although we are mainly interested on the property of ρ_t around $\varepsilon = 0$, the above results allow Taylor's expansion around nonzero values of ε as well.

In Example 1.1, limit operation, differentiation, and expectation were freely interchanged. The next assertion justifies this.

THEOREM 2.4. *Let Assumptions 1–3 and 5 hold and $\kappa = 1$ in Assumption 2.*

1. *There are $0 < c < C < \infty$, such that almost surely, $c \le \Lambda_{n,a}(\varepsilon) \le C$ for all $n \gg 1$, $a \in \mathcal{H}$ and ε, thus*

$$\mathsf{E}[\ln \mathsf{L}_a(\varepsilon)|\eta] = \lim_{n\to\infty} \mathsf{E}[\ln \Lambda_{n,a}(\varepsilon)|\eta].$$

2. *For ν with $1 \le |\nu| \le q$ and $a \in \mathcal{H}$,*

$$\mathsf{E}[\ln \mathsf{L}_a(\varepsilon)|\eta]^{(\nu)} = \mathsf{E}\big[(\ln \mathsf{L}_a)^{(\nu)}(\varepsilon)|\eta\big] = \lim_{n\to\infty} \mathsf{E}\big[(\ln \Lambda_{n,a})^{(\nu)}(\varepsilon)|\eta\big].$$

Similar results hold for $\Lambda_{-n,a}$ and $\bar{\mathsf{L}}_a$.

3. Binary state HMM with univariate parameters. In this section, we consider in more detail the case where η is a first order binary state Markov chain, with $\eta_t = \mathbf{1}\{H_t \text{ is false}\}$. Also, we suppose $\varepsilon \in \mathbb{R}$ and

$$\theta_0(0) = \theta_1(0) = 0, \tag{3.1}$$

that is, at $\varepsilon = 0$, false and true nulls are no longer distinguishable based on the data.

3.1. *Derivatives of likelihood ratio.* We shall focus $t = 0$. Analysis for other t can be done likewise. By (2.8), the full likelihood ratio (FLR) $\rho(\varepsilon)$ satisfies

$$\ln \frac{\rho(\varepsilon)}{\tilde\rho(\varepsilon)} = \mathsf{r}(\varepsilon) + \bar{\mathsf{r}}(\varepsilon) \qquad \text{with } \mathsf{r}(\varepsilon) = \ln \frac{\mathsf{L}_1(\varepsilon)}{\mathsf{L}_0(\varepsilon)},\ \bar{\mathsf{r}}(\varepsilon) = \ln \frac{\bar{\mathsf{L}}_1(\varepsilon)}{\bar{\mathsf{L}}_0(\varepsilon)}, \tag{3.2}$$

where $\tilde\rho(\varepsilon)$ is the local likelihood ratio (LLR) for η_0 only based on X_0:

$$\tilde\rho(\varepsilon) = \frac{\mathsf{Pr}\{\eta_0 = 1 \mid X_0\}}{\mathsf{Pr}\{\eta_0 = 0 \mid X_0\}} = \frac{P_0(1)\psi_0(\varepsilon, 1)}{P_0(0)\psi_0(\varepsilon, 0)},$$

with $\psi_t(\varepsilon, a)$ being defined in (2.3).

Consider $\mathsf{r}(\varepsilon)$. The treatment of $\bar{\mathsf{r}}(\varepsilon)$ is similar. By Theorem 2.2,

$$\mathsf{r}(\varepsilon) = \lim_{n\to\infty} \lambda_n(\varepsilon) \qquad \text{with } \lambda_n(\varepsilon) = \ln \frac{\mathsf{E}_\sigma[\prod_{s=1}^n \psi_s(\varepsilon, \sigma_s) \mid \sigma_0 = 1]}{\mathsf{E}_\sigma[\prod_{s=1}^n \psi_s(\varepsilon, \sigma_s) \mid \sigma_0 = 0]}. \tag{3.3}$$

By (3.1), for $t \in \mathbb{Z}$,

$$\psi_t(0, \sigma_t) = f(\varphi(Z_t, \theta_{\eta_t}(0)), \theta_{\sigma_t}(0)) = f(\varphi(Z_t, 0), 0) \tag{3.4}$$

is independent of σ, so $\lambda_n(0) = 0$, giving $\mathsf{r}(0) = 0$. Next, define

$$\begin{aligned} d_t(\varepsilon) &= \ln \psi_t(\varepsilon, 1) - \ln \psi_t(\varepsilon, 0), \\ D_{st} &= P_{st}(1, 1) - P_{st}(0, 1), \qquad s, t \in \mathbb{Z}. \end{aligned} \tag{3.5}$$

In general, unless η is stationary, $D_{st} \neq D_{ts}$ for $s \neq t$. By simple algebra, we have the following identity, which the next result relies upon

$$D_{rs} D_{st} = D_{rt}, \qquad D_{ts} D_{sr} = D_{tr}, \qquad r \le s \le t. \tag{3.6}$$

THEOREM 3.1. *Let Assumptions 1–4 hold. Then*

$$\mathsf{r}'(0) = \sum_{t=1}^{\infty} D_{0t} d_t'(0), \tag{3.7}$$

$$\begin{aligned} \mathsf{r}''(0) &= \sum_{t=1}^{\infty} D_{0t} \{ d_t''(0) + [P_{0t}(1,0) - P_{0t}(0,1)][d_t'(0)]^2 \} \\ &\quad + 2 \sum_{t=1}^{\infty} D_{0t} d_t'(0) \sum_{s=1}^{t-1} [P_{0s}(1,0) - P_{0s}(0,1)] d_s'(0), \end{aligned} \tag{3.8}$$

where $'$, $''$, $\dots$, *denote differentiations with respect to* ε.

Simplifications can be made when η is stationary and ergodic. In this case, $p_a = P_0(a) \in (0,1)$ and the transition matrix can be written as

$$Q = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (p_0, p_1) + r \begin{pmatrix} p_1 \\ -p_0 \end{pmatrix} (1, -1),$$

where $r \in (-1, 1)$ is the eigenvalue of Q different from 1. Then for $t \ge 1$,

$$Q^t = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (p_0, p_1) + r^t \begin{pmatrix} p_1 \\ -p_0 \end{pmatrix} (1, -1) = \begin{pmatrix} p_0 + r^t p_1 & p_1 - r^t p_1 \\ p_0 - r^t p_0 & p_1 + r^t p_0 \end{pmatrix},$$

so that in (3.7) and (3.8), $D_{0t} = r^t$ and $P_{0s}(1,0) - P_{0s}(0,1) = (p_0 - p_1)(1 - r^s)$.

3.2. *A univariate case.* In this section, suppose both X_t and $\theta_{\eta_t}(\varepsilon)$ are univariate. Suppose the following regularity conditions are satisfied:

1. $\lambda(x, \vartheta) = \ln f(x, \vartheta) \in C^{(2)}$ and $\varphi(z, v)$ as a function in v belongs to $C^{(2)}$, such that for any ϑ, v, and ν with $|\nu| \le 2$,
$$\mathsf{E}\big[\lambda^{(\nu)}(\varphi(Z_t, v), \vartheta)\big] = (\mathsf{E}[\lambda(\varphi(Z_t, v), \vartheta)])^{(\nu)},$$
where the differentiation is with respect to v and ϑ.
2. $\theta_a(\varepsilon) \in C^{(2)}$ for any $a \in \mathcal{H}$.

PROPOSITION 3.2. *Let Assumptions 1–4 hold. Then for each* t,

$$d_t'(0) = [\theta_1'(0) - \theta_0'(0)] \frac{\partial \lambda(x, 0)}{\partial \vartheta}, \tag{3.9}$$

$$\begin{aligned} d_t''(0) &= 2[\theta_1'(0) - \theta_0'(0)] \theta_{\eta_t}'(0) \frac{\partial^2 \lambda(x, 0)}{\partial x \, \partial \vartheta} \frac{\partial \varphi(Z_t, 0)}{\partial v} \\ &\quad + [\theta_1'(0)^2 - \theta_0'(0)^2] \frac{\partial^2 \lambda(x, 0)}{\partial \vartheta^2} + [\theta_1''(0) - \theta_0''(0)] \frac{\partial \lambda(x, 0)}{\partial \vartheta}, \end{aligned} \tag{3.10}$$

where the partial derivatives of λ *are evaluated at* $x = \varphi(Z_t, 0)$.

PROPOSITION 3.3. *Let Assumptions 1–3 and 5 hold and* $\kappa = 1$ *in Assumption 2. Then*

$$\mathsf{E}[\mathsf{r}'(0) \mid \eta] = 0, \tag{3.11}$$

$$\mathsf{E}[\mathsf{r}''(0) \mid \eta] = \mathsf{Var}[d_0'(0)] \sum_{t=1}^{\infty} D_{0t}[2\eta_t - P_{0t}(1,1) - P_{0t}(0,1)]. \tag{3.12}$$

Moreover, for all t, $\mathsf{E}[d_t'(0)] = 0$ *and* $\mathsf{Var}[d_t'(0)] = [\theta_1'(0) - \theta_0'(0)]^2 J(0)$, *where* $J(\vartheta)$ *is the Fisher information for the parametric family* $f(x, \vartheta)$.

Note that (3.12) implies $\mathsf{E}[\mathsf{r}''(0) \mid \eta_0] = (2\eta_0 - 1)\,\mathsf{Var}[d_0'(0)] \sum_{t=1}^{\infty} D_{0t}^2$, which is what we got toward the end of Example 1.1.

We next use the results to get a better view on the structure of $\rho(\varepsilon)$. Since $\mathsf{r}(0) = 0$, by Taylor's expansion and Theorem 3.1,

$$\begin{aligned}\mathsf{r}(\varepsilon) = {} & \sum_{t=1}^{\infty} D_{0t}\left[d_t'(0)\varepsilon + \frac{d_t''(0)\varepsilon^2}{2}\right] + \frac{\varepsilon^2}{2}\sum_{t=1}^{\infty} D_{0t}[P_{0t}(1,0) - P_{0t}(0,1)][d_t'(0)]^2 \\ & + \varepsilon^2 \sum_{t=1}^{\infty} D_{0t} d_t'(0) \sum_{s=1}^{t-1} [P_{0s}(1,0) - P_{0s}(0,1)] d_s'(0) + o(\varepsilon^2).\end{aligned}$$

Since $d_t(0) = 0$, then $d_t'(0)\varepsilon + d_t''(0)\varepsilon^2/2 = d_t(\varepsilon) + o(\varepsilon^2)$. Under the condition of Proposition 3.3, by Propositions 3.2 and 3.3, all $d_t'(0)$ are independent of η, have mean 0 and the same variance. Similar assertions can be made about the expansion of $\bar{\mathsf{r}}(\varepsilon)$. It follows that

$$\mathsf{r}(\varepsilon) + \bar{\mathsf{r}}(\varepsilon) = \sum_{t \neq 0} D_{0t} d_t(\varepsilon) + \varepsilon^2\,\mathsf{Var}[d_0'(0)] K + \varepsilon^2 \xi + o(\varepsilon^2),$$

where $K = (1/2)\sum_{t \neq 0} D_{0t}[P_{0t}(1,0) - P_{0t}(0,1)]$ and ξ is a random variable independent of η and has mean 0. Then by (3.2) and the definition of $d_t(\varepsilon)$ in (3.5),

$$\rho(\varepsilon) = \tilde{\rho}(\varepsilon) \prod_{t \neq 0} \left[\frac{\psi_t(\varepsilon, 1)}{\psi_t(\varepsilon, 0)}\right]^{D_{0t}} \times \exp\{\varepsilon^2 (\mathsf{Var}[d_0'(0)] K + \xi) + o(\varepsilon^2)\}. \tag{3.13}$$

According to (2.3), $\psi_t(\varepsilon, 1)/\psi_t(\varepsilon, 0)$ is the marginal likelihood ratio of X_t for the isolated test on $\eta_t = 1$ vs. $\eta_t = 0$, which completely ignores the dependence among the sites. The above expansion shows that all these likelihood ratios are factored into the FLR, with effects being adjusted by D_{0t}. For example, if D_{0t} is positive (resp., negative), then a large likelihood ratio at site t increases (resp., decreases) the FLR for the test on η_0. Also, by (3.6), if s has the same sign as t but farther away from 0, then the effect of the marginal likelihood ratio at site s on the test on η_0 is determined by D_{0t} and D_{ts}. In contrast, the LLR $\tilde{\rho}(\varepsilon)$ only takes into account the marginal likelihood ratio at site 0.

The above expansion is obtained for η_0. Taking into account explicitly the dependence on site location, the FLRs for the multiple tests on η_s, $s \in \mathbb{Z}$, are

$$\rho_s(\varepsilon) = \tilde{\rho}_s(\varepsilon) \prod_{t \neq s} \left[\frac{\psi_t(\varepsilon, 1)}{\psi_t(\varepsilon, 0)} \right]^{D_{st}} \times \exp\{\varepsilon^2 (\mathsf{Var}[d_0'(0)] K_s + \xi_s) + o(\varepsilon^2)\}, \tag{3.14}$$

where the LLR $\tilde{\rho}_s(\xi)$ and constants K_s are now expressed as

$$\tilde{\rho}_s(\varepsilon) = \frac{P_s(1)\psi_s(\varepsilon, 1)}{P_s(0)\psi_s(\varepsilon, 0)}, \qquad K_s = (1/2) \sum_{t \neq s} D_{st}[P_{st}(1, 0) - P_{st}(0, 1)]$$

and ξ_s are centered random variables independent of η. The conditional likelihoods of η_s can then be computed via $\mathsf{Pr}\{\eta_s = 0 \mid X\} = [1 + \rho_s(\varepsilon)]^{-1}$.

3.3. *Examples*.

EXAMPLE 3.1 (Translation). Suppose φ is defined on $\mathbb{R} \times \mathbb{R}$ such that $\varphi(z, v) = z + v$ and for $a = 0, 1$, $\theta_a(\varepsilon) = \varepsilon a$. Let each Z_t have density $h(z) = e^{-V(z)}$. Apparently, Example 1.1 belongs to this case.

For each $\vartheta \in \mathbb{R}$, $\varphi(Z_t, \vartheta) = Z_t + \vartheta$ has density $f(x, \vartheta) = h(x - \vartheta)$. Therefore, $\lambda(x, \vartheta) = \ln f(x, \vartheta) = -V(x - \vartheta)$. It is easy to check

$$\theta_a'(0) = a, \qquad \frac{\partial \varphi(z, 0)}{\partial v} = 1, \qquad \frac{\partial \lambda(x, \vartheta)}{\partial \vartheta} = V'(x - \vartheta),$$

$$\frac{\partial^2 \lambda(x, \vartheta)}{\partial x \, \partial \vartheta} = -\frac{\partial^2 \lambda(x, \vartheta)}{\partial \vartheta^2} = V''(x - \vartheta).$$

Provided necessary conditions are satisfied, by Proposition 3.2,

$$d_t'(0) = V'(Z_t), \qquad d_t''(0) = (2\eta_t - 1)V''(Z_t),$$

$$\mathsf{Var}[d_t'(0)] = \int V'(x)^2 e^{-V(x)} \, dx.$$

Then we can get $\mathsf{r}'(0)$, $\mathsf{r}''(0)$ and $\mathsf{E}[\mathsf{r}''(0) \mid \eta]$ by Theorem 3.1 and (3.12).

EXAMPLE 3.2 (Scaling). Suppose φ is defined on $\mathbb{R} \times \mathbb{R}$ such that $\varphi(z, v) = e^{-v} z$ and for $a = 0, 1$, $\theta_a(\varepsilon) = \varepsilon a$. Let each Z_t have density $h(z) = e^{-V(z)}$. For $v \in \mathbb{R}$, $\varphi(Z_t, v)$ has density $f(x, v) = e^v h(e^v x)$. Therefore, $\lambda(x, v) = v - V(e^v x)$. By Proposition 3.2,

$$d_t'(0) = 1 - Z_t V'(Z_t), \qquad d_t''(0) = (2\eta_t - 1) Z_t [V'(Z_t) + Z_t V''(Z_t)],$$

$$\mathsf{Var}[d_0'(0)] = \int [1 - x V'(x)]^2 e^{-V(x)} \, dx.$$

Then we can get $\mathsf{r}'(0)$, $\mathsf{r}''(0)$, and $\mathsf{E}[\mathsf{r}''(0) \mid \eta]$ by Theorem 3.1 and (3.12).

EXAMPLE 3.3 (t-statistics). Suppose the data observed at each time point t is a random vector $\xi_t = (\xi_{t,1}, \dots, \xi_{t,\nu+1})$, such that conditional on η, ξ_t are independent of each other, and for each t, $\xi_{t,j}$ are i.i.d. $\sim N(\varepsilon s_t \eta_t, s_t^2)$ for some $s_t = s_t(\eta) > 0$. Suppose s_t are completely intractable, that is, there is no information on the values of s_t or their interrelations. In this case, it is reasonable to use the t-statistics

$$X_t = \frac{\sqrt{\nu+1}\bar{\xi}_t}{\sqrt{S_t^2/\nu}}$$

to test on η_t, where $\bar{\xi}_t$ is the mean of $\xi_{t,j}$ and S_t^2 the sum of squares of $\xi_{t,j} - \bar{\xi}_t$.

By scaling, we assume without loss of generality that $s_t = 1$. Let $\zeta_t = \sqrt{\nu+1}(\bar{\xi}_t - \varepsilon\eta_t)$. Then, given η, $\zeta_t \sim N(0,1)$ and $S_t^2 \sim \chi_\nu^2$ are independent of each other. Define $Z_t = (\zeta_t, S_t)$ and, for $z = (r, s)$ and $a = 0, 1$, define

$$\varphi(z, v) = \sqrt{\nu}(r + v)/s, \qquad \theta_a(\varepsilon) = \sqrt{\nu+1}a\varepsilon.$$

Then $X_t = \sqrt{\nu}(\zeta_t + \sqrt{\nu+1}\eta_t\varepsilon)/S_t = \varphi(Z_t, \theta_{\eta_t}(\varepsilon))$. Conditional on η, $X_t \sim t_{\nu,\vartheta}(x)$ with $\vartheta = \theta_{\eta_t}(\varepsilon)$, that is, the noncentral t-distribution with ν degrees of freedom (df) and noncentrality parameter ϑ. In the notation of Assumption 1, $f(x, \vartheta) = t_{\nu,\vartheta}(x)$.

Recall

$$t_\nu(x) = \frac{C_\nu}{(\nu + x^2)^{(\nu+1)/2}} \qquad \text{with } C_\nu = \frac{\nu^{\nu/2}\Gamma((\nu+1)/2)}{\sqrt{\pi}\Gamma(\nu/2)},$$

$$t_{\nu,\vartheta}(x) = t_\nu(x) e^{-\vartheta^2/2}\left[1 + \sum_{k=1}^{\infty} \frac{c_k x^k}{(\nu+x^2)^{k/2}} \frac{\vartheta^k}{k!}\right]$$

with $c_k = \frac{\Gamma((\nu+k+1)/2)2^{k/2}}{\Gamma((\nu+1)/2)}$.

Therefore,

$$\lambda(x, \vartheta) = \ln f(x, \vartheta) = \ln t_\nu(x) - \frac{1}{2}\vartheta^2 + \ln\left[1 + \sum_{k=1}^{\infty} \frac{c_k x^k}{(\nu+x^2)^{k/2}} \frac{\vartheta^k}{k!}\right].$$

By $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots$,

$$\lambda(x, \vartheta) = \frac{c_1 x}{\sqrt{\nu + x^2}}\vartheta + \frac{1}{2}\left\{\frac{(c_2 - c_1^2)x^2}{\nu + x^2} - 1\right\}\vartheta^2 + \ln t_\nu(x) + O(\vartheta^3).$$

It follows that

$$\frac{\partial \lambda(x,0)}{\partial \vartheta} = \frac{c_1 x}{\sqrt{\nu + x^2}}, \qquad \frac{\partial^2 \lambda(x,0)}{\partial x \, \partial \vartheta} = \frac{c_1 \nu}{(\nu + x^2)^{3/2}},$$

$$\frac{\partial^2 \lambda(x,0)}{\partial \vartheta^2} = \frac{(c_2 - c_1^2)x^2}{\nu + x^2} - 1.$$

At $\varepsilon = 0$, $X_t = \sqrt{\nu}\zeta_t/S_t$. Since $\theta_a'(0) = \sqrt{\nu+1}a$, (3.9) yields

$$d_t'(0) = \frac{\sqrt{\nu+1}c_1 X_t}{\sqrt{\nu + X_t^2}} = \frac{\sqrt{2(\nu+1)}\Gamma(\nu/2+1)\zeta_t}{\Gamma((\nu+1)/2)\sqrt{\zeta_t^2 + S_t^2}}.$$

Next, since $\partial\varphi(Z_t, 0)/\partial v = \sqrt{\nu}/S_t$, by (3.10),

$$d_t''(0) = \frac{2c_1(\nu+1)\eta_t S_t^2}{(S_t^2+\zeta_t^2)^{3/2}} + (\nu+1)\left[\frac{(c_2 - c_1^2)\zeta_t^2}{S_t^2+\zeta_t^2} - 1\right].$$

Then $\mathsf{r}'(0)$ and $\mathsf{r}''(0)$ can be calculated by Theorem 3.1.

To apply Proposition 3.3, we need to check if Assumption 5 holds. It is not hard to see that for $g(\varepsilon) := \lambda(\varphi(Z_t, \theta_a(\varepsilon)), \theta_b(\varepsilon))$, $g^{(k)}(\varepsilon)$ is a linear combination of $S_t^{-j} \frac{\partial^j \lambda(x,\vartheta)}{\partial x^j} \frac{\partial^{k-j}\lambda(x,\vartheta)}{\partial \vartheta^{k-j}}$ evaluated at $x = \varphi(Z_t, \theta_a(\varepsilon))$ and $\vartheta = \theta_b(\varepsilon)$. It is also not hard to see that $\frac{\partial^j \lambda(x,\vartheta)}{\partial x^j}$ and $\frac{\partial^{k-j}\lambda(x,\vartheta)}{\partial \vartheta^{k-j}}$ are bounded, so as long as $\mathsf{E}[S_t^{-jq^2(q+1)/2}] < \infty$ for $j \le q$, Assumption 5 holds. Since here $q = 2$ and $S_t^2 \sim \chi_\nu^2$, it suffices to have $\nu > 12$. Under this condition,

$$\mathsf{Var}[d_0'(0)] = \left[\frac{\sqrt{2(\nu+1)}\Gamma(\nu/2+1)}{\Gamma((\nu+1)/2)}\right]^2 \mathsf{E}\left[\frac{\zeta_t^2}{\zeta_t^2 + S_t^2}\right].$$

Because S_t^2 is the sum of squares of ν i.i.d. $N(0,1)$ random variables that are independent of $\zeta_t \sim N(0,1)$, by symmetry,

$$\mathsf{E}\left[\frac{\zeta_t^2}{S_t^2+\zeta_t^2}\right] = \frac{1}{\nu+1} \quad\Longrightarrow\quad \mathsf{Var}[d_0'(0)] = \frac{1}{2}\left[\frac{\nu\Gamma(\nu/2)}{\Gamma((\nu+1)/2)}\right]^2.$$

Then $\mathsf{E}[\mathsf{r}''(0) \mid \eta]$ can be calculated by (3.12).

4. Technical details.

4.1. *Some inequalities.* For any set A, denote by $\#A$ the number of its elements.

LEMMA 4.1. *Let $\mathcal{H}$ be a finite set and $W_a \ge 0$, $V_a \ge 0$ for $a \in \mathcal{H}$ such that $W := \sum_a W_a > 0$ and $V := \sum_a V_a > 0$. Then for any x_a, $a \in \mathcal{H}$,*

$$\left|W^{-1}\sum_a W_a x_a - V^{-1}\sum_a V_a x_a\right| \le \max_{a,b\in\mathcal{H}} |x_a - x_b| \left[1 - \frac{\min_a(V_a/W_a)}{\max_a(V_a/W_a)}\right].$$

PROOF. Enumerate the elements in $\mathcal{H}$ in an arbitrary order. Then the left-hand side equals $|T|/D$, where

$$T = \sum_{a,b}(W_a V_b x_a - W_b V_a x_a) = \sum_{a<b}(W_a V_b - W_b V_a)(x_a - x_b),$$

$$D = \sum_{a,b}(W_a V_b + W_b V_a) \ge \sum_{a<b}(W_a V_b + W_b V_a).$$

Denote $\Delta = \max_{a,b} |x_a - x_b|$. Then

$$\frac{|T|}{D} \le \frac{\Delta \sum_{a<b} |W_a V_b - W_b V_a|}{\sum_{a<b}(W_a V_b + W_b V_a)} \le \Delta \max_{a,b} \frac{W_a V_b - W_b V_a}{W_a V_b + W_b V_a}$$
$$= \Delta\left[1 - \min_{a,b} \frac{2V_a/W_a}{V_a/W_a + V_b/W_b}\right] \le \Delta\left[1 - \frac{\min_a(V_a/W_a)}{\max_a(V_a/W_a)}\right],$$

completing the proof. □

LEMMA 4.2. *Let $\mathcal{A}$ and $\mathcal{B}$ be finite sets and $W_a, V_a, x_a > 0$ for $a \in \mathcal{A} \cup \mathcal{B}$. Then*

$$\left|\frac{\sum_{b\in\mathcal{B}} W_b x_b}{\sum_{a\in\mathcal{A}} W_a x_a} - \frac{\sum_{b\in\mathcal{B}} V_b x_b}{\sum_{a\in\mathcal{A}} V_a x_a}\right| \le \#B \times \left(\frac{\max_{b\in\mathcal{B}} x_b}{\min_{a\in\mathcal{A}} x_a}\right) \max_{a\in\mathcal{A}, b\in\mathcal{B}} \left|\frac{W_b}{W_a} - \frac{V_b}{V_a}\right|.$$

PROOF. The left-hand side equals $|T|/D$, where

$$T = \sum_{a\in\mathcal{A}, b\in\mathcal{B}} x_a x_b (W_b V_a - W_a V_b) = \sum_{a\in\mathcal{A}, b\in\mathcal{B}} x_a x_b W_a V_a \left(\frac{W_b}{W_a} - \frac{V_b}{V_a}\right),$$

$$D = \sum_{a,a'\in\mathcal{A}} x_a x_{a'} W_a V_{a'} \ge \left(\min_{a\in\mathcal{A}} x_a\right) \sum_{a\in\mathcal{A}} W_a V_a x_a.$$

Then by

$$|T| \le \#\mathcal{B}\left(\max_{b\in\mathcal{B}} x_b\right) \max_{a\in\mathcal{A}, b\in\mathcal{B}} \left|\frac{W_b}{W_a} - \frac{V_b}{V_a}\right| \sum_{a\in\mathcal{A}} W_a V_a x_a,$$

the lemma follows. □

LEMMA 4.3. *Let $\mathcal{H}$ be a finite set and $q \in \mathbb{N}$. For $a \in \mathcal{H}$, let $W_a : \mathbb{R}^p \to [0, \infty)$ and $g_a : \mathbb{R}^p \to \mathbb{R}$ be q times differentiable. Suppose $W := \sum_a W_a > 0$. Define function $\bar{g} = W^{-1} \sum_a W_a g_a$. Enumerate $\mathcal{H}$ in an arbitrary order. Then for ν with $|\nu| = 1$,*

$$\bar{g}^{(\nu)} = W^{-1} \sum_a W_a g_a^{(\nu)} + W^{-2} \sum_{a<b} (W_a^{(\nu)} W_b - W_a W_b^{(\nu)})(g_a - g_b), \tag{4.1}$$

and more generally, for ν with $|\nu| \le q$,

$$\bar{g}^{(\nu)} = W^{-1} \sum_a W_a g_a^{(\nu)} + \sum_{k=2}^{|\nu|+1} \sum_{0\le j<\nu} W^{-k} U_{k,\nu,j}, \tag{4.2}$$

where $U_{k,\nu,j}$ can be written as

$$U_{k,\nu,j} = \sum_{\substack{a_1,\dots,a_k\in\mathcal{H}, a_1<a_2 \\ i_1+\dots+i_k=\nu-j}} c_{k,\nu}(a_1, \dots, a_k, i_1, \dots, i_k) \prod_{s=1}^k W_{a_s}^{(i_s)} \times (g_{a_1}^{(j)} - g_{a_2}^{(j)}),$$

with $c_{k,\nu}(a_1, \dots, a_k, i_1, \dots, i_k)$ being constants.

PROOF. If $|\nu| = 1$, then

$$\begin{aligned}\bar{g}^{(\nu)} &= W^{-1}\sum_a W_a g_a^{(\nu)} + W^{-1}\sum_a W_a^{(\nu)} g_a - W^{-2}\sum_a W_a g_a \sum_b W_b^{(\nu)} \\ &= W^{-1}\sum_a W_a g_a^{(\nu)} + W^{-2}\sum_{a\neq b}(W_a^{(\nu)} W_b - W_a W_b^{(\nu)}) g_a \\ &= W^{-1}\sum_a W_a g_a^{(\nu)} + W^{-2}\sum_{a<b}(W_a^{(\nu)} W_b - W_a W_b^{(\nu)})(g_a - g_b),\end{aligned}$$

showing (4.1), and hence (4.2) for $|\nu| = 1$. Let $\nu = e + \mu$, where $|e| = 1$ and $0 \le \mu < \nu$. Suppose $\bar{g}^{(\mu)}$ has the form (4.2). Then

$$\bar{g}^{(\nu)} = (\bar{g}^{(\mu)})^{(e)} = \bar{f}^{(e)} + \sum_{k=2}^{|\nu|}\sum_{0\le j<\nu}(W^{-k} U_{k,\mu,j})^{(e)},$$

where $\bar{f} = W^{-1}\sum W_a f_a$, with $f_a = g_a^{(\mu)}$. By (4.1),

$$\begin{aligned}\bar{f}^{(e)} &= W^{-1}\sum_a W_a f_a^{(e)} + W^{-2}\sum_{a<b}(W_a^{(e)} W_b - W_a W_b^{(e)})(f_a - f_b) \\ &= W^{-1}\sum_a W_a g_a^{(\nu)} + W^{-2}\sum_{a<b}(W_a^{(e)} W_b - W_a W_b^{(e)})(g_a^{(\mu)} - g_b^{(\mu)}).\end{aligned}$$

On the other hand, for each $k = 2, \dots, |\nu|$ and $0 \le j < \nu$,

$$(W^{-k} U_{k,\mu,j})^{(e)} = -k W^{-k-1}\sum_{a\in\mathcal{H}} W_a^{(e)} U_{k,\mu,j} + W^{-k} U_{k,\mu,j}^{(e)}.$$

It is then not hard to see that $\bar{g}^{(\nu)}$ has the form (4.2). The proof is complete by induction. □

4.2. *Basic facts*. Define matrix-valued functions $L_n(\varepsilon) = (L_{n,ab}(\varepsilon), a, b \in \mathcal{H})$, such that for $n \ge 0$,

$$L_{\pm n,ab}(\varepsilon) = \mathsf{E}_\sigma\left[\mathbf{1}\{\sigma_{\pm n} = b\}\prod_{s=1}^{n}\psi_{\pm s}(\varepsilon, \sigma_{\pm s})\,\middle|\,\sigma_0 = a\right]. \tag{4.3}$$

Then from (2.5),

$$\Lambda_{n,a}(\varepsilon) = \frac{\sum_{b\in\mathcal{H}} L_{n,ab}(\varepsilon)}{\sum_{b\in\mathcal{H}} L_{n,\imath b}(\varepsilon)}. \tag{4.4}$$

For ease of notation, when there is no confusion, ε will be omitted.

LEMMA 4.4. *Let Assumptions 1–4 hold. Then for each n and a, $b \in \mathcal{H}$, $L_{n,ab} \in C^{(q)}$, and for $|n| \ge \kappa$, $L_{n,ab}$ is positive and finite.*

PROOF. By Assumption 4, $\psi_n(\varepsilon, a) \in C^{(q)}$ for each $n \in \mathbb{Z}$ and $a \in \mathcal{H}$, implying $L_{\pm n,ab} \in C^{(q)}$. For $n \ge \kappa$ and $a, b \in \mathcal{H}$, as $P_{0n}(a, b) > 0$, there is at least one $v = (v_1, \dots, v_n)$ with $v_n = b$ and $\mathsf{Pr}\{\sigma_1 = v_1, \dots, \sigma_n = v_n \mid \sigma_0 = a\} > 0$. For each such v and $t = 1, \dots, n$, by Assumption 3, $\psi_t(\varepsilon, v_t) \in (0, \infty)$. Therefore, $L_{n,ab}(\varepsilon) \in (0, \infty)$. The proof for $L_{-n,ab}$ is similar. $\square$

According to the lemma, $\Lambda_{n,a} \in (0, \infty)$ once $|n| \ge \kappa$. Also, by assumptions 2 and 3, $P_0(a) > 0$, $\psi_0(\varepsilon, a) > 0$. Therefore, $\rho_{mn}(\varepsilon) \in (0, \infty)$.

The following relation will be repeatedly used:

$$L_{n,ab} = \psi_n(\varepsilon, b) \sum_e L_{n-k,ae} I_{n,eb}^{(k)}, \qquad a, b \in \mathcal{H}, 1 \le k < n, \tag{4.5}$$

where

$$I_{n,eb}^{(k)} = I_{n,eb}^{(k)}(\varepsilon) = \mathsf{E}_\sigma \left[\mathbf{1}\{\sigma_n = b\} \prod_{n-k+1}^{n-1} \psi_t(\varepsilon, \sigma_t) \,\middle|\, \sigma_{n-k} = e \right]. \tag{4.6}$$

Similar relation holds when $n < 0$.

4.3. *Proof of Theorem 2.2.*

LEMMA 4.5. *Let Assumptions 1–3 hold.*

1. *Given a, $b \in \mathcal{H}$ and ε, for $|n| \ge \kappa$, $\min_e \frac{L_{n,be}(\varepsilon)}{L_{n,ae}(\varepsilon)}$ is strictly positive and increasing in n, and $\max_e \frac{L_{n,be}(\varepsilon)}{L_{n,ae}(\varepsilon)}$ is finite and decreasing in $|n|$.*

2. *There is an increasing deterministic function $r(\varepsilon_0) \in (0, 1)$, such that given $\varepsilon_0 > 0$, for almost all realizations of Z and η,*

$$\Delta_n(\varepsilon) := \max_{a,b,c,d} \left| \frac{L_{n,bc}(\varepsilon)}{L_{n,ac}(\varepsilon)} - \frac{L_{n,bd}(\varepsilon)}{L_{n,ad}(\varepsilon)} \right| \le C r(\varepsilon_0)^{|n|}, \qquad |n| \ge \kappa, |\varepsilon| \le \varepsilon_0, \tag{4.7}$$

where $C = C(\varepsilon_0, Z)$ is a random variable that only depends on ε_0 and Z and is finite almost surely. Additionally, for fixed ε, $\Delta_{\pm n}(\varepsilon)$ are decreasing in n.

PROOF. We only consider $n > 0$. The case $n < 0$ is similar. Given $a \ne b \in \mathcal{H}$, for $n \ge \kappa$ and $c \in \mathcal{H}$, by Lemma 4.4, $\frac{L_{n,bc}}{L_{n,ac}} \in (0, \infty)$. Then by (4.5),

$$\frac{L_{n,bc}}{L_{n,ac}} = \frac{\sum_e L_{n-k,be} I_{n,ec}^{(k)}}{\sum_e L_{n-k,ae} I_{n,ec}^{(k)}}. \tag{4.8}$$

Letting $k = 1$, it is easy to see that

$$\min_e \frac{L_{n-1,be}}{L_{n-1,ae}} \le \frac{L_{n,bc}}{L_{n,ac}} \le \max_e \frac{L_{n-1,be}}{L_{n-1,ae}} \qquad \text{all } c \in \mathcal{H},$$

which implies part 1.

Given $1 \le k < n$ and ε, for each $a, b, c, d \in \mathcal{H}$, apply Lemma 4.1 to $x_e = \frac{L_{n-k,be}}{L_{n-k,ae}}$, $W_e = L_{n-k,ae} I^{(k)}_{n,ec}$ and $V_e = L_{n-k,ae} I^{(k)}_{n,ed}$. Then by (4.8),

$$\left|\frac{L_{n,bc}}{L_{n,ac}} - \frac{L_{n,bd}}{L_{n,ad}}\right| \le \max_{c,d}\left|\frac{L_{n-k,bc}}{L_{n-k,ac}} - \frac{L_{n-k,bd}}{L_{n-k,ad}}\right| \times \left[1 - \frac{\min_e I^{(k)}_{n,ed}/I^{(k)}_{n,ec}}{\max_e I^{(k)}_{n,ed}/I^{(k)}_{n,ec}}\right].$$

Take maximum over c and d and then over a and b. It follows that

$$\Delta_n(\varepsilon) \le \gamma_n \Delta_{n-k}(\varepsilon) \qquad \text{with } \gamma_n = \gamma_n(\varepsilon, k) = 1 - \frac{\min_{c,d,e} I^{(k)}_{n,ed}/I^{(k)}_{n,ec}}{\max_{c,d,e} I^{(k)}_{n,ed}/I^{(k)}_{n,ec}}. \tag{4.9}$$

For $z = (z_1, \dots, z_{\kappa-1}) \in \mathcal{Z}^{\kappa-1}$, where κ is as in Assumption 2, define

$$\alpha(z, \varepsilon) = \min_{\substack{u_t, v_t \in \mathcal{H} \\ 1 \le t \le \kappa-1}} \prod_{t=1}^{\kappa-1} f(\varphi(z_t, \theta_{u_t}(\varepsilon)), \theta_{v_t}(\varepsilon)), \qquad \alpha_*(z, \varepsilon_0) = \inf_{|\varepsilon| \le \varepsilon_0} \alpha(z, \varepsilon),$$

$$\beta(z, \varepsilon) = \max_{\substack{u_t, v_t \in \mathcal{H} \\ 1 \le t \le \kappa-1}} \prod_{t=1}^{\kappa-1} f(\varphi(z_t, \theta_{u_t}(\varepsilon)), \theta_{v_t}(\varepsilon)), \qquad \beta^*(z, \varepsilon_0) = \sup_{|\varepsilon| \le \varepsilon_0} \beta(z, \varepsilon).$$

For $n \ge \kappa$, let

$$\zeta_n = \zeta_n(\varepsilon_0) = \alpha_*(Z_{n-\kappa+1}, \dots, Z_{n-1}, \varepsilon_0),$$
$$\xi_n = \xi_n(\varepsilon_0) = \beta^*(Z_{n-\kappa+1}, \dots, Z_{n-1}, \varepsilon_0).$$

Since $\psi_t(\varepsilon, \sigma_t) = f(\varphi(Z_t, \theta_{\eta_t}(\varepsilon)), \theta_{\sigma_t}(\varepsilon))$, then for $|\varepsilon| \le \varepsilon_0$,

$$\zeta_n \le \prod_{n-\kappa+1}^{n-1} \psi_t(\varepsilon, \sigma_t) \le \xi_n \tag{4.10}$$

$$\Longrightarrow \quad \zeta_n P_{n-\kappa,n}(e, c) \le I^{(\kappa)}_{n,ec}(\varepsilon) \le \xi_n P_{n-\kappa,n}(e, c). \tag{4.11}$$

Given $z \in \mathcal{Z}^{\kappa-1}$, by $\#H < \infty$ and Assumption 3, $\alpha(z, \varepsilon)$ and $\beta(z, \varepsilon)$ are continuous in ε and $0 < \alpha(z, \varepsilon) \le \beta(z, \varepsilon) < \infty$, yielding $0 < \alpha_*(z, \varepsilon_0) \le \beta^*(z, \varepsilon_0) < \infty$. As a result, $\Pr\{0 < \zeta_n \le \xi_n < \infty\} = 1$. Fix $0 < x < y < \infty$, such that $p_0 := \Pr\{x \le \zeta_\kappa \le \xi_\kappa \le y\} > 0$. Note that x and y can be chosen in such as way that they only depend on ε_0, the distribution of Z, and κ. Because Z_t are i.i.d., from the definitions of ζ_n and ξ_n, almost surely, there is an infinite sequence $n_s = n_s(Z, \varepsilon_0) \ge \kappa$, $s \ge 0$, such that

$$x \le \zeta_{n_s} \le \xi_{n_s} \le y \tag{4.12}$$

and furthermore, n_s can be chosen in such a way that

$$n_s \ge n_{s-1} + \kappa, \qquad |\{s : n_s \le n\}| \ge \frac{p_0 n}{2\kappa} \qquad \text{for } n \gg 1. \tag{4.13}$$

On the other hand, since $\#\mathcal{H} > 1$, Assumption 2 implies that

$$\phi_* \le P_{n-\kappa,n}(e,c) \le 1-\phi_* \qquad \text{all } c, e \in \mathcal{H}. \tag{4.14}$$

Combine (4.11), (4.12) and (4.14) to get

$$0 < \phi_* x \le I^{(\kappa)}_{n_s,ec}(\varepsilon) \le (1-\phi_*)y < \infty \qquad \forall c, e \in \mathcal{H}$$

and hence

$$\begin{aligned} \gamma_{n_s} &= 1 - \frac{\min_{c,d,e} I^{(\kappa)}_{n_s,ed}/I^{(\kappa)}_{n_s,ec}}{\max_{c,d,e} I^{(\kappa)}_{n_s,ed}/I^{(\kappa)}_{n_s,ec}} \le r_0 = r_0(\varepsilon_0) \\ &:= 1 - \left[\frac{\phi_* x}{(1-\phi_*)y}\right]^2 < 1. \end{aligned} \tag{4.15}$$

Now by (4.9), $\Delta_{n_s}(\varepsilon) \le \Delta_{n_s-\kappa}(\varepsilon) r_0$. Since $n_{s-1} \le n_s - \kappa$ while (4.9) implies that $\Delta_n(\varepsilon)$ is decreasing, $\Delta_{n_s}(\varepsilon) \le \Delta_{n_{s-1}}(\varepsilon) r_0$ and hence $\Delta_{n_s}(\varepsilon) \le \Delta_{n_1}(\varepsilon) r_0^{s-1}$ by induction. For any n, if $n_s \le n < n_{s+1}$, then $\Delta_n(\varepsilon) \le \Delta_{n_1}(\varepsilon) r_0^{s-1} \le \Delta_\kappa(\varepsilon) r_0^{s-1}$. Combining (4.13), for $n \gg 1$,

$$\Delta_n(\varepsilon) \le [\Delta_\kappa(\varepsilon)/r_0] r(\varepsilon_0)^n \qquad \text{with } r(\varepsilon_0) = r_0^{p_0/(2\kappa)}.$$

Note $\Delta_\kappa(\varepsilon) \le \max_{a,b,c} \frac{L_{\kappa,ac}(\varepsilon)}{L_{\kappa,bc}(\varepsilon)}$. Using (4.3) and (4.10) followed by assumption 2, it is seen that

$$\max_{a,b,c} \frac{L_{\kappa,ac}(\varepsilon)}{L_{\kappa,bc}(\varepsilon)} \le \frac{\xi_\kappa}{\zeta_\kappa} \max_{a,b,c} \frac{P_{0\kappa}(b,c)}{P_{0\kappa}(a,c)} \le \frac{(1-\phi_*)\xi_\kappa}{\phi_* \zeta_\kappa} < \infty.$$

Therefore, (4.7) is proved.

To make $r(\varepsilon_0)$ increasing, replace $r(\varepsilon_0)$ with, say, $[\inf_{c \ge \varepsilon_0} r(c) + 1]/2$. From the construction, $r(\varepsilon_0)$ only depends on the distributional properties of Z and η, but not specific realizations of the processes. Therefore, $r(\varepsilon_0)$ is deterministic. □

LEMMA 4.6. *Fix $a \in \mathcal{H}$ and ε.*

1. *For $a \in \mathcal{H}$,*

$$0 < \inf_{|n| \ge \kappa} \Lambda_{n,a}(\varepsilon) \le \sup_{|n| \ge \kappa} \Lambda_{n,a}(\varepsilon) < \infty.$$

2. *For $s \ge n \ge \kappa$ and $s \le n \le -\kappa$,*

$$|\Lambda_{n,a}(\varepsilon) - \Lambda_{s,a}(\varepsilon)| \le 2\Delta_n(\varepsilon) + \Delta_s(\varepsilon).$$

PROOF. From (4.4), for $s \ge n \ge \kappa$ and $s \le n \le -\kappa$,

$$\Lambda_{n,a}(\varepsilon), \frac{L_{s,ae}(\varepsilon)}{L_{s,\iota e}(\varepsilon)} \in \left[\min_e \frac{L_{n,ae}(\varepsilon)}{L_{n,\iota e}(\varepsilon)}, \max_e \frac{L_{n,ae}(\varepsilon)}{L_{n,\iota e}(\varepsilon)}\right].$$

Together with part 1 of Lemma 4.5, this yields the first part of the lemma and also

$$\left|\Lambda_{n,a}(\varepsilon) - \frac{L_{n,ab}(\varepsilon)}{L_{n,\imath b}(\varepsilon)}\right| \le \Delta_n, \qquad \left|\frac{L_{n,ab}(\varepsilon)}{L_{n,\imath b}(\varepsilon)} - \frac{L_{s,ab}(\varepsilon)}{L_{s,\imath b}(\varepsilon)}\right| \le \Delta_n,$$

where $b \in \mathcal{H}$ is arbitrary. Then by

$$|\Lambda_{n,a}(\varepsilon) - \Lambda_{s,a}(\varepsilon)| \le \left|\Lambda_{n,a}(\varepsilon) - \frac{L_{n,ab}(\varepsilon)}{L_{n,\imath b}(\varepsilon)}\right| + \left|\Lambda_{s,a}(\varepsilon) - \frac{L_{s,ab}(\varepsilon)}{L_{s,\imath b}(\varepsilon)}\right| + \left|\frac{L_{n,ab}(\varepsilon)}{L_{n,\imath b}(\varepsilon)} - \frac{L_{s,ab}(\varepsilon)}{L_{s,\imath b}(\varepsilon)}\right|,$$

the second part of the lemma follows. □

PROOF OF THEOREM 2.2. From Lemmas 4.5–4.6, it is seen that given $\varepsilon_0 > 0$, almost surely, as $n \to \infty$, $\Lambda_{n,a}(\varepsilon) \to \mathsf{L}_a(\varepsilon)$ and $\Lambda_{-n,a}(\varepsilon) \to \bar{\mathsf{L}}_a(\varepsilon)$ uniformly for $|\varepsilon| \le \varepsilon_0$, at rate $o(r(\varepsilon_0)^n)$. Since $\Lambda_{\pm n,a}(\varepsilon)$ are continuous, the uniform convergence implies that $\mathsf{L}_a(\varepsilon)$ and $\bar{\mathsf{L}}_a(\varepsilon)$ are continuous. Also, the lemmas imply that $\mathsf{L}_a(\varepsilon)$ and $\bar{\mathsf{L}}_a(\varepsilon)$ are strictly positive. By monotonicity argument, almost surely, the convergence holds simultaneously for all $\varepsilon_0 > 0$. □

4.4. *Proof of Theorem* 2.3. For $t \ne 0$, $n \ge 1$ and $\varepsilon_0 > 0$, define

$$V_{\pm n}(\varepsilon_0) = n \max_{1 \le t \le n} D_{\pm t}(\varepsilon_0) \quad \text{with } D_t(\varepsilon_0) = \max_{|\nu| \le q} \max_{a \in \mathcal{H}} \sup_{|\varepsilon| \le \varepsilon_0} \left|\frac{\psi_t^{(\nu)}(\varepsilon, a)}{\psi_t(\varepsilon, a)}\right|, \tag{4.16}$$

where $\psi_t^{(\nu)}$ is a derivative with respect to ε. Note $D_t(\varepsilon_0) \ge 1$ since the maximization in its definition takes into account $\nu = 0$.

LEMMA 4.7. *The following statements are true.*

1. *For* $\varepsilon_0 > 0$ *and* $n \ge 1$,

$$V_n(\varepsilon_0) \le n \max_{|t| \le n} [q + M_q(Z_t, \varepsilon_0)]^q. \tag{4.17}$$

2. *If Assumptions* 1–4 *hold, then* $\Pr\{\lim_n \beta^{-n} V_n(\varepsilon_0) = 0, \forall \beta > 1, \varepsilon_0 > 0\} = 1$.

PROOF. To show part 1, it suffices to show that for all ν with $|\nu| = l \le q$, and all $\varepsilon_0 > 0$ and $t \ne 0$,

$$d_{\nu,t}(\varepsilon_0) := \max_{a \in \mathcal{H}} \sup_{|\varepsilon| \le \varepsilon_0} \left|\frac{\psi_t^{(\nu)}(\varepsilon, a)}{\psi_t(\varepsilon, a)}\right| \le [l + M_l(Z_t, \varepsilon_0)]^l. \tag{4.18}$$

It is easily seen that (4.18) holds for $l = 0$. Suppose (4.18) holds if $|\nu| \le l$. Let $|\nu| = l + 1$. Without loss of generality, let $\nu = e + \mu$, where $e = (1, 0, \ldots, 0)$ and

$\mu = (\mu_1, \dots, \mu_p) \ge 0$. Let $\ell_{z,ab}(\varepsilon) = \ln f(\varphi(z, \theta_a(\varepsilon)), \theta_b(\varepsilon))$ as in Assumption 4. Then by $\psi_t^{(e)}(\varepsilon, a) = \psi_t(\varepsilon, a)\ell^{(e)}_{Z_t,\eta_t a}(\varepsilon)$,

$$\psi_t^{(\nu)}(\varepsilon, a) = \bigl[\psi_t(\varepsilon, a)\ell^{(e)}_{Z_t,\eta_t a}(\varepsilon)\bigr]^{(\mu)} = \sum_{i \le \mu} \binom{\mu}{i} \psi_t^{(i)}(\varepsilon, a)\ell^{(\nu-i)}_{Z_t,\eta_t a}(\varepsilon),$$

where $\binom{\mu}{i} = \binom{\mu_1}{i_1} \cdots \binom{\mu_p}{i_p}$.

For $i \le \mu$, $|\ell^{(\nu-i)}_{Z_t,\eta_t a}(\varepsilon)| \le M_l(Z_t, \varepsilon_0)$. Then, as $|\mu| = l$, by induction the hypothesis,

$$\begin{aligned}
\max_{a \in \mathcal{H}} \sup_{|\varepsilon| \le \varepsilon_0} \left| \frac{\psi_t^{(\nu)}(\varepsilon, a)}{\psi_t(\varepsilon, a)} \right| &\le M_l(Z_t, \varepsilon_0) \sum_{i \le \mu} \binom{\mu}{i} [|i| + M_l(Z_t, \varepsilon_0)]^{|i|} \\
&\le M_l(Z_t, \varepsilon_0) \sum_{i \le \mu} \binom{\mu}{i} [l + M_l(Z_t, \varepsilon_0)]^{|i|} \\
&= M_l(Z_t, \varepsilon_0)[|\nu| + M_l(Z_t, \varepsilon)]^l,
\end{aligned}$$

which implies (4.18). By induction, (4.18) holds for all $|\nu| \le q$.

Because $V_n(\varepsilon_0)$ is increasing in ε_0, to show part 2, it suffices to show that for each fixed $\varepsilon_0 > 0$ and $\beta > 1$, $\lim_n \beta^{-n} V_n(\varepsilon_0) = 0$ almost surely. Fix an arbitrary $c \in (1, \beta)$, such that $c^q < \beta$. By part 1 and Assumption 4, for some $p = p(\varepsilon_0) > 2$,

$$\begin{aligned}
\mathsf{Pr}\{V_n(\varepsilon_0) \ge nc^{qn}\} &\le \mathsf{Pr}\Bigl\{\max_{|t| \le n} M_q(Z_t, \varepsilon_0) \ge c^n - q\Bigr\} \\
&\le 2n\,\mathsf{Pr}\{M_q(Z_0, \varepsilon_0) \ge c^n - q\} = o(n^{-p+1}).
\end{aligned}$$

Then part 2 follows from the Borel–Cantelli lemma and $nc^{qn} = o(\beta^n)$. □

LEMMA 4.8. *Let Assumptions 1–4 hold. Fix $a, b, c \in \mathcal{H}$ and $k \ge 1$. Let*

$$W_n(\varepsilon) = L_{n-k,ab}(\varepsilon) I^{(k)}_{n,bc}(\varepsilon), \qquad n \ge k,$$

where $I^{(k)}_{n,bc}$ is defined in (4.6). Given $\nu > 0$ with $|\nu| \le q$ and $\varepsilon_0 > 0$, for $n \ge 0$,

$$\sup_{|\varepsilon| \le \varepsilon_0} \frac{|L^{(\nu)}_{n,ab}(\varepsilon)|}{L_{n,ab}(\varepsilon)} \le [V_n(\varepsilon_0)]^{|\nu|},$$

with $V_n(\varepsilon_0) := 0$ if $n = 0$, while for $n \ge k$,

$$\sup_{|\varepsilon| \le \varepsilon_0} \frac{|W^{(\nu)}_n(\varepsilon)|}{W_n(\varepsilon)} \le [V_{n-1}(\varepsilon_0)]^{|\nu|}.$$

PROOF. For $\nu = (\nu_1, \dots, \nu_p)$ with $1 \le |\nu| \le q$, it is not hard to get

$$L^{(\nu)}_{n,ab}(\varepsilon) = \mathsf{E}_\sigma\left[\mathbf{1}\{\sigma_n = b\} \sum_{l_1 + \cdots + l_n = \nu} \prod_{t=1}^n \psi_t^{(l_t)}(\varepsilon, \sigma_t) \,\middle|\, \sigma_0 = a\right].$$

For any sequence $l_1, \dots, l_n$ in the sum, at most $|\nu|$ of them are nonzero. For each $l_t > 0$, $|\psi_t^{(l_t)}(\varepsilon, \sigma_t)| \le D_t(\varepsilon_0)\psi_t(\varepsilon, \sigma_t)$ for $|\varepsilon| \le \varepsilon_0$. As a result,

$$\prod_{t=1}^{n} \bigl|\psi_t^{(l_t)}(\varepsilon, \sigma_t)\bigr| \le \Bigl[\max_{1\le t\le n} D_t(\varepsilon_0)\Bigr]^{|\nu|} \prod_{t=1}^{n} \psi_t(\varepsilon, \sigma_t).$$

On the other hand, there are $n^{\nu_1} \cdots n^{\nu_p} = n^{|\nu|}$ such sequences. Then

$$\begin{aligned}\bigl|L_{n,ab}^{(\nu)}(\varepsilon)\bigr| &\le \Bigl[n \max_{1\le t\le n} D_t(\varepsilon_0)\Bigr]^{|\nu|} \mathsf{E}_\sigma\Biggl[\mathbf{1}\{\sigma_n = b\} \prod_{t=1}^{n} \psi_t(\varepsilon, \sigma_t) \,\Bigg|\, \sigma_0 = a\Biggr] \\ &= \Bigl[n \max_{1\le t\le n} D_t(\varepsilon_0)\Bigr]^{|\nu|} L_{n,ab}(\varepsilon).\end{aligned}$$

This completes the proof of the first inequality. To show the second inequality, first,

$$W_n^{(\nu)}(\varepsilon) = \sum_{i\le\nu} \binom{\nu}{i} L_{n-k,ab}^{(i)}(\varepsilon)\bigl[I_{n,bc}^{(k)}\bigr]^{(\nu-i)}(\varepsilon).$$

Using the definition of $I_{n,bc}^{(k)}$ and following the treatment for $L_{n,ab}^{(\nu)}(\varepsilon)$,

$$\bigl|\bigl[I_{n,bc}^{(k)}\bigr]^{(\nu-i)}(\varepsilon)\bigr| \le (k-1)^{|\nu|-|i|}\Bigl[\max_{n-k+1\le t\le n-1} D_t(\varepsilon_0)\Bigr]^{|\nu|-|i|} I_{n,bc}^{(k)}(\varepsilon).$$

Combining the bound with the one for $L_{n-k,ab}^{(i)}(\varepsilon)$,

$$\begin{aligned}\bigl|W_n^{(\nu)}(\varepsilon)\bigr| &\le \Bigl[\max_{1\le t\le n-1} D_t(\varepsilon_0)\Bigr]^{|\nu|} \\ &\qquad \times \sum_{i\le\nu} \binom{\nu}{i} (n-k)^{|i|}(k-1)^{|\nu|-|i|} L_{n-k,ab}(\varepsilon) I_{n,bc}^{(k)}(\varepsilon) \\ &\le [V_{n-1}(\varepsilon_0)]^{|\nu|} W_n(\varepsilon).\end{aligned}$$

This finishes the proof. □

LEMMA 4.9. *Let Assumptions 1–4 hold. Define, for ν with $|\nu| = 1, \dots, q$,*

$$\Delta_{n,\nu}(\varepsilon) := \max_{a,b,c,d} \left| \left(\frac{L_{n,bc}}{L_{n,ac}}\right)^{(\nu)}(\varepsilon) - \left(\frac{L_{n,bd}}{L_{n,ad}}\right)^{(\nu)}(\varepsilon) \right|. \tag{4.19}$$

Then for each ν, there is an increasing deterministic function $0 \le r_\nu(\varepsilon_0) < 1$ in $\varepsilon_0 > 0$, such that almost surely, as $n \to \infty$,

$$\sup_{|\varepsilon|\le\varepsilon_0} \Delta_{n,\nu}(\varepsilon) = o\bigl(r_\nu(\varepsilon_0)^{|n|}\bigr) \qquad \text{all } \varepsilon_0 > 0.$$

PROOF. We only consider $n > 0$. The case $n < 0$ can be handled similarly. Given k, define $I^{(k)}_{n,ec}(\varepsilon)$ as in (4.6). Given $a \neq b \in \mathcal{H}$, write $W_{n,ec}(\varepsilon) = L_{n-k,ae}(\varepsilon) I^{(k)}_{n,ec}(\varepsilon)$, $W_{n,c}(\varepsilon) = \sum_e W_{n,ec}(\varepsilon)$. Then by (4.8), for $n \geq \kappa$,

$$\frac{L_{n,bc}}{L_{n,ac}} = W_{n,c}^{-1} \sum_e W_{n,ec} \frac{L_{n-k,be}}{L_{n-k,ae}}.$$

Fix $l = 1, \dots, q$. By Lemma 4.3, for $\nu \neq 0$ with $|\nu| = l$,

$$\left(\frac{L_{n,bc}}{L_{n,ac}}\right)^{(\nu)} = W_{n,c}^{-1} \sum_e W_{n,ec} \left(\frac{L_{n-k,be}}{L_{n-k,ae}}\right)^{(\nu)} + R_{n,\nu,c}, \tag{4.20}$$

where

$$R_{n,\nu,c} = \text{a linear combination of } \left[\prod_{s=1}^{m} \frac{W^{(i_s)}_{n,e_s c}}{W_{n,c}}\right]\left[\left(\frac{L_{n-k,be_1}}{L_{n-k,ae_1}}\right)^{(j)} - \left(\frac{L_{n-k,be_2}}{L_{n-k,ae_2}}\right)^{(j)}\right]$$

across $m = 2, \dots, |\nu| + 1$, $i_1, \dots, i_m \geq 0$, $0 \leq j < \nu$ with $i_1 + \cdots + i_m + j = \nu$, and $e_1, \dots, e_m \in \mathcal{H}$ with $e_1 < e_2$. Then, by the same argument that leads to (4.9),

$$\Delta_{n,\nu}(\varepsilon) \leq \gamma_n \Delta_{n-k,\nu}(\varepsilon) + 2 \max_c |R_{n,\nu,c}(\varepsilon)|, \tag{4.21}$$

where γ_n is given in (4.9).

The rest of the proof is by induction on l. First, let $|\nu| = 1$. By Lemma 4.3,

$$R_{n,\nu,c} = W_{n,c}^{-2} \sum_{e_1 < e_2} (W^{(\nu)}_{n,e_1 c} W_{n,e_2 c} - W_{n,e_1 c} W^{(\nu)}_{n,e_2 c}) \left(\frac{L_{n-k,be_1}}{L_{n-k,ae_1}} - \frac{L_{n-k,be_2}}{L_{n-k,ae_2}}\right).$$

Fix $\varepsilon_0 > 0$. By Lemma 4.8, for $|\varepsilon| \leq \varepsilon_0$, $|W^{(\nu)}_{n,ec}(\varepsilon)| \leq V_{n-1}(\varepsilon_0) W_{n,ec}(\varepsilon)$. Then

$$\begin{aligned} |R_{n,\nu,c}(\varepsilon)| &\leq W_{n,c}^{-2} \sum_{e_1 < e_2} 2 V_{n-1}(\varepsilon_0) W_{n,e_1 c} W_{n,e_2 c} \Delta_{n-k}(\varepsilon) \\ \Longrightarrow \quad & \max_c |R_{n,\nu,c}(\varepsilon)| \leq V_{n-1}(\varepsilon_0) \Delta_{n-k}(\varepsilon). \end{aligned} \tag{4.22}$$

By Lemma 4.5, there is increasing deterministic $r = r(\varepsilon_0) \in (0, 1)$, such that $\sup_{|\varepsilon| \leq \varepsilon_0} \Delta_n(\varepsilon) \leq r^n$ for $n \gg 1$. Fix $\beta \in (1, 1/r)$. Then by (4.21), (4.22) and part 2 of Lemma 4.7, almost surely, for $n \gg 1$ and $|\varepsilon| \leq \varepsilon_0$,

$$\Delta_{n,\nu}(\varepsilon) \leq \gamma_n \Delta_{n-k,\nu}(\varepsilon) + \beta^n r^{n-k} \leq \Delta_{n-k,\nu}(\varepsilon) + \beta^n r^{n-k}. \tag{4.23}$$

Let $k = 1$ to get $\Delta_{n,\nu}(\varepsilon) \leq \Delta_{n-1,\nu}(\varepsilon) + \beta^n r^{n-1}$. So by induction, for $s \leq n$,

$$\Delta_{n,\nu}(\varepsilon) \leq \Delta_{s,\nu}(\varepsilon) + \beta \sum_{t=s}^{n-1} (\beta r)^t \leq \Delta_{s,\nu}(\varepsilon) + \frac{\beta}{1 - \beta r} (\beta r)^s. \tag{4.24}$$

Next let $k = \kappa$. By the same argument that leads to (4.15), r can be chosen in such a way that there is a sequence $n_s = n_s(Z, \varepsilon_0)$ that satisfy (4.13) and $\gamma_{n_s} \leq r$. By the first inequality in (4.23), for $s \gg 1$,

$$\Delta_{n_s,\nu}(\varepsilon) \leq r \Delta_{n_s-\kappa,\nu}(\varepsilon) + \beta^{n_s} r^{n_s-\kappa} \leq r \Delta_{n_s-\kappa,\nu}(\varepsilon) + \beta^{\kappa} (\beta r)^{n_{s-1}}.$$

Let $n = n_s - \kappa$ and $s = n_{s-1}$ in (4.24) and combine it with the last equality to get

$$\Delta_{n_s,\nu}(\varepsilon) \le r\Delta_{n_{s-1},\nu}(\varepsilon) + c(\beta r)^{n_{s-1}},$$

where $c = \beta^\kappa + \beta/(1 - r\beta)$. Then by induction and the fact that $n_s \ge \kappa s$,

$$\begin{aligned}\Delta_{n_s,\nu}(\varepsilon) &\le r^{s-1}\Delta_{n_1,\nu}(\varepsilon) + c\sum_{t=1}^{s-1} r^{s-t-1}(\beta r)^{n_t} \\ &\le r^{s-1}\Delta_{n_1,\nu}(\varepsilon) + c\sum_{t=1}^{s-1} r^{s-t-1}(\beta r)^{t} \\ &\le r^{s-1}\Delta_{n_1,\nu}(\varepsilon) + cs(\beta r)^{s-1}.\end{aligned}$$

Now for any $n_s \le n < n_{s+1}$, by (4.24) and the above inequality,

$$\Delta_{n,\nu}(\varepsilon) \le r^{s-1}\Delta_{n_1,\nu} + \left(\frac{\beta}{1 - r\beta} + cs\right)(\beta r)^{s-1}.$$

Since for $s \gg 1$, $s + 1 \ge \frac{p_0}{2\kappa} n_{s+1} \ge \frac{p_0}{2\kappa} n$, it can be seen that $\Delta_{n,\nu}(\varepsilon) = O(c^n)$, with $c = (\beta r)^{p_0/(2\kappa)} < 1$. Since $\beta \in (1, 1/r)$ is arbitrary, it follows that for a given ε_0 and any $r_1 > r_* := r^{p_0/(2\kappa)}$, say $r_1 = r_1(\varepsilon_0) = (1 + r_*)/2$, $\sup_{|\varepsilon|\le\varepsilon_0} \Delta_{n,\nu}(\varepsilon) = o(r_1^n)$ almost surely. By monotonicity, it can be seen that almost surely, the exponentially fast convergence holds simultaneously for all ε_0.

Now let $|\nu| > 1$. To bound $R_{n,\nu,c}(\varepsilon)$, for $s = 2, \dots, |\nu| + 1$, and p-tuples of nonnegative integers, $i_1, \dots, i_s$, j, $i_1 + \cdots + i_s = \nu - j < \nu$, and $e_1, \dots, e_s \in \mathcal{H}$, by Lemma 4.8, for $|\varepsilon| \le \varepsilon_0$,

$$\left|W_{n,e_1c}^{(i_1)} \cdots W_{n,e_sc}^{(i_s)}\right| \le \prod_{k=1}^{s} [V_{n-1}(\varepsilon_0)]^{|i_k|} W_{n,e_kc} \le W_{n,c}^s [V_{n-1}(\varepsilon_0)]^{|\nu|}$$

so in place of (4.22), we have

$$\max_c |R_{n,\nu,c}(\varepsilon)| \le C_\nu [V_{n-1}(\varepsilon_0)]^{|\nu|} \times \sum_{j<\nu} \Delta_{n-k,j}(\varepsilon), \tag{4.25}$$

where $\Delta_{n-k,0}(\varepsilon) := \Delta_{n-k}(\varepsilon)$ and $C_\nu > 0$ is some constant only depending on ν.

Suppose it has been shown that for each $j < \nu$, there is $r_j = r_j(\varepsilon_0) \in (0, 1)$, such that $\sup_{|\varepsilon|\le\varepsilon_0} \Delta_{n,j}(\varepsilon) = o(r_j^n)$. Then using (4.21) and (4.25) and following the argument for $\Delta_{n,j}(\varepsilon)$ with $|j| = 1$, $\sup_{|\varepsilon|\le\varepsilon_0} \Delta_{n,\nu}(\varepsilon) = o(r_\nu^n)$ for some $r_\nu = r_\nu(\varepsilon_0) \in (0, 1)$. By induction, the exponential rate of convergence holds for all ν with $|\nu| \le q$. Again, from the construction, r_ν only depends on the distributional properties of Z and η and hence is deterministic. $\square$

Set $k = 1$ in (4.20). For $n \ge \kappa$ and $a, b, c \in \mathcal{H}$,

$$\min_e \left(\frac{L_{n-1,be}}{L_{n-1,ae}}\right)^{(\nu)} - |R_{n,\nu,c}| \le \left(\frac{L_{n,bc}}{L_{n,ac}}\right)^{(\nu)} \le \max_e \left(\frac{L_{n-1,be}}{L_{n-1,ae}}\right)^{(\nu)} + |R_{n,\nu,c}|,$$

giving

$$\left|\left(\frac{L_{n,bc}}{L_{n,ac}}\right)^{(\nu)}(\varepsilon) - \left(\frac{L_{n-1,bc}}{L_{n-1,ac}}\right)^{(\nu)}(\varepsilon)\right| \le \Delta_{n-1,\nu}(\varepsilon) + 2|R_{n,\nu,c}(\varepsilon)|. \tag{4.26}$$

COROLLARY 4.10. *Let assumptions 1–4 hold. Then almost surely, as* $s \ge n \to \infty$,

$$\max_{a\in\mathcal{H}} \sup_{|\varepsilon|\le\varepsilon_0} |R_{n,\nu,c}(\varepsilon)| = o(r_\nu^n(\varepsilon_0))$$

$$\max_{a,b,c\in\mathcal{H}} \sup_{|\varepsilon|\le\varepsilon_0} \left|\left(\frac{L_{n,bc}}{L_{n,ac}}\right)^{(\nu)}(\varepsilon) - \left(\frac{L_{s,bc}}{L_{s,ac}}\right)^{(\nu)}(\varepsilon)\right| = o(r_\nu^n(\varepsilon_0)),$$

for all $\varepsilon_0 > 0$ *and* ν *with* $1 \le |\nu| \le, q$, *and likewise for* $\bar{L}_{n,ab}$, *where* $r_\nu(\varepsilon_0)$ *are given in Lemma 4.9.*

PROOF. The first inequality is already shown in the proof of Lemma 4.9. The second one follows from summing the inequality in (4.26) over $n+1, \ldots, s$ and applying the first inequality and Lemma 4.9. □

PROOF OF THEOREM 2.3. Let $r_\nu(\varepsilon_0)$ be as in Lemma 4.9. For $e \in \mathcal{H}$, denote

$$\omega_{n,e} = L_{n,\imath e}, \qquad \omega_n = \sum_e \omega_{n,e}.$$

Then for $a \in \mathcal{H}$, $\Lambda_{n,a} = \omega_n^{-1} \sum_e \omega_{n,e}(\frac{L_{n,ae}}{L_{n,\imath e}})$ and similar to (4.20),

$$\Lambda_{n,a}^{(\nu)} = \omega_n^{-1} \sum_e \omega_{n,e} \left(\frac{L_{n,ae}}{L_{n,\imath e}}\right)^{(\nu)} + T_{n,\nu}, \tag{4.27}$$

where $T_{n,\nu}$ is a linear combination of

$$\omega_n^{-m} \omega_{n,e_1}^{(i_1)} \cdots \omega_{n,e_m}^{(i_m)} \left[\left(\frac{L_{n-k,ae_1}}{L_{n-k,\imath e_1}}\right)^{(j)} - \left(\frac{L_{n-k,ae_2}}{L_{n-k,\imath e_2}}\right)^{(j)}\right]$$

across $m = 2, \ldots, |\nu|+1$, $0 \le j < \nu$, $i_1, \ldots, i_m \ge 0$ with $i_1 + \cdots + i_m + j = \nu$, and $e_1, \ldots, e_m \in \mathcal{H}$ with $e_1 < e_2$. Fix any $b \in \mathcal{H}$. From the above formulas,

$$\left|\Lambda_{n,a}^{(\nu)} - \left(\frac{L_{n,ab}}{L_{n,\imath b}}\right)^{(\nu)}\right| \le \Delta_{n,\nu} + |T_{n,\nu}|. \tag{4.28}$$

Following the treatment of $R_{n,\nu,c}$ in (4.25), except that we have to use the first inequality in Lemma 4.8, it can be seen that

$$|T_{n,\nu}(\varepsilon)| \le C_\nu [V_n(\varepsilon_0)]^{|\nu|} \times \sum_{j<\nu} \Delta_{n-k,j}(\varepsilon), \qquad |\varepsilon| \le \varepsilon_0, \tag{4.29}$$

yielding $\max_{|\varepsilon|\le\varepsilon_0} |T_{n,\nu}(\varepsilon)| = o(r_\nu^n)$. Now for $s \ne n$, by (4.28), it is not hard to get

$$\begin{aligned}(4.30)\qquad |\Lambda_{s,a}^{(\nu)} - \Lambda_{n,a}^{(\nu)}| &\le \Delta_{s,\nu} + |T_{s,\nu}| + \Delta_{n,\nu} + |T_{n,\nu}| \\ &\quad + \left|\left(\frac{L_{s,ab}}{L_{s,\imath b}}\right)^{(\nu)} - \left(\frac{L_{n,ab}}{L_{n,\imath b}}\right)^{(\nu)}\right|.\end{aligned}$$

Then by Lemma 4.9 and Corollary 4.10,

$$\sup_{|\varepsilon|\le\varepsilon_0} |\Lambda_{s,a}^{(\nu)}(\varepsilon) - \Lambda_{n,a}^{(\nu)}(\varepsilon)| = o(r_\nu^{s\wedge n}(\varepsilon_0)), \qquad a \in \mathcal{H}.$$

Since $\#\mathcal{H} < \infty$, almost surely, the rate holds simultaneously for all $a \in \mathcal{H}$. □

4.5. *Proof of Theorem* 2.4. Since the parameter κ in Assumption 2 equals 1, $P_{n-1,n}(a,b) \in [\phi_*, 1-\phi_*]$ for $a, b \in \mathcal{H}$ and $n \in \mathbb{Z}$, with $0 < \phi_* < 1$ as in Assumption 2. Consequently,

$$(4.31)\qquad \gamma = 1 - \inf_n \frac{\min_{c,d,e}((P_{n-1,n}(e,d))/(P_{n-1,n}(e,c)))}{\max_{c,d,e}((P_{n-1,n}(e,d))/(P_{n-1,n}(e,c)))} \in \left[0, 1 - \frac{\phi_*}{1-\phi_*}\right].$$

For $a, e \in \mathcal{H}$, by (4.3), $L_{1,ae}(\varepsilon) = P_{01}(a,e)\psi_1(\varepsilon, e)$, giving

$$(4.32)\qquad \frac{L_{1,be}(\varepsilon)}{L_{1,ae}(\varepsilon)} \equiv \frac{P_{01}(b,e)}{P_{01}(a,e)} \le \frac{1-\phi_*}{\phi_*} \qquad \forall \varepsilon.$$

Then by Lemma 4.5,

$$(4.33)\qquad \frac{\phi^*}{1-\phi^*} \le \Lambda_{n,a}(\varepsilon) \le \frac{1-\phi^*}{\phi^*}.$$

This together with dominated convergence shows part 1 of Theorem 2.4. To prove part 2, we need several lemmas.

LEMMA 4.11. *Fix* $\varepsilon_0 > 0$. *Let* γ *and* ϕ_* *be as in* (4.31) *and* $\alpha = \phi_*^{-1} - 1$. *There is a constant* $C > 0$, *such that if* $1 \le |\nu| = l \le q$, $|\varepsilon| \le \varepsilon_0$ *and* $n \ge 1$, *then*

$$\begin{aligned}(4.34)\qquad &|\Lambda_{n,a}^{(\nu)}(\varepsilon) - \Lambda_{n-1,a}^{(\nu)}(\varepsilon)| \\ &\quad \le C\alpha\gamma^{(n-l-1)\vee 1} n^{l(l+2)} \sum_{t=1}^n [q + M_q(Z_t, \varepsilon_0)]^{ql(l+1)/2}.\end{aligned}$$

PROOF. First, by (4.32) and the definitions of Δ_n and $\Delta_{n,\nu}$ in (4.7) and (4.19),

$$\begin{aligned}(4.35)\qquad &\Delta_1(\varepsilon) \equiv \max_{a,b,c,d} \left|\frac{P_{01}(b,c)}{P_{01}(a,c)} - \frac{P_{01}(b,d)}{P_{01}(a,d)}\right| \le \frac{\gamma(1-\phi_*)}{\phi_*}, \\ &\Delta_{1,\nu}(\varepsilon) \equiv 0, \qquad \nu > 0.\end{aligned}$$

By (4.6), $I^{(1)}_{n,ec} = P_{n-1,n}(e, c)$, so (4.9) gives $\Delta_n(\varepsilon) \le \gamma \Delta_{n-1}(\varepsilon)$. Thus,

$$\Delta_n(\varepsilon) \le \alpha\gamma^n \qquad \forall n \ge 1, \varepsilon > 0. \tag{4.36}$$

Let $R_{n,\nu,c}(\varepsilon)$ be as in (4.20) and

$$\bar{\Delta}_{n,l}(\varepsilon) = \max_{|\nu|=l} \Delta_{n,\nu}(\varepsilon).$$

Recall the definition of $V_n(\varepsilon_0)$ in (4.16). For brevity, write $v_n = V_n(\varepsilon_0)$. By (4.25), there are constants $c_l > 1$, such that

$$\max_{|\nu|=l,c} |R_{n,\nu,c}(\varepsilon)| \le \frac{1}{2} c_l v^l_{n-1} \sum_{i=0}^{l-1} \bar{\Delta}_{n-1,i}(\varepsilon), \tag{4.37}$$

for $l = 1, \dots, q$, $n \ge 1$, $\varepsilon_0 > 0$ and $|\varepsilon| \le \varepsilon_0$. Then by (4.21), for $n \ge 0$,

$$\bar{\Delta}_{n+1,l}(\varepsilon) \le \gamma \bar{\Delta}_{n,l}(\varepsilon) + c_l v^l_n \sum_{i=0}^{l-1} \bar{\Delta}_{n,i}(\varepsilon). \tag{4.38}$$

We show by induction that for $l \ge 1$ and $n \ge 0$,

$$\bar{\Delta}_{n+1,l}(\varepsilon) \le \alpha\gamma^{(n+1-l)\vee 1} n c_l v^l_n \prod_{i=1}^{l-1} (1 + n c_i v^i_n), \tag{4.39}$$

where $\bar{\Delta}_{n+1,0}(\varepsilon) = \Delta_{n+1}(\varepsilon)$.

By (4.35), (4.39) holds for $n = 0$ and $l \ge 1$. Let $n \ge 1$ next. If $l = 1$, then by (4.36) and (4.38),

$$\bar{\Delta}_{n+1,1}(\varepsilon) \le \gamma \bar{\Delta}_{n,1}(\varepsilon) + c_1 v_n \Delta_n(\varepsilon) \le \gamma \bar{\Delta}_{n,1}(\varepsilon) + \alpha\gamma^n c_1 v_n,$$

and by induction on n and (4.35),

$$\bar{\Delta}_{n+1,1}(\varepsilon) \le \gamma^n \bar{\Delta}_{1,1}(\varepsilon) + \alpha\gamma^n c_1 \sum_{s=1}^{n} v_s = \alpha\gamma^n c_1 \sum_{s=1}^{n} v_s \le \alpha\gamma^n c_1 n v_n.$$

So (4.39) holds for $l = 1$. Suppose (4.39) holds for $1 \le l < k$. By $\gamma \in (0, 1)$,

$$\begin{aligned}
\sum_{i=0}^{k-1} \bar{\Delta}_{n,i}(\varepsilon_0) &= \Delta_n(\varepsilon_0) + \sum_{i=1}^{k-1} \bar{\Delta}_{n,i}(\varepsilon_0) \\
&\le \alpha \left\{ \gamma^n + \sum_{i=1}^{k-1} \gamma^{(n-i)\vee 1} c_i (n-1) v^i_{n-1} \prod_{h=1}^{i-1} [1 + c_h (n-1) v^h_{n-1}] \right\} \\
&\le \alpha\gamma^{(n+1-k)\vee 1} \left\{ 1 + \sum_{i=1}^{k-1} c_i n v^i_n \prod_{h=1}^{i-1} (1 + c_h n v^h_n) \right\} \\
&= \alpha\gamma^{(n+1-k)\vee 1} \prod_{i=1}^{k-1} (1 + c_i n v^i_n),
\end{aligned} \tag{4.40}$$

so by (4.38),

$$\bar{\Delta}_{n+1,k}(\varepsilon) \le \gamma \bar{\Delta}_{n,k}(\varepsilon) + \alpha\gamma^{(n+1-k)\vee 1} c_k v_n^k \prod_{i=1}^{k-1}(1 + c_i n v_n^i).$$

By induction on n, it is seen that $\bar{\Delta}_{n,k}(\varepsilon)$ satisfies (4.39). As a by-product, by (4.37) and (4.40),

$$\max_{|\nu|=l,c} |R_{n,\nu,c}(\varepsilon)| \le \frac{1}{2}\alpha\gamma^{(n-l)\vee 1} c_l v_{n-1}^l \prod_{i=1}^{l-1}(1 + c_i n v_{n-1}^i). \tag{4.41}$$

Combining (4.26), (4.39) and (4.41), for any $|\nu| = l$,

$$\begin{aligned}\left| \left(\frac{L_{n,bc}}{L_{n,ac}}\right)^{(\nu)}(\varepsilon) - \left(\frac{L_{n-1,bc}}{L_{n-1,ac}}\right)^{(\nu)}(\varepsilon) \right| &\le \Delta_{n-1,l}(\varepsilon) + 2|R_{n,\nu,c}(\varepsilon)| \\ &\le \alpha\gamma^{(n-l-1)\vee 1} n c_l v_{n-1}^l \prod_{i=1}^{l-1}(1 + c_i n v_{n-1}^i).\end{aligned}$$

Let $T_{n,\nu}$ be as in (4.27). With (4.39) being established now, by (4.29), we get the following bounds similar to (4.41):

$$\max_{|\nu|=l} |T_{n,\nu}(\varepsilon)| \le \frac{1}{2}\alpha\gamma^{(n-l)\vee 1} c_l v_n^l \prod_{i=1}^{l-1}(1 + n c_i v_{n-1}^i). \tag{4.42}$$

Combine (4.26), (4.30) and the above inequalities. It is seen that for some constants $C_l > 1$,

$$|\Lambda_{n,a}^{(\nu)} - \Lambda_{n-1,a}^{(\nu)}| \le C_l \alpha\gamma^{(n-l-1)\vee 1} n^l v_n^{l(l+1)/2}.$$

Then applying Lemma 4.7 to $v_n = V_n(\varepsilon_0)$, the lemma is proved. □

Now for $n \ge 1$, $|\Lambda_{n,a}^{(\nu)}(\varepsilon)| \le |\Lambda_{1,a}^{(\nu)}(\varepsilon)| + \sum_{k=2}^n |\Lambda_{n,a}^{(\nu)}(\varepsilon) - \Lambda_{n-1,a}^{(\nu)}(\varepsilon)|$. Letting $k = 1$ in (4.28) and (4.29) and combining them with (4.32) and (4.35), it is seen that $|\Lambda_{1,a}^{(\nu)}(\varepsilon)| \le C|V_1(\varepsilon)|^{|\nu|}$, where C is a constant. Together with (4.34), this implies there is a constant $C_l = C_l(\gamma, \phi_*)$, such that for ν with $1 \le |\nu| = l \le q$,

$$|\Lambda_{n,a}^{(\nu)}(\varepsilon)| \le C_l \sum_{t=1}^{\infty} \beta_{l,t}[q + M_q(Z_t, \varepsilon_0)]^{ql(l+1)/2}, \qquad |\varepsilon| \le \varepsilon_0, \tag{4.43}$$

where $\beta_{l,t} = \sum_{k=t+1}^{\infty} \gamma^k k^{l(l+1)} = o((c\gamma)^t)$ for any $0 < c < 1/\gamma$.

Part 2 of Theorem 2.4 is an immediate consequence of the next result.

LEMMA 4.12. *Let $\varepsilon_0 > 0$. Almost surely, the following statements hold for all $|\varepsilon| \le \varepsilon_0$, $n \ge 1$ and ν with $1 \le |\nu| \le q$.*

1. $\mathsf{E}[(\ln \Lambda_{n,a})^{(\nu)}(\varepsilon) \mid \eta]$ *and* $\mathsf{E}[(\ln \Lambda_{n,a})(\varepsilon) \mid \eta]^{(\nu)}$ *both exist and are equal.*
2. *As* $n \to \infty$, $\mathsf{E}[(\ln \Lambda_{n,a})^{(\nu)}(\varepsilon) \mid \eta] \to \mathsf{E}[(\ln \mathsf{L}_a)^{(\nu)}(\varepsilon) \mid \eta]$.
3. *As* $n \to \infty$, $(\mathsf{E}[\ln \Lambda_{n,a}(\varepsilon) \mid \eta])^{(\nu)} \to (\mathsf{E}[\ln \mathsf{L}_a(\varepsilon) \mid \eta])^{(\nu)}$.

PROOF. 1. It is not hard to see that $(\ln \Lambda_{n,a})^{(\nu)}(\varepsilon)$ is a linear combination of products of the form

$$h_{n,\nu_1,\dots,\nu_s}(\varepsilon) := \frac{\Lambda_{n,a}^{(\nu_1)}(\varepsilon)\cdots\Lambda_{n,a}^{(\nu_s)}(\varepsilon)}{\Lambda_{n,a}(\varepsilon)^s}, \qquad \nu_k > 0, \qquad \nu_1 + \cdots + \nu_s = \nu.$$

By (4.33) and (4.43),

$$|h_{n,\nu_1,\dots,\nu_s}(\varepsilon)| \le \zeta := C \prod_{k=1}^{s} \sum_{t=1}^{\infty} \beta_{l,t}[q + M_q(Z_t, \varepsilon_0)]^{q|\nu_k|(|\nu_k|+1)/2}, \qquad |\varepsilon| \le \varepsilon_0,$$

with $C = C(\gamma, \phi_*)$ a constant. As $\sum_k |\nu_k|(|\nu_k| + 1) \le |\nu|(|\nu| + 1)$, by Assumption 5 and the independence of Z_t, $\mathsf{E}\zeta < \infty$. Note that ζ is independent of η. It follows that $(\ln \Lambda_{n,a})^{(\nu)}(\varepsilon)$ for all n and $|\varepsilon| \le \varepsilon_0$ are bounded by a single random variable that has a finite expectation and is independent of η. This implies $\mathsf{E}[(\ln \Lambda_{n,a})^{(\nu)}(\varepsilon) \mid \eta]$ exists, and together with $\ln \Lambda_{n,a} \in C^{(q)}$, implies the rest of part 1 through dominated convergence.

2. By Theorems 2.1 and 2.2, $\Lambda_{n,a}^{(\nu)}(\varepsilon)$ converges as $n \to \infty$ for all ε. By Lemma 4.11 and (4.33), $(\ln \Lambda_{n,a})^{(\nu)}(\varepsilon)$ converges. Then the claim follows from dominated convergence.

3. Consider $h_{n,\nu_1,\dots,\nu_s}(\varepsilon)$ again. By Lemma 4.11 and (4.33), it can be seen that for $\nu_1, \dots, \nu_s > 0$ with $\nu_1 + \cdots + \nu_s = \nu$, $|h_{n+1,\nu_1,\dots,\nu_s}(\varepsilon) - h_{n,\nu_1,\dots,\nu_s}(\varepsilon)| \le C\gamma_1^n \zeta$ holds for $|\varepsilon| \le \varepsilon_0$, where $C > 0$ $\gamma_1 \in (\gamma, 1)$ are constants and $\zeta > 0$ is a random variable independent of η with $\mathsf{E}\zeta < \infty$. As a result, $\mathsf{E}[(\ln \Lambda_{n,a})^{(\nu)}(\varepsilon) \mid \eta]$ converges uniformly on each compact set of ε. Together with part 1, this implies part 3. $\square$

4.6. *Proof for the binary case.* The following simple identity will be repeatedly used. For any function F on $\{0, 1\}$, denote $dF = F(1) - F(0)$. Then for s, $t \in \mathbb{Z}$,

$$\mathsf{E}_\sigma[F(\sigma_t) \mid \sigma_s = 1] - \mathsf{E}_\sigma[F(\sigma_t) \mid \sigma_s = 0] = D_{st}\, dF, \tag{4.44}$$

$$F(0) - \mathsf{E}_\sigma[F(\sigma_t) \mid \sigma_s = 0] = -P_{st}(0, 1)\, dF. \tag{4.45}$$

Define for $t \in \mathbb{Z}$ and $n \ge 1$,

$$\ell_t(\varepsilon, a) = \ln \psi_t(\varepsilon, a), \qquad S_n(\varepsilon) = \sum_{t=1}^{n} \ell_t(\varepsilon, \sigma_t).$$

Then $\lambda_n(\varepsilon) = \ln \mathsf{E}_\sigma[e^{S_n(\varepsilon)} \mid \sigma_0 = 1] - \ln \mathsf{E}_\sigma[e^{S_n(\varepsilon)} \mid \sigma_0 = 0]$ in (3.3).

PROOF OF THEOREM 3.1. For $n \ge 1$, by (4.44),

$$\begin{aligned}\lambda_n'(0) &= \mathsf{E}_\sigma[S_n'(0) \mid \sigma_0 = 1] - \mathsf{E}_\sigma[S_n'(0) \mid \sigma_0 = 0] \\ &= \sum_{t=1}^n \{\mathsf{E}_\sigma[\ell_t'(0, \sigma_t) \mid \sigma_0 = 1] - \mathsf{E}_\sigma[\ell_t'(0, \sigma_t) \mid \sigma_0 = 0]\} \\ &= \sum_{t=1}^n D_{0t} d_t'(0).\end{aligned}$$

By Theorems 2.2 and 2.3, letting $n \to \infty$, (3.7) follows. To get $\mathsf{r}''(0)$, for $n \ge 1$,

$$\begin{aligned}\lambda_n''(0) = {}& \mathsf{E}_\sigma[S_n''(0) \mid \sigma_0 = 1] - \mathsf{E}_\sigma[S_n''(0) \mid \sigma_0 = 0] + \mathsf{Var}_\sigma[S_n'(0) \mid \sigma_0 = 1] \\ &- \mathsf{Var}_\sigma[S_n'(0) \mid \sigma_0 = 0].\end{aligned}$$

Similar to the calculation of $\mathsf{r}'(0)$,

$$\lim_{n\to\infty} \{\mathsf{E}_\sigma[S_n''(0) \mid \sigma_0 = 1] - \mathsf{E}_\sigma[S_n''(0) \mid \sigma_0 = 0]\} = \sum_{t=1}^\infty D_{0t} d_t''(0).$$

On the other hand, denoting by δ_t the random variable $\ell_t'(0, \sigma_t)$,

$$\mathsf{Var}_\sigma[S_n'(0) \mid \sigma_0] = \sum_{t=1}^n \mathsf{Var}_\sigma(\delta_t \mid \sigma_0) + 2 \sum_{1\le s<t\le n} \mathsf{Cov}_\sigma(\delta_s, \delta_t \mid \sigma_0).$$

Given $1 \le s \le t \le n$, let $F(\sigma_s) = \delta_s \mathsf{E}_\sigma[\delta_t \mid \sigma_s]$. By $\mathsf{E}_\sigma[\delta_s\delta_t \mid \sigma_0] = \mathsf{E}_\sigma[F(\sigma_s) \mid \sigma_0]$ and (4.44),

$$\mathsf{E}_\sigma[\delta_s\delta_t \mid \sigma_0 = 1] - \mathsf{E}_\sigma[\delta_s\delta_t \mid \sigma_0 = 0] = D_{0s}\, dF.$$

Similarly, by (4.44), $\mathsf{E}_\sigma[\delta_t \mid \sigma_s = 1] = \mathsf{E}_\sigma[\delta_t \mid \sigma_s = 0] + D_{st} d_t'(0)$. Then, as $\ell_s'(0, 1) = \ell_s'(0, 0) + d_s'(0)$,

$$\begin{aligned}dF &= F(1) - F(0) = \ell_s'(0, 1)\mathsf{E}_\sigma(\delta_t \mid \sigma_s = 1) - \ell_s'(0, 0)\mathsf{E}_\sigma(\delta_t \mid \sigma_s = 0) \\ &= \mathsf{E}_\sigma(\delta_t \mid \sigma_s = 0) d_s'(0) + D_{st}\ell_s'(0, 0) d_t'(0) + D_{st} d_s'(0) d_t'(0)\end{aligned}$$

and likewise,

$$\begin{aligned}&\mathsf{E}_\sigma(\delta_s \mid \sigma_0 = 1)\mathsf{E}_\sigma(\delta_t \mid \sigma_0 = 1) - \mathsf{E}_\sigma(\delta_s \mid \sigma_0 = 0)\mathsf{E}_\sigma(\delta_t \mid \sigma_0 = 0) \\ &\quad = D_{0s}\mathsf{E}_\sigma(\delta_t \mid \sigma_0 = 0) d_s'(0) + D_{0t}\mathsf{E}_\sigma(\delta_s \mid \sigma_0 = 0) d_t'(0) + D_{0s} D_{0t} d_s'(0) d_t'(0).\end{aligned}$$

Combining the above identities,

$$\mathsf{Cov}_\sigma(\delta_s, \delta_t \mid \sigma_0 = 1) - \mathsf{Cov}_\sigma(\delta_s, \delta_t \mid \sigma_0 = 0) = I_1 + I_2 + I_3,$$

with

$$\begin{cases} I_1 = D_{0s}[\mathsf{E}_\sigma(\delta_t \mid \sigma_s = 0) - \mathsf{E}_\sigma(\delta_t \mid \sigma_0 = 0)]d_s'(0), \\ I_2 = [D_{0s}D_{st}\ell_s'(0,0) - D_{0t}\mathsf{E}_\sigma(\delta_s \mid \sigma_0 = 0)]d_t'(0), \\ I_3 = D_{0s}(D_{st} - D_{0t})d_s'(0)d_t'(0). \end{cases}$$

By conditioning on σ_s,

$$\begin{aligned} &\mathsf{E}_\sigma(\delta_t \mid \sigma_s = 0) - \mathsf{E}_\sigma(\delta_t \mid \sigma_0 = 0) \\ &\quad = \mathsf{E}_\sigma(\delta_t \mid \sigma_s = 0) - \mathsf{E}_\sigma[\mathsf{E}_\sigma(\delta_t \mid \sigma_s) \mid \sigma_0 = 0] \\ &\quad \overset{\text{(a)}}{=} -P_{0s}(0,1)[\mathsf{E}_\sigma(\delta_t \mid \sigma_s = 1) - \mathsf{E}_\sigma(\delta_t \mid \sigma_s = 0)] \\ &\quad \overset{\text{(b)}}{=} -D_{st}P_{0s}(0,1)d_t'(0), \end{aligned}$$

where (a) is due to (4.45), and (b) due to (4.44). By (3.6), $D_{0s}D_{st} = D_{0t}$. Therefore, $I_1 = -D_{0t}P_{0s}(0,1)d_s'(0)d_t'(0)$. Likewise,

$$I_2 = D_{0t}[\ell_s'(0,0) - \mathsf{E}_\sigma(\delta_s \mid \sigma_0 = 0)]d_t'(0) = -D_{0t}P_{0s}(0,1)d_s'(0)d_t'(0)$$

and $I_3 = D_{0t}(1 - D_{0s})d_s'(0)d_t'(0)$. Then (3.8) follows from

$$\begin{aligned} &\mathsf{Cov}_\sigma(\delta_s, \delta_t \mid \sigma_0 = 1) - \mathsf{Cov}_\sigma(\delta_s, \delta_t \mid \sigma_0 = 0) \\ &\quad = D_{0t}[P_{0s}(1,0) - P_{0s}(0,1)]d_s'(0)d_t'(0) \end{aligned}$$

and Theorems 2.2 and 2.3. □

To prove the rest of the results, recall $\lambda(x, \vartheta) = \ln f(x, \vartheta)$.

PROOF OF PROPOSITION 3.2. Given t, Z and η, $\ell_t(\varepsilon, a)$ is a composite of functions $\lambda(x, \vartheta)$, $\varphi(Z_t, v)$, $\theta_a(\varepsilon)$ and $\theta_{\eta_t}(\varepsilon)$, such that

$$\ell_t(\varepsilon, a) = \lambda(\varphi(Z_t, \theta_{\eta_t}(\varepsilon)), \theta_a(\varepsilon)),$$

so by the chain rule for differentiation,

$$\ell_t'(\varepsilon, a) = \frac{\partial\lambda(x,\vartheta)}{\partial x}\frac{\partial\varphi(Z_t, v)}{\partial v}\theta_{\eta_t}'(\varepsilon) + \frac{\partial\lambda(x,\vartheta)}{\partial\vartheta}\theta_a'(\varepsilon),$$

where the right-hand side is evaluated at $x = \varphi(Z_t, v)$, $v = \theta_{\eta_t}(\varepsilon)$, and $\vartheta = \theta_a(\varepsilon)$. Since $\theta_1(0) = \theta_0(0) = 0$, the first summand on the right-hand side takes the same value for $a = 0, 1$. Therefore, (3.9) holds.

Likewise,

$$\begin{aligned} \ell_t''(\varepsilon, a) = {} & \frac{\partial^2\lambda}{\partial x^2}\left[\frac{\partial\varphi}{\partial v}\right]^2\theta_{\eta_t}'(\varepsilon)^2 + 2\frac{\partial^2\lambda}{\partial x\,\partial\vartheta}\frac{\partial\varphi}{\partial v}\theta_{\eta_t}'(\varepsilon)\theta_a'(\varepsilon) + \frac{\partial^2\lambda}{\partial\vartheta^2}\theta_a'(\varepsilon)^2 \\ & + \frac{\partial\lambda}{\partial x}\frac{\partial^2\varphi}{\partial v^2}\theta_{\eta_t}'(\varepsilon)^2 + \frac{\partial\lambda}{\partial x}\frac{\partial\varphi}{\partial v}\theta_{\eta_t}''(\varepsilon) + \frac{\partial\lambda}{\partial\vartheta}\theta_a''(\varepsilon), \end{aligned}$$

where again the right-hand side is evaluated at $x = \varphi(Z_t, v)$, $v = \theta_{\eta_t}(\varepsilon)$, and $\vartheta = \theta_a(\varepsilon)$. Then (3.10) follows. □

PROOF OF PROPOSITION 3.3. We shall first show for any t,

$$\mathsf{E}[d_t'(0) \mid \eta] = 0, \tag{4.46}$$

$$\mathsf{Var}[d_t'(0) \mid \eta] = [\theta_1'(0) - \theta_0'(0)]^2 J(0), \tag{4.47}$$

$$\mathsf{E}[d_t''(0) \mid \eta] = [\theta_1'(0) - \theta_0'(0)][2\theta_{\eta_t}'(0) - \theta_0'(0) - \theta_1'(0)]J(0). \tag{4.48}$$

Denote $\xi_t = \varphi(Z_t, 0)$. Then ξ_t has density $f(x, 0)$ and log-density $\lambda(x, 0)$. Take expectation conditional on η on both sides of (3.9) to get

$$\mathsf{E}[d_t'(0) \mid \eta] = [\theta_1'(0) - \theta_0'(0)]\mathsf{E}\left[\frac{\partial \lambda(\xi_t, 0)}{\partial \vartheta}\right].$$

Then (4.46) follows from the property of score function.

For the same reason, (4.47) follows as well and, taking expectation conditional on η on both sides of (3.10),

$$\mathsf{E}[d_t''(0) \mid \eta] = 2[\theta_1'(0) - \theta_0'(0)]\theta_{\eta_t}'(0)\mathsf{E}\left[\frac{\partial^2 \lambda(\xi_t, 0)}{\partial x\, \partial \vartheta}\, \frac{\partial \varphi(Z_t, 0)}{\partial v}\right] - [\theta_1'(0)^2 - \theta_0'(0)^2]J(0).$$

Therefore, to prove (4.48), it suffices to show

$$\mathsf{E}\left[\frac{\partial^2 \lambda(\xi_t, 0)}{\partial x\, \partial \vartheta}\, \frac{\partial \varphi(Z_t, 0)}{\partial v}\right] = J(0). \tag{4.49}$$

Define

$$g(v, Z_t) = \frac{\partial \lambda(\varphi(Z_t, v), \vartheta)}{\partial \vartheta}\bigg|_{\vartheta=0} = \frac{1}{f(\varphi(Z_t, v), 0)}\, \frac{\partial f(\varphi(Z_t, v), 0)}{\partial \vartheta}.$$

Observe that

$$\frac{\partial g(v, Z_t)}{\partial v}\bigg|_{v=0} = \frac{\partial^2 \lambda(\xi_t, 0)}{\partial x\, \partial \vartheta}\, \frac{\partial \varphi(Z_t, 0)}{\partial v}.$$

Therefore, the left-hand side of (4.49) is equal to

$$\mathsf{E}\left[\frac{\partial g(v, Z_t)}{\partial v}\bigg|_{v=0}\right] = \frac{\partial \mathsf{E}[g(v, Z_t)]}{\partial v}\bigg|_{v=0}.$$

By assumption 1, $\varphi(Z_t, v)$ has density $f(x, v)$. Therefore, the right-hand side of the above identity is equal to

$$\frac{\partial}{\partial v}\left[\int \frac{1}{f(x, 0)}\, \frac{\partial f(x, 0)}{\partial \vartheta} f(x, v)\, dx\right]_{v=0} = \int \frac{1}{f(x, 0)}\left[\frac{\partial f(x, 0)}{\partial \vartheta}\right]^2 dx = J(0),$$

which gives (4.49).

From Theorem 2.4, (3.7) and (4.46),

$$\mathsf{E}[\mathsf{r}'(0) \mid \eta] = \sum_{t=1}^{\infty} D_{0t} \mathsf{E}[d_t'(0) \mid \eta] = 0$$

showing (3.11). On the other hand, given η, since Z_t are independent, $d_s'(0)$ are independent of $d_t'(0)$ for $s < t$. Then by $\mathsf{E}[d_t'(0) \mid \eta] = 0$ and (3.8),

$$\begin{aligned}\mathsf{E}[\mathsf{r}''(0) \mid \eta] &= \sum_{t=1}^{\infty} D_{0t} \{\mathsf{E}[d_t''(0) \mid \eta] + [P_{0t}(1,0) - P_{0t}(0,1)] \operatorname{Var}[d_t'(0) \mid \eta]\} \\ &= [\theta_1'(0) - \theta_0'(0)] J(0) \sum_{t=1}^{\infty} D_{0t} f_t,\end{aligned}$$

where

$$\begin{aligned} f_t &= 2\theta_{\eta_t}'(0) - \theta_0'(0) - \theta_1'(0) + [P_{0t}(1,0) - P_{0t}(0,1)][\theta_1'(0) - \theta_0'(0)] \\ &= [\theta_1'(0) - \theta_0'(0)][2\eta_t - P_{0t}(1,1) - P_{0t}(0,1)].\end{aligned}$$

Therefore, (3.12) holds. □

APPENDIX

In this Appendix, we make a general statement on the conditional likelihood under the FDR criterion. Let $H_1, \ldots, H_n$ be a set of hypotheses being tested and let X be the available data. Let $p_k = \Pr\{H_k \text{ is false} \mid X\}$. For any testing procedure based on X, let R be the total number of rejected H_k and V that of rejected true H_k. Then the number of rejected false nulls is $R - V$.

PROPOSITION A.1. *Given* $\alpha \in (0, 1)$, *among all testing procedures whose rejection decisions are uniquely determined by* X *and which satisfy the FDR control criterion*

$$\mathrm{FDR} = \mathsf{E}\left[\frac{V}{R \vee 1} \middle| X\right] \le \alpha,$$

the following Benjamini–Hochberg procedure [4] *has the largest* $\mathsf{E}[R - V \mid X]$:

1. *sort* $q_i = 1 - p_i$ *into* $q_{(1)} \le q_{(2)} \le \cdots \le q_{(n)}$;
2. *let* $r = \max\{j : q_{(1)} + \cdots + q_{(j)} \le \alpha j\}$ *and reject* H_k *if* $q_k \le q_{(r)}$.

PROOF. Given a procedure with $R > 0$, let H_{i_k}, $k = 1, \ldots, R$ be the rejected nulls. Then, as in [6], $\mathrm{FDR} = \sum_{j=1}^R q_{i_j}/R \ge \sum_{j=1}^R q_{(j)}/R$, while $\mathsf{E}[R - V \mid X] = R - \sum_{j=1}^R q_{i_j} \le R - \sum_{j=1}^R q_{(j)}$. It is then not hard to see that under the FDR control criterion, the procedure in the proposition attains the largest value of $E[R - V \mid X]$. □

REFERENCES

[1] ATAR, R. and ZEITOUNI, O. (1997). Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Probab. Statist.* **33** 697–725. MR1484538

[2] ATAR, R. and ZEITOUNI, O. (1997). Lyapunov exponents for finite state nonlinear filtering. *SIAM J. Control Optim.* **35** 36–55. MR1430282

[3] BAXENDALE, P., CHIGANSKY, P. and LIPTSER, R. (2004). Asymptotic stability of the Wonham filter: Ergodic and nonergodic signals. *SIAM J. Control Optim.* **43** 643–669 (electronic). MR2086177

[4] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392

[5] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245

[6] CHI, Z. (2008). False discovery rate control with multivariate p-values. *Electron. J. Stat.* **2** 368–411. MR2411440

[7] CHIGANSKY, P. and LIPTSER, R. (2004). Stability of nonlinear filters in nonmixing case. *Ann. Appl. Probab.* **14** 2038–2056. MR2099662

[8] CHIGANSKY, P. and LIPTSER, R. (2006). On a role of predictor in the filtering stability. *Electron. Comm. Probab.* **11** 129–140 (electronic). MR2240706

[9] DI MASI, G. B. and STETTNER, Ł. (2005). Ergodicity of hidden Markov models. *Math. Control Signals Systems* **17** 269–296. MR2177319

[10] DOUC, R. and MATIAS, C. (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* **7** 381–420. MR1836737

[11] DOUC, R., MOULINES, É. and RYDÉN, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.* **32** 2254–2304. MR2102510

[12] EFRON, B. (2007). Correlation and large-scale simultaneous significance testing. *J. Amer. Statist. Assoc.* **102** 93–103. MR2293302

[13] EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. G. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160. MR1946571

[14] FARCOMENI, A. (2007). Some results on the control of the false discovery rate under dependence. *Scand. J. Statist.* **34** 275–297. MR2346640

[15] FLEMING, W. H. and HERNÁNDEZ-HERNÁNDEZ, D. (1999). Risk-sensitive control of finite state machines on an infinite horizon. II. *SIAM J. Control Optim.* **37** 1048–1069 (electronic). MR1680932

[16] HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. MR2387976

[17] KAIJSER, T. (1975). A limit theorem for partially observed Markov chains. *Ann. Probab.* **3** 677–696. MR0383536

[18] KOCHMAN, F. and REEDS, J. (2006). A simple proof of Kaijser's unique ergodicity result for hidden Markov α-chains. *Ann. Appl. Probab.* **16** 1805–1815. MR2288705

[19] LE GLAND, F. and MEVEL, L. (2000). Basic properties of the projective product with application to products of column-allowable nonnegative matrices. *Math. Control Signals Systems* **13** 41–62. MR1742139

[20] LE GLAND, F. and MEVEL, L. (2000). Exponential forgetting and geometric ergodicity in hidden Markov models. *Math. Control Signals Systems* **13** 63–93. MR1742140

[21] MÜLLER, P., PARMIGIANI, G., ROBERT, C. and ROUSSEAU, J. (2004). Optimal sample size for multiple testing: The case of gene expression microarrays. *J. Amer. Statist. Assoc.* **99** 990–1001. MR2109489

[22] OWEN, A. B. (2005). Variance of the number of false discoveries. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 411–426. MR2155346

[23] QIU, X., KLEBANOV, L. and YAKOVLEV, A. (2005). Correlation between gene expression levels and limitations of the empirical Bayes methodology for finding differentially expressed genes. *Stat. Appl. Genet. Mol. Biol.* **4** Art. 34, 32 pp. (electronic). MR2183944

[24] RUDIN, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill, New York. MR0385023

[25] SARKAR, S. K. (2006). False discovery and false nondiscovery rates in single-step multiple testing procedures. *Ann. Statist.* **34** 394–415. MR2275247

[26] STOREY, J. D. (2007). The optimal discovery procedure: A new approach to simultaneous significance testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 347–368. MR2323757

[27] SUN, W. and CAI, T. T. (2009). Large scale multiple testing under dependency. *J. Roy. Statist. Soc. Ser. B* **71** 393–424.

[28] TADIĆ, V. B. and DOUCET, A. (2005). Exponential forgetting and geometric ergodicity for optimal filtering in general state-space models. *Stochastic Process. Appl.* **115** 1408–1436. MR2152381

[29] WU, W. B. (2008). On false discovery control under dependence. *Ann. Statist.* **36** 364–380. MR2387975

DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
215 GLENBROOK ROAD, U-4120
STORRS, CONNECTICUT 06269
USA
E-MAIL: zchi@stat.uocnn.edu