ALEA, Lat. Am. J. Probab. Math. Stat. 12 (1), 533–550 (2015)

A conditional strong large deviation result and a functional central limit theorem for the rate function

Anton Bovier and Hannah Mayer

Institut für Angewandte Mathematik
Rheinische Friedrich-Wilhelms-Universität Bonn
Endenicher Allee 60
53115 Bonn, Germany
E-mail address: bovier@uni-bonn.de, hannah.mayer@uni-bonn.de
URL: http://wt.iam.uni-bonn.de/~bovier, http://wt.iam.uni-bonn.de/~mayer

Abstract. We study the large deviation behaviour of $S_n = \sum_{j=1}^n W_j Z_j$, where $(W_j)_{j\in\mathbb{N}}$ and $(Z_j)_{j\in\mathbb{N}}$ are sequences of real-valued, independent and identically distributed random variables satisfying certain moment conditions, independent of each other. More precisely, we prove a conditional strong large deviation result and describe the fluctuations of the random rate function through a functional central limit theorem.

1. Introduction and Results

Let $(Z_j)_{j\in\mathbb{N}}$ be independent, identically distributed (i.i.d.) random variables and let $(W_j)_{j\in\mathbb{N}}$ be i.i.d. random variables as well. Define the σ-fields $\mathcal{Z} \equiv \sigma(Z_j, j \in \mathbb{N})$ and $\mathcal{W} \equiv \sigma(W_j, j \in \mathbb{N})$ and let $\mathcal{Z}$ and $\mathcal{W}$ be independent. Furthermore, define

$$S_n \equiv \sum_{j=1}^{n} Z_j W_j. \tag{1.1}$$

Received by the editors November 21, 2014; accepted June 23, 2015.

2010 *Mathematics Subject Classification.* 60F05, 60F10, 60F17, 60K37.

Key words and phrases. Large deviations; exact asymptotics; functional central limit theorem; local limit theorem.

We are very grateful to Frank den Hollander for numerous discussions and valuable input. A.B. is partially supported through the German Research Foundation in the Priority Programme 1590 "Probabilistic Structures in Evolution", the Hausdorff Center for Mathematics (HCM), and the Cooperative Research Centre 1060 "The Mathematics of Emergent Effects". He is member of the Cluster of Excellence "ImmunoSensation" at Bonn University. H.M. is supported by the German Research Foundation in the Bonn International Graduate School in Mathematics (BIGS) and the Cluster of Excellence "ImmunoSensation".

In this paper we derive *strong (local) large deviation* estimates on S_n *conditioned* on the σ- field $\mathcal{W}$. The random variables W_j can be interpreted as a random environment weighting the summands of S_n. Conditioning on $\mathcal{W}$ can thus be understood as fixing the environment. Comets (1989) investigates conditional large deviation estimates of such sums in the more general setup of i.i.d. random fields of random variables taking values in a Polish Space. His results concern, however, only the standard *rough* large deviation estimates. Local limit theorems have been obtained in the case $S_n \in \mathbb{R}$ (see e.g. Bahadur and Ranga Rao (1960); Chaganty and Sethuraman (1993)) and for the case $S_n \in \mathbb{R}^d$ (see Iltis (1995)), but these have, to our knowledge, not been applied to conditional laws of sums of the form (1.1).

Our result consists of two parts. The first part is an almost sure local limit theorem for the conditional tail probabilities $\mathbb{P}(S_n \geq an|\mathcal{W})$, $a \in \mathbb{R}$. The second part is a functional central limit theorem for the *random* rate function.

1.1. *Strong large deviations.* For a general review of *large deviation theory* see for example den Hollander (2000) or Dembo and Zeitouni (2010). A *large deviation principle* for a family of real-valued random variables S_n roughly says that, for $a > \mathbb{E}\left[\frac{1}{n}S_n\right]$,

$$\mathbb{P}(S_n \geq an) = \exp\left[-nI(a)(1+\mathrm{o}(1))\right]. \tag{1.2}$$

The Gärtner-Ellis theorem asserts that the *rate function*, $I(a)$, is obtained as the limit of the Fenchel-Legendre transformation of the logarithmic moment generating function of S_n, to wit $I(a) = \lim_{n\to\infty} I_n(a)$, where $I_n(a)$ is defined by

$$I_n(a) \equiv \sup_{\vartheta}(a\vartheta - \Psi_n(\vartheta)) = a\vartheta_n - \Psi_n(\vartheta_n), \tag{1.3}$$

where $\Psi_n(\vartheta) \equiv \frac{1}{n}\log \mathbb{E}[\exp(\vartheta S_n)]$ and ϑ_n satisfies $\Psi_n'(\vartheta_n) = a$. Furthermore, define $\Phi_n(\vartheta) \equiv \mathbb{E}[\exp(\vartheta S_n)]$.

Strong large deviations estimates refine this *exponential asymptotics*. They provide estimates of the form

$$\mathbb{P}(S_n \geq an) = \frac{\exp(-nI_n(a))}{\vartheta_n \sigma_n \sqrt{2\pi n}}[1+\mathrm{o}(1)], \tag{1.4}$$

where $\sigma_n^2 \equiv \Psi_n''(\vartheta_n)$ denotes the variance of $\frac{1}{\sqrt{n}}S_n$ under the *tilted* law $\widetilde{\mathbb{P}}$ that has density

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \frac{\mathrm{e}^{\vartheta_n S_n}}{\mathbb{E}\left[\mathrm{e}^{\vartheta_n S_n}\right]}. \tag{1.5}$$

The standard theorem for S_n a sum of i.i.d. random variables is due to Bahadur and Ranga Rao (1960). The generalisation, which we summarise by Theorem 1.3, is a result of Chaganty and Sethuraman (1993). We abusively refer to $I_n(a)$ as the *rate function*. The following theorem is based on 2 assumptions.

Assumption 1.1. There exist $\vartheta_* \in (0,\infty)$ and $\beta < \infty$ such that

$$|\Psi_n(\vartheta)| < \beta, \text{ for all } \vartheta \in \{\vartheta \in \mathbb{C} : |\vartheta| < \vartheta_*\} \tag{1.6}$$

for all $n \in \mathbb{N}$ large enough.

Assumption 1.2. $(a_n)_{n\in\mathbb{N}}$ is a bounded real-valued sequence such that the equation

$$a_n = \Psi_n'(\vartheta) \tag{1.7}$$

has a solution $\vartheta_n \in (0, \vartheta_{**})$ with $\vartheta_{**} \in (0, \vartheta_*)$ for all $n \in \mathbb{N}$ large enough.

Theorem 1.3 (Theorem 3.3 in Chaganty and Sethuraman (1993))**.** *Let S_n be a sequence of real-valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let Ψ_n be their logarithmic moment generating function defined above and assume that Assumptions 1.1 and 1.2 hold for Ψ_n. Assume furthermore that*

(i) $\lim_{n\to\infty} \vartheta_n \sqrt{n} = \infty$,

(ii) $\liminf_{n\to\infty} \sigma_n^2 > 0$, *and*

(iii) $\lim_{n\to\infty} \sqrt{n} \sup_{\delta_1 \le |t| \le \delta_2 \vartheta_n} \left| \frac{\Phi_n(\vartheta_n + it)}{\Phi_n(\vartheta_n)} \right| = 0 \quad \forall\, 0 < \delta_1 < \delta_2 < \infty$,

are satisfied. Then

$$\mathbb{P}\left(S_n \ge n a_n\right) = \frac{\exp(-n I_n(a_n))}{\vartheta_n \sigma_n \sqrt{2\pi n}} \left[1 + o(1)\right], \; n \to \infty. \tag{1.8}$$

This result is deduced from a local central limit theorem for $\frac{S_n - n a_n}{\sqrt{n\sigma_n^2}}$ under the tilted law $\widetilde{\mathbb{P}}$ defined in (1.5).

Remark 1.4. There are estimates for $\mathbb{P}(S_n \in n\Gamma)$, where $S_n \in \mathbb{R}^d$ and $\Gamma \subset \mathbb{R}^d$, see Iltis (1995). Then the leading order prefactor depends on d and the geometry of the set Γ.

1.2. *Application to the conditional scenario.* Throughout the following we write $I_n^{\mathcal{W}}(a)$, $\vartheta_n^{\mathcal{W}}(a)$, $\Phi_n^{\mathcal{W}}(\vartheta)$, $\Psi_n^{\mathcal{W}}(\vartheta)$ and $\mathbb{E}^{\mathcal{W}}[\cdot]$ for the random analogues of the quantities defined in the previous section, e.g. $\Phi_n^{\mathcal{W}}(\vartheta) \equiv \mathbb{E}[\exp(\vartheta S_n)|\mathcal{W}]$.

Remark 1.5. One could also condition on a different σ-field $\mathcal{Y}$ as in the application to financial mathematics and an immunological model described in Section 2. In the proofs we just need the fact that $\mathcal{W} \subset \mathcal{Y}$ and $\mathcal{Z}$ is independent of $\mathcal{Y}$.

Theorem 1.6. *Let S_n be defined in (1.1). Assume that the random variables W_1 and Z_1 satisfy the following conditions:*

(i) Z_1 is not concentrated on one point.

(a) If Z_1 is lattice valued, W_1 has an absolutely continuous part and there exists an interval $[c, d]$ such that the density of W_1 on $[c, d]$ is bounded from below by $p > 0$.

(b) If Z_1 has a density, $\mathbb{P}(|W_1| > 0) > 0$.

(ii) The moment generating function of Z_1, $M(\vartheta) \equiv \mathbb{E}[\exp(\vartheta Z_1)]$, is finite for all $\vartheta \in \mathbb{R}$.

(iii) For $f(\vartheta) \equiv \log M(\vartheta)$, both $\mathbb{E}[f(\vartheta W_1)]$ and $\mathbb{E}[W_1 f'(\vartheta W_1)]$ are finite for all $\vartheta \in \mathbb{R}$.

(iv) There exists a function $F : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[F(W_1)]$ is finite and $W_1^2 f''(\vartheta W_1) \le F(W_1)$ for all $\vartheta \in \mathbb{R}$.

Let $\vartheta_* \in \mathbb{R}_+$ be arbitrary but fixed. Let $J \equiv (\mathbb{E}[W_1]\mathbb{E}[Z_1], \mathbb{E}[W_1 f'(\vartheta_* W_1)])$ and let $a \in J$. Then

$$\mathbb{P}\left(\forall a \in J : \mathbb{P}(S_n \ge an|\mathcal{W}) = \frac{\exp(-n I_n^{\mathcal{W}}(a))}{\sqrt{2\pi n}\,\vartheta_n^{\mathcal{W}}(a)\sigma_n^{\mathcal{W}}(a)}(1 + o(1))\right) = 1, \tag{1.9}$$

where

$$I_n^{\mathcal{W}}(a) = a\vartheta_n^{\mathcal{W}}(a) - \frac{1}{n}\sum_{j=1}^{n} f\left(W_j \vartheta_n^{\mathcal{W}}(a)\right) \tag{1.10}$$

and $\vartheta_n^{\mathcal{W}}(a)$ solves $a = \frac{d}{d\vartheta}(\frac{1}{n}\sum_{j=1}^{n} f(W_j\vartheta))$.

This theorem is proven in Section 3.

Remark 1.7. The precise requirements on the distribution of W_1 depend on the distribution of Z_1. In particular, Condition (iii) does not in general require the moment generating function of W_1 to be finite for all $\vartheta \in \mathbb{R}$. Condition (iv) looks technical. It is used to establish Condition (ii) of Theorem 1.3 for all a at the same time. For most applications, it is not very restrictive, see Section 1.4 for examples.

1.3. *Functional central limit theorem for the random rate function.* Note that the rate function $I_n^{\mathcal{W}}(a)$ is *random*. Even if we may expect that $I_n^{\mathcal{W}}(a)$ converges to a deterministic function $I(a)$, almost surely, due to the fact that it is multiplied by n in the exponent in Equation (1.9), its fluctuations are relevant. To control them, we prove a functional central limit theorem. We introduce the following notation.

$$g(\vartheta) \equiv \mathbb{E}[f(W_1\vartheta)] \quad \text{and} \quad X_n(\vartheta) \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(f(W_j\vartheta) - \mathbb{E}[f(W_j\vartheta)] \right). \tag{1.11}$$

Moreover, define $\vartheta(a)$ as the solution of the equation $a = g'(\vartheta)$.

In addition to the assumptions made in Theorem 1.6, we need the following assumption on the covariance structure of the summands appearing in the definition of $X_n(\vartheta)$ and their derivatives.

Assumption 1.8. There exists $C < \infty$, such that, for all $a, a' \in \bar{J}$, where $\bar{J}$ is the closure of the interval J,

$$\begin{aligned}
&\mathrm{Cov}\left(f(\vartheta(a)W_j), f(\vartheta(a')W_j)\right), \quad \mathrm{Cov}\left(W_j f'(\vartheta(a)W_j), W_j f'(\vartheta(a')W_j)\right),\\
&\mathrm{Cov}\left(W_j^2 f''(\vartheta(a)W_j), W_j^2 f''(\vartheta(a')W_j)\right), \quad \mathrm{Cov}\left(f(\vartheta(a)W_j), W_j f'(\vartheta(a')W_j)\right),\\
&\mathrm{Cov}\left(W_j f'(\vartheta(a)W_j), W_j^2 f''(\vartheta(a')W_j)\right), \quad \mathrm{Cov}\left(f(\vartheta(a)W_j), W_j^2 f''(\vartheta(a')W_j)\right) \text{ and}\\
&\mathbb{V}\left[W_j^3 f'''(\vartheta(a)W_j)\right]
\end{aligned}$$

are all smaller than C.

Theorem 1.9. *If $g''(\vartheta(a)) > c$ for some $c > 0$ and Assumption 1.8 is satisfied, then the rate function satisfies*

$$I_n^{\mathcal{W}}(a) = I(a) + n^{-1/2} X_n(\vartheta(a)) + n^{-1} r_n(a), \tag{1.12}$$

where

$$I(a) \equiv a\vartheta(a) - g(\vartheta(a)), \tag{1.13}$$

$$(X_n(\vartheta(a)))_{a\in\bar{J}} \xrightarrow{\mathcal{D}} (X_a)_{a\in\bar{J}}, \quad \text{as}\, n \to \infty, \tag{1.14}$$

where X is the Gaussian process with mean zero and covariance

$$\mathrm{Cov}(X_a, X_{a'}) = \mathbb{E}[f(W_1\vartheta(a)) f(W_1\vartheta(a'))] - \mathbb{E}[f(W_1\vartheta(a))]\mathbb{E}[f(W_1\vartheta(a'))], \tag{1.15}$$

and

$$r_n(a) = \frac{(X_n'(\vartheta(a)))^2}{2\left[g''(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n''(\vartheta(a))\right]} + o(1), \tag{1.16}$$

uniformly in $a \in \bar{J}$.

To prove Theorem 1.9 we show actually more, namely that the process

$$(X_n(\vartheta(a)), X_n'(\vartheta(a)), X_n''(\vartheta(a)))_{a\in\bar{J}} \xrightarrow{\mathcal{D}} (X_a, X_a', X_a'')_{a\in\bar{J}}, \tag{1.17}$$

(see Lemma 4.1 below). The proof of the theorem is given in Section 4.

1.4. *Examples.* In the following we list some examples in which the conditions of the preceding theorems are satisfied.

(1) Let Z_1 be a Gaussian random variable with mean zero and variance σ^2. In this case,

$$\begin{aligned} f(\vartheta) &= \log(\mathbb{E}[\exp(\vartheta Z_1)]) = \frac{1}{2}\sigma^2\vartheta^2, \quad f'(\vartheta) = \sigma^2\vartheta & (1.18) \\ f''(\vartheta) &= \sigma^2, \quad \text{and} \quad f'''(\vartheta) = 0 & (1.19) \end{aligned}$$

This implies that W_1 must have finite fourth moments to satisfy Assumption 1.8. Under this requirement Conditions (iii) and (iv) of Theorem 1.6 are met. According to Condition (ib) of Theorem 1.6, W_1 may not be concentrated at 0. Moreover,

$$g''(\vartheta) = \sigma^2 > c \tag{1.20}$$

independent of the distribution of W_1.

(2) Let Z_1 be a binomially distributed random variable, $Z_1 \sim B(m, p)$. Thus

$$\begin{aligned} f(\vartheta) &= m\log(1 - p + p\mathrm{e}^{\vartheta}) & (1.21) \\ f'(\vartheta) &= m\frac{p\mathrm{e}^{\vartheta}}{1 - p + p\mathrm{e}^{\vartheta}} \le m & (1.22) \\ f''(\vartheta) &= m(p - p^2)\frac{\mathrm{e}^{\vartheta}}{(1 - p + p\mathrm{e}^{\vartheta})^2} \le f''\left(\log\left(\frac{3p - 1}{p}\right)\right) & (1.23) \\ f'''(\vartheta) &= m(p - p^2)\mathrm{e}^{\vartheta}\frac{1 - 3p + p\mathrm{e}^{\vartheta}}{(1 - p + p\mathrm{e}^{\vartheta})^3} \in C_0. & (1.24) \end{aligned}$$

Then W_1 has to satisfy (ia) of Theorem 1.6 and must have finite sixth moments. One can show that $f'(\vartheta), f''(\vartheta)$ and $f'''(\vartheta)$ are bounded, $\mathbb{E}[f(\vartheta W_1)]$ and the moments depending on $f(\vartheta W_1)$ in Assumption 1.8 are finite. Furthermore, the assumption $0 < \mathbb{E}[W_1^2] < \infty$ implies that $g(\vartheta(a)) > c$ as required in Theorem 1.9.

Remark 1.10. In both cases it is not necessary that the moment generating function of W_1 exists.

1.5. *Related results.* After posting our manuscript on arXiv, Ioannis Kontoyiannis informed us about the papers Dembo and Kontoyiannis (1999) and Dembo and Kontoyiannis (2002), where some similar results on conditional large deviations are obtained. They concern sums of the form

$$\rho_n \equiv \frac{1}{n}\sum_{j=1}^{n}\rho(W_j, Z_j), \tag{1.25}$$

where $\mathbf{W} = (W_j)_{j\in\mathbb{N}}$ and $\mathbf{Z} = (Z_j)_{j\in\mathbb{N}}$ are two stationary processes with W_j and Z_j taking values in Polish spaces A_W and A_Z, respectively, and $\rho : A_W \times A_Z \to [0, \infty)$ is some measurable function. Their main motivation is to estimate the frequency with which subsequences of length n in the process Z occur that are "close" to W. To do this, they estimate conditional probabilities of the form

$$\mathbb{P}\left(\rho_n \le D|\mathcal{W}\right), \tag{1.26}$$

obtaining, under suitable assumptions, refined large deviation estimates of the form

$$\frac{1}{n}\log \mathbb{P}\left(\rho_n \leq D|\mathcal{W}\right) = R_n(D) + \frac{1}{\sqrt{n}}\Lambda_n(D) + o(1/\sqrt{n}), \tag{1.27}$$

almost surely, where they show that $R_n(D)$ converges a.s. while $\Lambda_n(D)$ converges in distribution to a Gaussian random variable.

2. Applications

2.1. *Stochastic model of T-cell activation.* The immune system defends the body against dangerous intrusion, e.g. bacteria, viruses and cancer cells. The interaction of so-called T-cells and antigen presenting cells plays an important rôle in performing this task. Van den Berg, Rand and Burroughs developed a stochastic model of T-cell activation in van den Berg et al. (2001) which was further investigated in Zint et al. (2008) and Mayer and Bovier (2015). Let us briefly explain this model.

The antigen presenting cells display on their surface a mixture of peptides present in the body. During a bond between a T-cell and a presenting cell the T-cell scans the presented mixture of peptides. The T-cell is stimulated during this process, and if the sum of all stimuli exceeds a threshold value, the cell becomes activated and triggers an immune response. The signal received by the T-cell is represented by

$$S_n \equiv \sum_{j=1}^{n} Z_j W_j + z_{\mathsf{f}} W_{\mathsf{f}}, \tag{2.1}$$

where W_j represents the stimulation rate elicited by a peptide of type j and Z_j represents the random number of presented peptides of type j. The sum describes the signal due to self peptides, $z_{\mathsf{f}} W_{\mathsf{f}}$ is the signal due to one foreign peptide type. From the biological point of view, T-cell activations are rare events and thus large deviation theory is called for to investigate $\mathbb{P}(S_n \geq na|\mathcal{Y})$, where $\mathcal{Y}$ is a σ-field such that W_j are measurable with respect to $\mathcal{Y}$ and Z_j are independent of $\mathcal{Y}$. For two examples of distributions discussed in Zint et al. (2008), Theorems 1.6 and 1.9 can be applied. In both examples, the random variables Z_j are binomially distributed, and thus their moment generating function exists everywhere. W_j is defined by $W_j \equiv \frac{1}{\tau_j}\exp(-\frac{1}{\tau_j})$, where τ_j are exponentially distributed or logarithmic normally distributed, i.e. W_j are bounded and the required moments exist. Furthermore, W_1 has a density and Condition (ia) of Theorem 1.6 is met. Using Theorems 1.6 and 1.9, one can prove that the probability of T-cell activation for a given type of T-cell grows exponentially with the number of presented foreign peptides, z_{f}, if the corresponding stimulation rate W_{f} is sufficiently large. It is then argued that a suitable activation threshold can be set that allows significantly differentiate between the presence or absence of foreign peptides. For more details see Mayer and Bovier (2015).

2.2. *Large portfolio losses.* Dembo, Deuschel, and Duffie investigate in Dembo et al. (2004) the probability of large financial losses on a bank portfolio or the total claims against an insurer conditioned on a macro environment. The random variable S_n represents the total loss on a portfolio consisting of many positions, W_j is a $\{0, 1\}$-valued random variable and indicates if position j experiences a loss, whereas the random variable Z_j is for example exponentially distributed and represents the amount of loss. They consider the probability conditioned on a common macro environment $\mathcal{Y}$ and assume that $Z_1, W_1, \ldots, Z_n, W_n$ are conditionally independent. Furthermore, they work in the slightly generalised setup of finitely many blocks of different distributions. That is

$$S_n \equiv \sum_{\alpha=1}^{K} \sum_{j=1}^{Q_\alpha} Z_{\alpha,j} W_{\alpha,j}, \tag{2.2}$$

where $Z_{\alpha,j} \overset{\mathcal{D}}{=} Z_\alpha$ and $W_{\alpha,j} \overset{\mathcal{D}}{=} W_\alpha$ for each $\alpha \in \{1, \ldots, K\}$ and $\sum_{\alpha=1}^{K} Q_\alpha = n$. Moreover, the conditional probability of losses for each position is calculated and the influence of the length of the time interval, in which the loss occurs, is investigated. For more details see Dembo et al. (2004). This analysis was generalised later in a paper by Spiliopoulos and Sowers (2011).

Remark 2.1. In general, the exponential distribution for Z_1 causes problems because the moment generating function does not exist everywhere. Evaluating at ϑW_j thus might yield to an infinite term depending on the range of W_j. In this application there is no problem because W_j is $\{0, 1\}$-valued.

3. Proof of Theorem 1.6

Proof of Theorem 1.6: We prove Theorem 1.6 by showing that the conditional law of S_n given $\mathcal{W}$ satisfies the assumptions of Theorem 1.3 uniformly in $a \in J$, almost surely.

Assumption 1.1 is satisfied due to Conditions (ii) and (iii) of Theorem 1.6: For each $n \in \mathbb{N}$ and each realisation of $(W_j)_{j\in\mathbb{N}}$ $\Psi_n^{\mathcal{W}}(\vartheta)$ is a convex function. Furthermore,

$$\Psi_n^{\mathcal{W}}(\vartheta) \leq \max\{\Psi_n^{\mathcal{W}}(\vartheta_*), \Psi_n^{\mathcal{W}}(-\vartheta_*)\} \tag{3.1}$$

and

$$\lim_{n\to\infty} \max\{\Psi_n^{\mathcal{W}}(\vartheta_*), \Psi_n^{\mathcal{W}}(-\vartheta_*)\} = \max\{\mathbb{E}\left[f(W_1\vartheta_*)\right], \mathbb{E}\left[f(-W_1\vartheta_*))\right]\}, \quad \text{a.s.} \tag{3.2}$$

This implies that Assumption 1.1 is satisfied. To prove that Assumption 1.2 holds, note that, by the law of large numbers,

$$\lim_{n\to\infty} \frac{d}{d\vartheta} \Psi_n^{\mathcal{W}}(0) = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}^{\mathcal{W}}[S_n] = \mathbb{E}[W_1]\mathbb{E}[Z_1], \quad \text{a.s.} \tag{3.3}$$

Next, by convexity, and again the law of large numbers

$$\liminf_{n\to\infty} \sup_{\vartheta\in[0,\vartheta_*]} \frac{d}{d\vartheta} \Psi_n^{\mathcal{W}}(\vartheta) = \liminf_{n\to\infty} \frac{d}{d\vartheta} \Psi_n^{\mathcal{W}}(\vartheta_*) = \mathbb{E}[W_1 f'(\vartheta_* W_1)], \quad \text{a.s.} \tag{3.4}$$

Recall that $\vartheta_n^{\mathcal{W}}(a)$ is defined as the solution of

$$a = \frac{1}{n} \sum_{j=1}^{n} \frac{d}{d\vartheta} \log M(W_j \vartheta) = \frac{1}{n} \sum_{j=1}^{n} W_j f'(\vartheta W_j). \tag{3.5}$$

For n large enough, the solution $\vartheta_n^{\mathcal{W}}(a)$ exists for $a \in J$ and is unique since the logarithmic moment generating function $\Psi_n^{\mathcal{W}}$ is strictly convex. Again by monotonicity of $\frac{d}{d\vartheta}\Psi_n^{\mathcal{W}}(\vartheta)$ in ϑ, and because of (3.3) and (3.4), for $a \in J$, $\vartheta_n^{\mathcal{W}}(a) \in (0, \vartheta_*)$, almost surely, for n large enough. Thus Assumption 1.2 is satisfied.

In order to establish Condition (i) of Theorem 1.3 we prove the following

Lemma 3.1. $\mathbb{P}\big(\forall a \in J : \lim_{n\to\infty} \vartheta_n^{\mathcal{W}}(a) = \vartheta(a)\big) = 1.$

Proof: First, using that $g'(\vartheta)$ is continuous and monotone increasing

$$\begin{aligned}
&\mathbb{P}\left(\forall a \in J : \lim_{n\to\infty} |\vartheta_n^{\mathcal{W}}(a) - \vartheta(a)| = 0\right) \\
&= \mathbb{P}\Bigg(\forall a \in J : \lim_{n\to\infty} \Bigg| g'(\vartheta_n^{\mathcal{W}}(a)) - \frac{1}{n}\sum_{j=1}^{n} W_j f'(\vartheta_n^{\mathcal{W}}(a)W_j) \\
&\qquad\qquad -g'(\vartheta(a)) + \frac{1}{n}\sum_{j=1}^{n} W_j f'(\vartheta_n^{\mathcal{W}}(a)W_j)\Bigg| = 0\Bigg) \\
&= \mathbb{P}\left(\forall a \in J : \lim_{n\to\infty} \left| g'(\vartheta_n^{\mathcal{W}}(a)) - \frac{1}{n}\sum_{j=1}^{n} W_j f'(\vartheta_n^{\mathcal{W}}(a)W_j)\right| = 0\right),
\end{aligned} \tag{3.6}$$

where we used that, by definition of $\vartheta(a)$ and $\vartheta_n(a)$,

$$\frac{1}{n}\sum_{j=1}^{n} W_j f'(\vartheta_n^{\mathcal{W}}(a)W_j) = g'(\vartheta(a)) = a. \tag{3.7}$$

Since we have seen that for $a \in J$, $\vartheta(a) \in [0, \vartheta_*]$ and, for n large enough, $\vartheta_n^{\mathcal{W}}(a) \in [0, \vartheta_*]$, the last line in (3.6) is bounded from below by

$$\mathbb{P}\left(\sup_{\vartheta\in[0,\vartheta_*]} \lim_{n\to\infty} \left| \frac{1}{n}\sum_{j=1}^{n} W_j f'(\vartheta W_j) - \mathbb{E}[W_1 f'(\vartheta W_1)] \right| = 0\right). \tag{3.8}$$

Denote the open ball of radius δ around ϑ by $B_\delta(\vartheta) \equiv \{\bar\vartheta \in \mathbb{R} : |\vartheta - \bar\vartheta| < \delta\}$. The following facts are true:

(1) By Condition (iii) of Theorem 1.6 $W_1 f'(\vartheta W_1)$ is integrable, for each $\vartheta \in [0, \vartheta_*]$.
(2) $W_1(\omega) f'(\vartheta W_1(\omega))$ is a continuous function of ϑ, $\forall \omega \in \Omega$.
(3) $W_1 f'(\vartheta W_1)$ is monotone increasing in ϑ since $\frac{d}{d\vartheta}(W_1 f'(\vartheta W_1)) > 0$.
(4) (1), (2), and (3) imply, by dominated convergence, that, for all $\vartheta \in [0, \vartheta_*]$,

$$\lim_{\delta\downarrow 0} \mathbb{E}\left[\sup_{\bar\vartheta\in B_\delta(\vartheta)} W_1 f'(\bar\vartheta W_1) - \inf_{\bar\vartheta\in B_\delta(\vartheta)} W_1 f'(\bar\vartheta W_1)\right] = 0. \tag{3.9}$$

Note that (4) implies that, for all $\vartheta \in [0, \vartheta_*]$ and for all $\varepsilon > 0$, there exists a $\delta = \delta(\varepsilon, \vartheta)$, such that

$$\left|\mathbb{E}\left[\sup_{\bar\vartheta\in B_{\delta(\varepsilon,\vartheta)}(\vartheta)} W_1 f'(\bar\vartheta W_1) - \inf_{\bar\vartheta\in B_{\delta(\varepsilon,\vartheta)}(\vartheta)} W_1 f'(\bar\vartheta W_1)\right]\right| < \varepsilon. \tag{3.10}$$

The collection $\{B_{\delta(\varepsilon,\vartheta)}(\vartheta)\}_{\vartheta\in[0,\vartheta_*]}$ is an open cover of $[0,\vartheta_*]$, and since $[0,\vartheta_*]$ is compact we can choose a finite subcover, $\{B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)\}_{1\le k\le K}$. Therefore

$$\begin{aligned}
&\sup_{\vartheta\in[0,\vartheta_*]}\left\{\left|\frac{1}{n}\sum_{j=1}^{n}W_jf'(\vartheta W_j)-\mathbb{E}[W_1f'(\vartheta W_1)]\right|\right\} \qquad (3.11)\\
&=\max_{1\le k\le K}\sup_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}\left\{\left|\frac{1}{n}\sum_{j=1}^{n}W_jf'(\vartheta W_j)-\mathbb{E}[W_1f'(\vartheta W_1)]\right|\right\}\\
&=\max_{1\le k\le K}\max\left\{\left|\sup_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}\left(\frac{1}{n}\sum_{j=1}^{n}W_jf'(\vartheta W_j)-\mathbb{E}[W_1f'(\vartheta W_1)]\right)\right|,\right.\\
&\qquad\left.\left|\inf_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}\left(\frac{1}{n}\sum_{j=1}^{n}W_jf'(\vartheta W_j)-\mathbb{E}[W_1f'(\vartheta W_1)]\right)\right|\right\}.
\end{aligned}$$

It suffices to show that for all $1\le k\le K$ and n large enough almost surely

$$\begin{aligned}
-\varepsilon\quad &< \inf_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}\left(\frac{1}{n}\sum_{j=1}^{n}W_jf'(\vartheta W_j)-\mathbb{E}[W_1f'(\vartheta W_1)]\right)\\
&\le \sup_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}\left(\frac{1}{n}\sum_{j=1}^{n}W_jf'(\vartheta W_j)-\mathbb{E}[W_1f'(\vartheta W_1)]\right)<\varepsilon. \qquad (3.12)
\end{aligned}$$

Note that

$$\begin{aligned}
&\sup_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}\left(\frac{1}{n}\sum_{j=1}^{n}W_jf'(\vartheta W_j)-\mathbb{E}[W_1f'(\vartheta W_1)]\right) \qquad (3.13)\\
&\le\frac{1}{n}\sum_{j=1}^{n}\sup_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}W_jf'(\vartheta W_j)-\inf_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}\mathbb{E}[W_1f'(\vartheta W_1)]
\end{aligned}$$

Since by convexity of f

$$W_jf'(-\vartheta_*W_j)\le\sup_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}W_jf'(\vartheta W_j)\le W_jf'(\vartheta_*W_j) \qquad (3.14)$$

and since these bounds are integrable by Condition (iii) of Theorem 1.6 also the supremum itself is integrable. Thus, the strong law of large numbers applies and (3.13) converges almost surely to

$$\mathbb{E}\left[\sup_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}W_1f'(\vartheta W_1)\right]-\inf_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}\mathbb{E}[W_1f'(\vartheta W_1)], \qquad (3.15)$$

which in turn, due to (3.10), is bounded from above by

$$\mathbb{E}\left[\sup_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}W_1f'(\vartheta W_1)-\inf_{\vartheta\in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)}W_1f'(\vartheta W_1)\right]<\varepsilon. \qquad (3.16)$$

With a similar argument it can be shown that for all $1 \le k \le K$ and n large enough almost surely

$$\inf_{\vartheta \in B_{\delta(\varepsilon,\vartheta_k)}(\vartheta_k)} \left(\frac{1}{n} \sum_{j=1}^{n} W_j f'(\vartheta W_j) - \mathbb{E}[W_1 f'(\vartheta W_1)] \right) > -\varepsilon. \tag{3.17}$$

Thus, $\vartheta_n^{\mathcal{W}}(a)$ converges almost surely to $\vartheta(a)$. □

But for $a \in J$, we know that $\vartheta(a) > 0$, and since $\vartheta_n^{\mathcal{W}}(a)$ converges to $\vartheta(a)$, a.s., a fortiori, Condition (i) of Theorem 1.3 is satisfied, a.s.

Next we show that Condition (ii) of Theorem 1.3 is also satisfied, almost surely. To see this, write

$$\begin{aligned}
&\left(\frac{d^2}{d\vartheta^2}\Psi_n^{\mathcal{W}}(\vartheta)\right)\Bigg|_{\vartheta=\vartheta_n^{\mathcal{W}}(a)} \\
&= \frac{1}{n}\sum_{j=1}^{n} \frac{\mathbb{E}[W_j^2 Z_j^2 e^{\vartheta_n^{\mathcal{W}}(a) W_j Z_j}|\mathcal{W}]\mathbb{E}[e^{\vartheta_n^{\mathcal{W}}(a) W_j Z_j}|\mathcal{W}] - \left(\mathbb{E}[W_j Z_j e^{\vartheta_n^{\mathcal{W}}(a) W_j Z_j}|\mathcal{W}]\right)^2}{(\mathbb{E}[e^{\vartheta_n^{\mathcal{W}}(a) W_j Z_j}|\mathcal{W}])^2} \\
&= \frac{1}{n}\sum_{j=1}^{n} \mathbb{V}^{\vartheta_n^{\mathcal{W}}(a)}[W_j Z_1|\mathcal{W}].
\end{aligned} \tag{3.18}$$

The conditional variance $\mathbb{V}^{\vartheta_n^{\mathcal{W}}(a)}[W_j Z_j|\mathcal{W}]$ is clearly positive with positive probability, since we assumed the distribution of Z_1 to be non-degenerate and W_j is non-zero with positive probability. We need to show that also the infimum over $n \in \mathbb{N}$ is strictly positive. Note that

$$\Psi''(\vartheta(a)) = \mathbb{E}[\mathbb{V}^{\vartheta(a)}[W_1 Z_1|\mathcal{W}]] > 0. \tag{3.19}$$

We need the following lemma.

Lemma 3.2. $\mathbb{P}\left(\forall a \in J : \lim_{n\to\infty} \Psi_n''(\vartheta_n^{\mathcal{W}}(a)) = \Psi''(\vartheta(a))\right) = 1.$

Proof: Since trivially

$$\begin{aligned}
&|\Psi_n''(\vartheta_n^{\mathcal{W}}(a)) - \Psi''(\vartheta(a))| \\
&\le |\Psi_n''(\vartheta_n^{\mathcal{W}}(a)) - \Psi''(\vartheta_n^{\mathcal{W}}(a))| + |\Psi''(\vartheta_n^{\mathcal{W}}(a)) - \Psi''(\vartheta(a))|,
\end{aligned} \tag{3.20}$$

Lemma 3.2 follows if both

$$\mathbb{P}\left(\forall a \in J : \lim_{n\to\infty} |\Psi_n''(\vartheta_n^{\mathcal{W}}(a)) - \Psi''(\vartheta_n^{\mathcal{W}}(a))| = 0\right) = 1 \tag{3.21}$$

and

$$\mathbb{P}\left(\forall a \in J : \lim_{n\to\infty} |\Psi''(\vartheta_n^{\mathcal{W}}(a)) - \Psi''(\vartheta(a))| = 0\right) = 1. \tag{3.22}$$

Now, $\Psi''(\vartheta)$ is a continuous function of ϑ and uniformly continuous on the compact interval $[0, \vartheta_*]$. This implies that

$$\forall \varepsilon > 0 \, \exists \delta = \delta(\varepsilon) : \forall \vartheta, \vartheta' : |\vartheta - \vartheta'| < \delta \quad |\Psi''(\vartheta) - \Psi''(\vartheta')| < \varepsilon. \tag{3.23}$$

From the uniform almost sure convergence of $\vartheta_n^{\mathcal{W}}(a)$ to $\vartheta(a)$, it follows that

$$\forall \delta > 0 \, \exists N = N(\omega, \delta) : |\vartheta_n^{\mathcal{W}}(a) - \vartheta(a)| < \delta, \tag{3.24}$$

which in turn implies that

$$\forall n \ge N : |\Psi''(\vartheta_n^{\mathcal{W}}(a)) - \Psi''(\vartheta(a))| < \varepsilon. \tag{3.25}$$

Therefore, Equation (3.22) holds. The proof of (3.21) is very similar to that of Lemma 3.1. The difference is that we cannot use monotonicity to obtain a majorant for $W_1^2 f''(\vartheta W_1)$, but instead use Condition (iv) of Theorem 1.6. Again, as in (3.8),

$$\begin{aligned}
&\mathbb{P}\left(\forall a \in J : \lim_{n\to\infty} |\Psi_n''(\vartheta_n^{\mathcal{W}}(a)) - \Psi''(\vartheta_n^{\mathcal{W}}(a))| = 0\right) \\
&\geq \mathbb{P}\left(\sup_{\vartheta\in[0,\vartheta_*]} \lim_{n\to\infty} |\Psi_n''(\vartheta) - \Psi''(\vartheta)| = 0\right). \qquad (3.26)
\end{aligned}$$

Moreover, the following facts are true:

(1) By Condition (iv) of Theorem 1.6 and the convexity of f, $0 \leq W_1^2 f''(\vartheta W_1) \leq F(W_1)$ and $\mathbb{E}[F(W_1)] < \infty$.
(2) $W_1^2(\omega) f''(\vartheta W_1(\omega))$ is a continuous function of ϑ $\forall \omega \in \Omega$.
(3) From (1) and (2) it follows by dominated convergence that for all $\vartheta \in [0, \vartheta_*]$ that

$$\lim_{\delta\downarrow 0} \mathbb{E}\left[\sup_{\bar\vartheta\in B_\delta(\vartheta)} W_1^2 f''(\bar\vartheta W_1) - \inf_{\bar\vartheta\in B_\delta(\vartheta)} W_1^2 f''(\bar\vartheta W_1)\right] = 0. \qquad (3.27)$$

The proof of Lemma 3.2 proceeds from here exactly as the proof of Lemma 3.1, just replacing f' by f'' and W_1 by W_1^2. □

Condition (ii) of Theorem 1.3 now follows immediately.

Next we show that Condition (iii) is satisfied. We want to show that $\forall 0 < \delta_1 < \delta_2 < \infty$

$$\mathbb{P}\left(\forall a \in J : \lim_{n\to\infty} \sqrt{n} \sup_{\delta_1 \leq |t| \leq \delta_2 \vartheta_n^{\mathcal{W}}(a)} \left|\frac{\Phi_n^{\mathcal{W}}(\vartheta_n^{\mathcal{W}}(a) + it)}{\Phi_n^{\mathcal{W}}(\vartheta_n^{\mathcal{W}}(a))}\right| = 0\right) = 1. \qquad (3.28)$$

As above we bound the probability in (3.28) from below by

$$\mathbb{P}\left(\lim_{n\to\infty} \sqrt{n} \sup_{\vartheta\in[0,\vartheta_*]} \sup_{\delta_1\leq|t|\leq\delta_2\vartheta} \left|\frac{\Phi_n^{\mathcal{W}}(\vartheta + it)}{\Phi_n^{\mathcal{W}}(\vartheta)}\right| = 0\right). \qquad (3.29)$$

Therefore, (3.28) follows from the first Borel-Cantelli lemma if, for each $\delta > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\sqrt{n} \sup_{\vartheta\in[0,\vartheta_*]} \sup_{\delta_1\leq|t|\leq\delta_2\vartheta} \left|\frac{\Phi_n^{\mathcal{W}}(\vartheta + it)}{\Phi_n^{\mathcal{W}}(\vartheta)}\right| > \delta\right) < \infty. \qquad (3.30)$$

Note that

$$\left|\frac{\Phi_n^{\mathcal{W}}(\vartheta + it)}{\Phi_n^{\mathcal{W}}(\vartheta)}\right| = \prod_{j=1}^{n} \left|\frac{M(W_j(\vartheta + it))}{M(W_j\vartheta)}\right| \qquad (3.31)$$

is a product of functions with absolute value less or equal to 1. Each factor is the characteristic function of a tilted Z_j. According to a result of Feller (1971) there are 3 classes of characteristic functions.

Lemma 3.3 (Lemma 4 in Chapter XV in Feller (1971))**.** *Let ϕ be the characteristic function of a probability distribution function F. Then one of the following must hold:*

(1) $|\phi(\zeta)| < 1$ *for all* $\zeta \neq 0$.
(2) $|\phi(\lambda)| = 1$ *and* $|\phi(\zeta)| < 1$ *for* $0 < \zeta < \lambda$. *In this case* ϕ *has period* λ *and there exists a real number* b *such that* $F(x + b)$ *is arithmetic with span* $h = {}^{2\pi}/_{\lambda}$.

(3) $|\phi(\zeta)| = 1$ *for all* ζ*. In this case* $\phi(\zeta) = \mathrm{e}^{\mathrm{i}b\zeta}$ *and* F *is concentrated at the point* b*.*

Case (3) is excluded by assumption. Under Condition (ia) of Theorem 1.6 we are in Case (1). In this case it is rather easy to verify Equation (3.28). Namely, observe that there exists $0 < \rho < 1$, such that for all $\vartheta \in [0, \vartheta_*]$, for all $\delta_1 \le t \le \delta_2 \vartheta_*$, whenever $K^{-1} \le |W_j| \le K$, for some $0 < K < \infty$,

$$\left| \frac{M(W_j(\vartheta + it))}{M(W_j\vartheta)} \right| < 1 - \rho. \tag{3.32}$$

This implies that, for ϑ as specified,

$$\left| \frac{M(W_j(\vartheta + it))}{M(W_j\vartheta)} \right| \le (1-\rho)^{\mathbb{1}_{\{\frac{1}{K} \le |W_j| \le K\}}}. \tag{3.33}$$

Therefore,

$$\begin{aligned} &\mathbb{P}\left(\sqrt{n} \sup_{\vartheta \in [0,\vartheta_*]} \sup_{\delta_1 \le |t| \le \delta_2 \vartheta} \prod_{j=1}^{n} \left| \frac{M(W_j(\vartheta + it))}{M(W_j\vartheta)} \right| > \delta \right) \\ &\le \mathbb{P}\left(\sqrt{n}(1-\rho)^{\sum_{j=1}^{n} \mathbb{1}_{\{\frac{1}{K} \le |W_j| \le K\}}} > \delta\right), \end{aligned} \tag{3.34}$$

where K is chosen such that $\mathbb{P}\left(\frac{1}{K} \le |W_j| \le K\right) > 0$. With $c_n \equiv \frac{\log \delta - \frac{1}{2}\log n}{\log(1-\rho)}$, the probability in the second line of (3.34) is equal to

$$\begin{aligned} &\mathbb{P}\left(\sum_{j=1}^{n} \mathbb{1}_{\{\frac{1}{K} \le |W_j| \le K\}} < c_n\right) \\ &\le \sum_{k=1}^{\lceil c_n \rceil} \binom{n}{k} \mathbb{P}\left(\frac{1}{K} \le |W_j| \le K\right)^k \left[1 - \mathbb{P}\left(\frac{1}{K} \le |W_j| \le K\right)\right]^{n-k} \\ &\le \lceil c_n \rceil \binom{n}{\lceil c_n \rceil} \left[1 - \mathbb{P}\left(\frac{1}{K} \le |W_j| \le K\right)\right]^{n - \lfloor c_n \rfloor}. \end{aligned} \tag{3.35}$$

Since $\binom{n}{\lceil c_n \rceil} \sim n^{C \log n}$ for a constant C, this is summable in n and (3.30) holds.

Case (2) of lattice-valued random variables Z_j, which corresponds to Condition (ib) of Theorem 1.6, is more subtle. Each of the factors in the product in (3.31) is a periodic function, which is equal to 1 if and only if $W_j t \in \{k\lambda, k \in \mathbb{Z}\}$, where λ is the period of this function. This implies that each factor is smaller than 1 if $W_j \notin \{k\lambda/t, k \in \mathbb{Z}\}$. The points of this set do not depend on ϑ and have the smallest distance to each other if t is maximal, i.e. $t = \delta_2 \vartheta_*$. Each factor is *strictly* smaller than 1 if tW_j does not lie in a finite interval around one of these points. We choose these intervals as follows. Let

$$\widetilde{\delta} \equiv \min\left\{\frac{\lambda}{8\delta_2\vartheta_*}, \frac{d-c}{4}\right\} \tag{3.36}$$

and define the intervals

$$I(k, t, \widetilde{\delta}) \equiv \left[\frac{k\lambda}{t} - \widetilde{\delta}, \frac{k\lambda}{t} + \widetilde{\delta}\right]. \tag{3.37}$$

These disjoint and consecutive intervals are separated by a distance at least $6\widetilde{\delta}$ from each other. Then, for all $\vartheta \in [0, \vartheta_*]$ there exists $0 < \rho(\vartheta) < 1$, independent of t, such that

$$\left| \frac{M(W_j(\vartheta + it))}{M(W_j\vartheta)} \right| \leq (1 - \rho(\vartheta))^{\mathbb{1}_{\{|W_j| \notin \cup_{k\in\mathbb{Z}} I(k,t,\widetilde{\delta})\}}} . \tag{3.38}$$

Furthermore, $\left|\frac{M(\theta+it)}{M(\theta)}\right|$ is continuous in θ, and thus its supremum over compact intervals is attained. Thus, for any $C > 0$ there exists $\bar{\rho} = \bar{\rho}(C, \vartheta_*) > 0$ such that, for all $\vartheta \in [0, \vartheta_*]$,

$$\left| \frac{M(W_j(\vartheta + it))}{M(W_j\vartheta)} \right| \leq (1 - \bar{\rho})^{\mathbb{1}_{\{W_j \in [-C,C] \setminus \cup_{k\in\mathbb{Z}} I(k,t,\widetilde{\delta})\}}} . \tag{3.39}$$

We choose C such that the interval $[c, d]$ from Hypothesis (ia) is contained in $[-C, C]$. Then we get with Equations (3.38) and (3.39) that

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{n} \sup_{\vartheta\in[0,\vartheta_*]} \sup_{\delta_1 \leq |t| \leq \delta_2\vartheta} \prod_{j=1}^{n} \left| \frac{M(W_j(\vartheta + it))}{M(W_j\vartheta)} \right| > \delta \right) \\
&\leq \mathbb{P}\left(\sqrt{n} \sup_{\vartheta\in[0,\vartheta_*]} \sup_{\delta_1 \leq |t| \leq \delta_2\vartheta_*} \prod_{j=1}^{n} (1 - \bar{\rho})^{\mathbb{1}_{\{W_j \in [-C,C] \setminus \cup_{k\in\mathbb{Z}} I(k,t,\widetilde{\delta})\}}} > \delta \right) \\
&= \mathbb{P}\left(\sqrt{n}\, (1 - \bar{\rho})^{\inf_{\delta_1 \leq |t| \leq \delta_2\vartheta_*} \sum_{j=1}^{n} \mathbb{1}_{\{W_j \in [-C,C] \setminus \cup_{k\in\mathbb{Z}} I(k,t,\widetilde{\delta})\}}} > \delta \right) .
\end{aligned} \tag{3.40}$$

With $c_n \equiv \frac{\log\delta - \frac{1}{2}\log n}{\log(1-\bar{\rho})}$ Equation (3.40) can be rewritten as

$$\begin{aligned}
&\mathbb{P}\Big(\inf_{\delta_1 \leq |t| \leq \delta_2\vartheta_*} \sum_{j=1}^{n} \mathbb{1}_{\{W_j \in [-C,C] \setminus \cup_{k\in\mathbb{Z}} I(k,t,\widetilde{\delta})\}} < c_n \Big) \\
&\leq \mathbb{P}\Big(\inf_{\delta_1 \leq |t| \leq \delta_2\vartheta_*} \sum_{j=1}^{n} \mathbb{1}_{\{W_j \in ([-C,C]\cap[c,d]) \setminus \cup_{k\in\mathbb{Z}} I(k,t,\widetilde{\delta})\}} < c_n \Big) .
\end{aligned} \tag{3.41}$$

(3.41) is summable over n since the number of W_j contained in the “good” sets is of order n, i.e. $\#\{j : W_j \in [c,d] \setminus \cup_{k\in\mathbb{Z}} I(k,t,\widetilde{\delta})\} = \mathcal{O}(n)$. Define

$$K(t) = \#\{k : I(k,t,\widetilde{\delta}) \cap [c,d] \neq \emptyset\}, \tag{3.42}$$

and let $k_1, \ldots k_{K(t)}$ enumerate the intervals contained in $[c, d]$. Let $m_1(t), \ldots, m_{K(t)}(t)$ be chosen such that $W_{m_i(t)} \in I(k_i, t, \widetilde{\delta})$. Note that $m_i(t)$ are random. The probability in the last line of (3.41) is bounded from above by

$$\mathbb{P}\left(\inf_{\delta_1 \leq |t| \leq \delta_2\vartheta_*} \sum_{j=1}^{n} \mathbb{1}_{\{W_j \in [c,d], |W_j - W_{m_1(t)}| > 2\widetilde{\delta}, \ldots, |W_j - W_{m_{K(t)}(t)}| > 2\widetilde{\delta}\}} \leq c_n \right) . \tag{3.43}$$

Since there are only finitely many intervals of length $2\widetilde{\delta}$ with distance $6\widetilde{\delta}$ to each other in $[c, d]$, there exists $K < \infty$ such that $\sup_{t\in[\delta_1,\delta_2\vartheta_*]} K(t) < K$. Thus, the probability in (3.43) is not larger than

$$\begin{aligned}
&\mathbb{P}\left(\exists_{m_1,\dots,m_K\in\{1,\dots,n\}}:\sum_{j=1}^{n}\mathbb{1}_{\left\{W_j\in[c,d],|W_j-W_{m_1}|>2\widetilde{\delta},\dots,|W_j-W_{m_K}|>2\widetilde{\delta}\right\}}\leq c_n\right)\\
&\leq\sum_{m_1,\dots,m_K=1}^{n}\mathbb{P}\left(\sum_{j=1}^{n}\mathbb{1}_{\left\{W_j\in[c,d],|W_j-W_{m_1}|>2\widetilde{\delta},\dots,|W_j-W_{m_K}|>2\widetilde{\delta}\right\}}\leq c_n\right)\\
&\leq n^K\mathbb{P}\left(\sum_{j=1}^{n}\mathbb{1}_{\left\{W_j\in[c,d],|W_j-W_{m_1}|>2\widetilde{\delta},\dots,|W_j-W_{m_K}|>2\widetilde{\delta}\right\}}\leq c_n\right).
\end{aligned}\tag{3.44}$$

The indicator function vanishes whenever $j=m_i$ with $i\in\{1,\dots,K\}$. Thus,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{j=1}^{n}\mathbb{1}_{\left\{W_j\in[c,d],|W_j-W_{m_1}|>2\widetilde{\delta},\dots,|W_j-W_{m_K}|>2\widetilde{\delta}\right\}}\leq c_n\right)\\
&=\mathbb{P}\left(\sum_{j\notin\{m_1,\dots,m_K\}}^{n}\mathbb{1}_{\left\{W_j\in[c,d],|W_j-W_{m_1}|>2\widetilde{\delta},\dots,|W_j-W_{m_K}|>2\widetilde{\delta}\right\}}\leq c_n\right)\\
&=\mathbb{P}\left(\sum_{j=K}^{n}\mathbb{1}_{\left\{W_j\in[c,d],|W_j-W_1|>2\widetilde{\delta},\dots,|W_j-W_K|>2\widetilde{\delta}\right\}}\leq c_n\right)
\end{aligned}\tag{3.45}$$

due to the i.i.d. assumption. (3.45) is equal to

$$\sum_{l=0}^{\lceil c_n\rceil}\binom{n-K}{l}\mathbb{P}(A)^l\left(1-\mathbb{P}(A)\right)^{n-K-l}\leq\lceil c_n\rceil\binom{n-K}{\lceil c_n\rceil}(1-\mathbb{P}(A))^{n-K-\lceil c_n\rceil}.\tag{3.46}$$

Here A is the event

$$A=\left\{W\in[c,d],|W-W_1|>2\widetilde{\delta},\dots,|W-W_K|>2\widetilde{\delta}\right\},\tag{3.47}$$

where W is an independent copy of W_1. We show that $\mathbb{P}(A)$ is strictly positive.

$$\begin{aligned}
\mathbb{P}(A)\ &=\ \int_{[c,d]}\mathbb{P}\left(|W-W_1|>2\widetilde{\delta},\dots,|W-W_K|>2\widetilde{\delta}\,\middle|\,W\right)dP_W \qquad (3.48)\\
&\geq\ \int_{[c,d]}\mathbb{P}\left(W_i\in[W-2\widetilde{\delta},W+2\widetilde{\delta}]^c\cap[c,d],\forall i\in\{1,\dots,K\}\right)dP_W,
\end{aligned}$$

where P_W denotes the distribution of W. Since the random variables $W_1,\dots,W_K,W$ are independent of each other, this is equal to

$$\int_{[c,d]}\mathbb{P}\left(W_1\in[W-2\widetilde{\delta},W+2\widetilde{\delta}]^c\cap[c,d]|W\right)^K dP_W,\tag{3.49}$$

and due to the lower bound on the density of P_W postulated in Hypothesis (ia), this in turn is bounded from below by

$$(p(d-c-4\widetilde{\delta}))^K\int_{[c,d]}dP_W\geq(d-c)p^{K+1}(d-c-4\widetilde{\delta})^K\equiv\widetilde{p}\in(0,1].\tag{3.50}$$

Combining Equations (3.46) and (3.50) we obtain

$$(3.44)\leq n^K\lceil c_n\rceil\binom{n}{\lceil c_n\rceil}\widetilde{p}^{\,n-K-\lceil c_n\rceil}\tag{3.51}$$

which is summable over n, as desired. Thus all hypotheses of Theorem 1.3 are satisfied with probability one, uniformly in $a \in J$, and so the conclusion of Theorem 1.6 follows. $\square$

4. Proof of Theorem 1.9

In order to prove Theorem 1.9 we need the joint weak convergence of the process X_n, defined in (1.11) and its derivatives, as stated in Lemma 4.1. Define on the closure, $\bar{J}$ of the interval J (recall the definition of J in Theorem 1.6), the processes $(\widehat{X}_a^n)_{a\in\bar{J}}$, $n \in \mathbb{N}$, via

$$\widehat{X}_a^n \equiv (X_n(\vartheta(a)), X_n'(\vartheta(a)), X_n''(\vartheta(a))). \tag{4.1}$$

Lemma 4.1. *The family of processes $(\widehat{X}_a^n)_{a\in\bar{J}}$ defined on $\big(C(\bar{J},\mathbb{R}^3), \mathcal{B}(C(\bar{J},\mathbb{R}^3))\big)$, converges weakly, as $n \to \infty$, to a process $(\widehat{X}_a)_{a\in\bar{J}}$ on the same space, if there exists $c > 0$, such that, for all $a \in \bar{J}$, $g''(\vartheta(a)) > c$, and if Assumption 1.8 is satisfied.*

Proof: As usual, we prove convergence of the finite dimensional distributions and tightness.

More precisely, we have to check that:

(1) $(\widehat{X}_a^n)_{a\in\bar{J}}$ converges in finite dimensional distribution.
(2) The family of initial distributions, i.e. the distributions of $\widehat{X}_b^n$, where $b \equiv \mathbb{E}[Z_1W_1]$, is tight.
(3) There exists $C > 0$ independent of a and n such that
$$\mathbb{E}\left[\|\widehat{X}_{a+h}^n - \widehat{X}_a^n\|^2\right] \le C|h|^2, \tag{4.2}$$
which is a Kolmogorov-Chentsov criterion for tightness, see Kallenberg (2002, Corollary 14.9).

First, we consider the finite dimensional distributions. Let

$$\begin{aligned} Y_{a,j} &\equiv f(\vartheta(a)W_j) - \mathbb{E}\left[f(\vartheta(a)W_j)\right] \\ Y_{a,j}' &\equiv W_j f'(\vartheta(a)W_j) - \mathbb{E}[W_j f'(\vartheta(a)W_j)] \quad \text{and} \\ Y_{a,j}'' &\equiv W_j^2 f''(\vartheta(a)W_j) - \mathbb{E}[W_j^2 f''(\vartheta(a)W_j)]. \end{aligned} \tag{4.3}$$

Moreover, let $\ell \in \mathbb{N}$, $a_1 < a_2 < \cdots < a_\ell \in \bar{J}$ and

$$\chi_j \equiv \left(Y_{a_1,j}, Y_{a_1,j}', Y_{a_1,j}'', \ldots, Y_{a_\ell,j}, Y_{a_\ell,j}', Y_{a_\ell,j}''\right) \in \mathbb{R}^{3\ell}. \tag{4.4}$$

These vectors are independent for different j and the components $(\chi_j)_k, 1 \le k \le 3\ell$, have covariances $\mathrm{Cov}((\chi_j)_k, (\chi_j)_m) = \mathcal{C}_{km} < C$ for all $k, m \in \{1, \ldots, 3\ell\}$, according to Assumption 1.8. Therefore, $\frac{1}{\sqrt{n}}\sum_{j=1}^n \chi_j$ converges, as $n \to \infty$, to the 3ℓ-dimensional Gaussian vector with mean zero and covariance matrix $\mathcal{C}$ by the central limit theorem. This proves convergence of the finite dimensional distributions of $(\widehat{X}_a^n)_{a\in\bar{J}}$.

The family of initial distributions is given by the random variables evaluated at $\vartheta(b)$. This family is seen to be tight using Chebychev's inequality

$$\begin{aligned} \mathbb{P}\left(\|X_b^n\|_2 > C\right) &\le \frac{\mathbb{V}\left[\sqrt{X_n(\vartheta(b))^2 + (X_n'(\vartheta(b)))^2 + (X_n''(\vartheta(b)))^2}\right]}{C^2} \\ &\le \frac{\mathbb{E}\left[X_n(\vartheta(b))^2 + (X_n'(\vartheta(b)))^2 + (X_n''(\vartheta(b)))^2\right]}{C^2}. \end{aligned} \tag{4.5}$$

which is finite by Assumption 1.8. For each ε we can choose C large enough such that (4.5) $< \varepsilon$. It remains to check Condition (3). Since

$$\begin{aligned} \mathbb{E}\left[\|\widehat{X}^n_{a+h} - \widehat{X}^n_a\|^2\right] &= \mathbb{E}\left[(X_n(\vartheta(a+h)) - X_n(\vartheta(a)))^2\right] \\ &+\mathbb{E}\left[[X'_n(\vartheta(a+h)) - X'_n(\vartheta(a)))^2\right] \\ &+\mathbb{E}\left[(X''_n(\vartheta(a+h)) - X''_n(\vartheta(a)))^2\right], \end{aligned} \tag{4.6}$$

we need to show that each of the three terms on the right-hand side is of order h^2. Note that $\mathbb{E}\left[[X_n(\vartheta(a+h)) - X_n(\vartheta(a))]^2\right] \leq C|h|^2$ if

$$\mathbb{E}\left[\left(\tfrac{d}{da}X_n(\vartheta(a))\right)^2\right] \leq C. \tag{4.7}$$

Since $X_n(\vartheta(a)) = \frac{1}{\sqrt{n}}\sum_{j=1}^n Y_{a,j}$,

$$\mathbb{E}\left[\left(\tfrac{d}{da}X_n(\vartheta(a))\right)^2\right] = \frac{1}{n}\sum_{j=1}^n \mathbb{E}\left[\left(\tfrac{d}{da}Y_{a,j}\right)^2\right]. \tag{4.8}$$

Each summand can be controlled by

$$\begin{aligned} &\mathbb{E}\left[\left(\tfrac{d}{da}Y_{a,j}\right)^2\right] = \mathbb{E}\left[\left(\tfrac{d}{da}f(\vartheta(a)W_j)\right)^2\right] - \left(\mathbb{E}\left[\tfrac{d}{da}f(\vartheta(a)W_j)\right]\right)^2 \\ &= \left(\tfrac{d}{da}\vartheta(a)\right)^2\left(\mathbb{E}\left[W_j^2 f'(\vartheta(a)W_j)^2\right] - \left(\mathbb{E}\left[W_j f'(\vartheta(a)W_j)\right]\right)^2\right) \\ &= \left(\tfrac{d}{da}\vartheta(a)\right)^2 \mathbb{V}\left[W_j f'(\vartheta(a)W_j)\right]. \end{aligned} \tag{4.9}$$

By the implicit function theorem,

$$\frac{d}{da}\vartheta(a) = \left(g''(\vartheta(a))\right)^{-1}. \tag{4.10}$$

Thus, Equation (4.7) holds since $g''(\vartheta(a)) > c$ by assumption and $\mathbb{V}\left[W_j f'(\vartheta(a)W_j)\right]$ is bounded by Assumption 1.8. The bounds for the remaining terms follow in the same way by controlling the derivatives of $X'_n(\vartheta(a))$ and $X''_n(\vartheta(a))$. We obtain

$$\begin{aligned} \mathbb{E}\left[\left(\tfrac{d}{da}Y'_{a,j}\right)^2\right] &= \mathbb{E}\left[\left(\tfrac{d}{da}W_j f'(\vartheta(a)W_j)\right)\right] - \left(\mathbb{E}\left[\tfrac{d}{da}W_j f'(\vartheta(a)W_j)\right]\right)^2 \\ &= \left(\tfrac{d}{da}\vartheta(a)\right)^2 \mathbb{V}\left[W_j^2 f''(\vartheta(a)W_j)\right] \end{aligned} \tag{4.11}$$

and

$$\begin{aligned} \mathbb{E}\left[\left(\tfrac{d}{da}Y''_{a,j}\right)^2\right] &= \mathbb{E}\left[\left(\tfrac{d}{da}W_j^2 f''(\vartheta(a)W_j)\right)\right] - \left(\mathbb{E}\left[\tfrac{d}{da}W_j^2 f''(\vartheta(a)W_j)\right]\right)^2 \\ &= \left(\tfrac{d}{da}\vartheta(a)\right)^2 \mathbb{V}\left[W_j^3 f'''(\vartheta(a)W_j)\right]. \end{aligned} \tag{4.12}$$

In both formulae the right hand sides are bounded due to Assumption 1.8. This proves the lemma. □

Proof of Theorem 1.9: Recall that $\vartheta_n^{\mathcal{W}}(a)$ is determined as the solution of the equation

$$a = g'(\vartheta) + \frac{1}{\sqrt{n}}X'_n(\vartheta). \tag{4.13}$$

Write $\vartheta_n^{\mathcal{W}}(a) \equiv \vartheta(a) + \delta^n(a)$, where $\vartheta(a)$ is defined as the solution of

$$a = g'(\vartheta). \tag{4.14}$$

Note that $\vartheta(a)$ is deterministic while $\delta^n(a)$ is random and $\mathcal{W}$-measurable . The rate function can be rewritten as

$$I_n^{\mathcal{W}}(a) = a(\vartheta(a) + \delta^n(a)) - g(\vartheta(a) + \delta^n(a)) - \frac{1}{\sqrt{n}} X_n(\vartheta(a) + \delta^n(a)). \tag{4.15}$$

A second order Taylor expansion and reordering of the terms yields

$$\begin{aligned} I_n^{\mathcal{W}}(a) = \quad & \underbrace{a\vartheta(a) - g(\vartheta(a))}_{\equiv I(a)} - \frac{1}{\sqrt{n}} X_n(\vartheta(a)) \\ & + \underbrace{(a - g'(\vartheta(a)))}_{=0} \delta^n(a) - \frac{1}{\sqrt{n}} \delta^n(a) X_n'(\vartheta(a)) \\ & - \frac{1}{2} (\delta^n(a))^2 \left(g''(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n''(\vartheta(a)) \right) + \mathrm{o}((\delta^n(a))^2). \end{aligned} \tag{4.16}$$

Note that the leading terms on the right-hand side involve the three components of the processes $\widehat{X}^n$ whose convergence we have just proven. We obtain the following equation for $\delta^n(a)$ using a first order Taylor expansion.

$$\begin{aligned} a &= g'(\vartheta(a) + \delta^n(a)) + \frac{1}{\sqrt{n}} X_n'(\vartheta(a) + \delta^n(a)) \qquad (4.17) \\ &= g'(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n'(\vartheta(a)) + \delta^n(a) \left(g''(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n''(\vartheta(a)) \right) + \mathrm{o}(\delta^n(a)), \end{aligned}$$

which implies

$$\delta^n(a) = \frac{-\frac{1}{\sqrt{n}} X_n'(\vartheta(a))}{g''(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n''(\vartheta(a))} + \mathrm{o}(\delta^n(a)). \tag{4.18}$$

Lemma 4.1 combined with $g''(\vartheta(a)) = \mathcal{O}(1)$ yields $\delta^n(a) = \mathcal{O}(1/\sqrt{n})$. We insert the expression for $\delta^n(a)$ into Equation (4.16) to obtain

$$\begin{aligned} & I_n^{\mathcal{W}}(a) \\ & = I(a) - \frac{1}{\sqrt{n}} X_n(\vartheta(a)) - \frac{1}{2} \left(g''(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n''(\vartheta(a)) \right) \\ & \times \left[\frac{\frac{1}{n} (X_n'(\vartheta(a)))^2}{\left(g''(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n''(\vartheta(a)) \right)^2} - \frac{\frac{1}{\sqrt{n}} X_n'(\vartheta(a)) \mathrm{o}(\delta^n)}{\left(g''(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n''(\vartheta(a)) \right)} + \mathrm{o}((\delta^n)^2) \right] \\ & + \frac{\frac{1}{n} (X_n'(\vartheta(a)))^2}{g''(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n''(\vartheta(a))} + \frac{1}{\sqrt{n}} X_n'(\vartheta(a)) \mathrm{o}(\delta^n) + \mathrm{o}((\delta^n)^2). \end{aligned} \tag{4.19}$$

Combining this with the bound (4.18), it follows that

$$I_n^{\mathcal{W}}(a) = I(a) - \frac{1}{\sqrt{n}} X_n(\vartheta(a)) + \frac{1}{n} r_n(a), \tag{4.20}$$

where

$$r_n(a) \equiv \frac{\frac{1}{2} (X_n'(\vartheta(a)))^2}{g''(\vartheta(a)) + \frac{1}{\sqrt{n}} X_n''(\vartheta(a))} + \mathrm{o}(1). \tag{4.21}$$

$r_n(a)$ converges weakly due to the continuous mapping theorem and the joint weak convergence of $X_n'(\vartheta(a))$ and $X_n''(\vartheta(a))$. This completes the proof of the theorem. □

References

R.R. Bahadur and R. Ranga Rao. On deviations of the sample mean. *Ann. Math. Statist.* **31**, 1015–1027 (1960). MR0117775.

H.A. van den Berg, D.A. Rand and N.J. Burroughs. A Reliable and Safe T Cell Repertoire based on Low-affinity T Cell Receptors. *J. Theor. Biol.* **209** (4), 465 – 486 (2001). DOI: 10.1006/jtbi.2001.2281.

N. R. Chaganty and J. Sethuraman. Strong large deviation and local limit theorems. *Ann. Probab.* **21** (3), 1671–1690 (1993). MR1235434.

F. Comets. Large deviation estimates for a conditional probability distribution. Applications to random interaction Gibbs measures. *Probab. Theory Related Fields* **80** (3), 407–432 (1989). MR976534.

A. Dembo, J.-D. Deuschel and D. Duffie. Large portfolio losses. *Finance Stoch.* **8** (1), 3–16 (2004). MR2022976.

A. Dembo and I. Kontoyiannis. The asymptotics of waiting times between stationary processes, allowing distortion. *Ann. Appl. Probab.* **9** (2), 413–429 (1999). MR1687410.

A. Dembo and I. Kontoyiannis. Source coding, large deviations, and approximate pattern matching. *IEEE Trans. Inform. Theory* **48** (6), 1590–1615 (2002). Special issue on Shannon theory: perspective, trends, and applications. MR1909475.

A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin (2010). ISBN 978-3-642-03310-0. Corrected reprint of the second (1998) edition. MR2571413.

W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney (1971). MR0270403.

F. den Hollander. *Large deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI (2000). ISBN 0-8218-1989-5. MR1739680.

M. Iltis. Sharp asymptotics of large deviations in $\mathbf{R}^d$. *J. Theoret. Probab.* **8** (3), 501–522 (1995). MR1340824.

O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition (2002). ISBN 0-387-95313-2. MR1876169.

H. Mayer and A. Bovier. Stochastic modelling of T-cell activation. *J. Math. Biol.* **70** (1-2), 99–132 (2015). MR3294967.

K. Spiliopoulos and R.B. Sowers. Recovery rates in investment-grade pools of credit assets: a large deviations analysis. *Stochastic Process. Appl.* **121** (12), 2861–2898 (2011). MR2844544.

N. Zint, E. Baake and F. den Hollander. How T-cells use large deviations to recognize foreign antigens. *J. Math. Biol.* **57** (6), 841–861 (2008). MR2439675.