

## POISSON APPROXIMATION FOR TWO SCAN STATISTICS WITH RATES OF CONVERGENCE

BY XIAO FANG<sup>1</sup> AND DAVID SIEGMUND<sup>2</sup>

*National University of Singapore and Stanford University*

As an application of Stein's method for Poisson approximation, we prove rates of convergence for the tail probabilities of two scan statistics that have been suggested for detecting local signals in sequences of independent random variables subject to possible change-points. Our formulation deals simultaneously with ordinary and with large deviations.

**1. Introduction.** Let  $\{X_1, \dots, X_n\}$  be a sequence of random variables. A widely studied problem is to test the hypothesis that the  $X$ 's are independent and identically distributed against the alternative that for some  $0 \leq i < j \leq n$ ,  $\{X_{i+1}, \dots, X_j\}$  have a distribution that differs from the distribution of the other  $X$ 's. If  $t := j - i$  is assumed known and the change in distribution is a shift in the mean, one common suggestion to detect the change is the statistic

$$(1.1) \quad M_{n;t} = \max_{1 \leq i \leq n-t+1} (X_i + \dots + X_{i+t-1}).$$

See Glaz, Naus and Wallenstein (2001) for an introduction to scan statistics.

When  $t$  is unknown but the distributions of the  $X$ 's are otherwise completely specified, the maximum log likelihood ratio statistic is

$$(1.2) \quad \max_{0 \leq i < j \leq n} (S_j - S_i),$$

where

$$(1.3) \quad S_i = \sum_{k=1}^i \log[f_1(X_k)/f_0(X_k)]$$

and  $f_0$  ( $f_1$ , resp.) is the density function of  $X$  under the null hypothesis (alternative hypothesis resp.). Appropriate statistics when the distributions involve unknown parameters can be found, for example, in Yao (1993).

---

Received March 2014; revised May 2015.

<sup>1</sup>Supported in part by the NUS-Overseas Postdoctoral Fellowship from the National University of Singapore.

<sup>2</sup>Supported in part by the National Science Foundation.

*MSC2010 subject classifications.* Primary 60F05, 62E20; secondary 62L10.

*Key words and phrases.* Stein's method, Poisson approximation, total variation distance, relative error, rate of convergence, scan statistics, change-point analysis, exponential family.

Asymptotic  $p$  values of test statistics (1.1) and (1.2) have been derived as  $n \rightarrow \infty$  under certain distributional assumptions on  $X_1$ ; see, for example, Arratia, Gordon and Waterman (1990), Haiman (2007) and Siegmund (1988). The statistic (1.2) has also been studied for its role in queueing theory, where it has the interpretation of the maximum waiting time among the first  $n$  customers of a single server queue [cf. Iglehart (1972)]. However, except for (1.1) in the special case when  $X_1$  is a Bernoulli variable [cf. Arratia, Gordon and Waterman (1990) and Haiman (2007)], and for (1.2) when the problem is scaled so that the probability is approximately zero [cf. Siegmund (1988)], the rate of convergence for these approximations is unknown. In this paper, we establish rate of convergence of tail approximations for both statistics (1.1) and (1.2) under the assumption that  $X_1$  comes from an exponential family of distributions. The error in our approximation is relative error, hence is applicable when the probability is small as well as when it converges to a positive limit.

In practice, simulations have been widely used to justify the accuracy of the approximations suggested here. The sample size  $n$  used in those simulations is typically a few thousands at most, partly because the simulation would take too long for larger  $n$ . We have not seen any related work trying to infer a convergence rate by simulation results. The constants arising from our calculations are undoubtedly much too large to be an alternative source to justify use of the approximations in practice. We view the value of our approximations as providing understanding of the relations of various parameters involved in the approximations, and in particular the uniformity of the validity of the approximation for both large and ordinary deviations.

In the next section, we state our main results. Section 3 contains an introduction to our main technique, Stein’s method and the proof of our main results. We discuss related problems in Section 4.

**2. Main results.**

2.1. *Scan statistics with fixed window size.* Let  $\{X_1, \dots, X_n\}$  be independent, identically distributed random variables with distribution function  $F$  and  $\mathbb{E}X_1 = \mu_0$ . Suppose the distribution of  $X_1$  can be imbedded in an exponential family of probability measures  $\{F_\theta : \theta \in \Theta\}$  where  $\Theta$  is an open interval in  $\mathbb{R}$  containing 0, and

$$(2.1) \quad dF_\theta(x) = e^{\theta x - \Psi(\theta)} dF(x).$$

It is known that the mean and variance of  $F_\theta$  are  $\Psi'(\theta)$  and  $\Psi''(\theta)$ , respectively. We assume  $F_\theta(x)$  is nondegenerate, that is,  $\Psi''(\theta) > 0$  for all  $\theta \in \Theta$ . From  $X_1 \sim F$  and  $\mathbb{E}X_1 = \mu_0$ , we have  $\Psi(0) = 0$  and  $\Psi'(0) = \mu_0$ .

Let  $a > \mu_0$  be given. Assume that there exists  $\theta_a \in \Theta$  such that  $\Psi'(\theta_a) = a$ . For a positive integer  $t < n$ , and for  $M_{n;t}$  defined in (1.1), we are interested in calculating approximately the probability  $\mathbb{P}(M_{n;t} \geq at)$ . In the following theorem,

we provide a Poisson approximation result with rate of convergence. We consider the following two cases:

Case 1: There exists  $\theta_{a'} \in \Theta$  such that  $\theta_{a'} > \theta_a$  [thus  $\Psi'(\theta_{a'}) = a' > a$ ] and

$$(2.2) \quad \sup_{\theta_a \leq \theta \leq \theta_{a'}} \int_{-\infty}^{\infty} |\varphi_{\theta}(x)|^p dx < \infty \quad \text{for some positive integer } p,$$

where  $\varphi_{\theta}$  is the characteristic function of  $F_{\theta}$ .

Case 2:  $X_1$  is an integer-valued random variable with span 1, where the span is defined to be the largest value of  $\Delta$  such that

$$(2.3) \quad \sum_{k \in \mathbb{Z}} \mathbb{P}(X_i = k\Delta + w) = 1 \quad \text{for some } w \in \mathbb{Z}.$$

We remark that (2.2) is a smoothness condition on  $F$  [cf. Condition 1.4 of Diaconis and Freedman (1988)]. Note also that any lattice random variable, i.e., that satisfying (2.3) with  $w \in \mathbb{R}$  instead of  $w \in \mathbb{Z}$ , can be reduced to case 2 by linear transformation.

For the statement in the following theorem, we define  $\sigma_a^2 = \Psi''(\theta_a)$  and  $\lceil at \rceil = \inf\{v \in \mathbb{Z} : v \geq at\}$ . Let  $\{X_i, X_i^a : i \geq 1\}$  be independent with  $X_i \sim F$  and  $X_i^a \sim F_{\theta_a}$ , and let  $D_k = \sum_{i=1}^k (X_i^a - X_i)$  for  $k \geq 1$ .

**THEOREM 2.1.** *Under the assumptions given above, for some constant  $C$  depending only on the exponential family (2.1),  $\mu_0$ , and  $a$ , we have*

$$(2.4) \quad \begin{aligned} &|\mathbb{P}(M_{n;t} \geq at) - (1 - e^{-\lambda})| \\ &\leq C \left( \frac{(\log t)^2}{t} + \frac{(\log t \wedge \log(n-t))}{n-t} \right) (\lambda \wedge 1), \end{aligned}$$

where for case 1,

$$(2.5) \quad \lambda = \frac{(n-t+1)e^{-[a\theta_a - \Psi(\theta_a)]t}}{\theta_a \sigma_a (2\pi t)^{1/2}} \exp \left[ - \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}(e^{-\theta_a D_k^+}) \right],$$

and for case 2,

$$(2.6) \quad \lambda = \frac{(n-t+1)e^{-(a\theta_a - \Psi(\theta_a))t} e^{-\theta_a(\lceil at \rceil - at)}}{(1 - e^{-\theta_a}) \sigma_a (2\pi t)^{1/2}} \exp \left[ - \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}(e^{-\theta_a D_k^+}) \right].$$

**REMARK 2.1.** The various expressions entering into  $\lambda$  will be explained below. Here, it is important to note that provided  $n - t$  and  $t$  are large the error of approximation is relative error, valid when  $n$  is relatively small, so  $\lambda$  is near zero, and when  $\lambda$  is bounded away from zero. Although it is possible to trace through the proof of Theorem 2.1 and obtain a numerical value for the constant  $C$  in (2.4), it would be too large for practical purposes. Therefore, we do not pursue it here.

REMARK 2.2. [Arratia, Gordon and Waterman \(1990\)](#) obtained a bound for  $|\mathbb{P}(M_{n;t} \geq at) - (1 - e^{-\lambda})|$  for independent, identically distributed Bernoulli random variables. They do not restrict the threshold ( $at$  in our case) to grow linearly in  $t$  with fixed slope. For fixed  $a$ , their bound is of the form [cf. equations (11)–(13) of [Arratia, Gordon and Waterman \(1990\)](#)]

$$C \left( e^{-ct} + \frac{t}{n} \right) (\lambda \wedge 1).$$

Compared to their result, [Theorem 2.1](#) applies to more general distributions and recovers typical limit theorems in the literature on scan statistics. As  $t, n - t \rightarrow \infty$ , [Theorem 2.1](#) guarantees the relative error in (2.4) goes to 0; see, for example, [Theorem 1 of Chan and Zhang \(2007\)](#).

REMARK 2.3. The infinite series appearing in the definition of  $\lambda$  is derived as an application of classical random walk results of [Spitzer](#). It arises probabilistically in the proof of [Theorem 2.1](#) in the form  $\mathbb{E}[1 - \exp\{-\theta_a D_{\tau_+}\}]/\mathbb{E}(\tau_+)$ , where  $\tau_+ = \inf\{t : D_t > 0\}$ . The series form is useful for numerical computation. Let  $g(x) = \mathbb{E}e^{ixD_1}$  and  $\xi(x) = \log\{1/[1 - g(x)]\}$ . [Woodroffe \(1979\)](#) proved that for case 1 of [Theorem 2.1](#),

$$(2.7) \quad \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}(e^{-\theta_a D_k^+}) = -\log[(a - \mu_0)\theta_a] - \frac{1}{\pi} \int_0^{\infty} \frac{\theta_a^2 [I\xi(x) - \pi/2]}{x(\theta_a^2 + x^2)} dx + \frac{1}{\pi} \int_0^{\infty} \frac{\theta_a \{R\xi(x) + \log[(a - \mu_0)x]\}}{\theta_a^2 + x^2} dx,$$

where  $R$  and  $I$  denote real and imaginary parts. [Tu and Siegmund \(1999\)](#) proved that for case 2 of [Theorem 2.1](#),

$$(2.8) \quad \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}(e^{-\theta_a D_k^+}) = -\log(a - \mu_0) + \frac{1}{2\pi} \int_0^{2\pi} \left\{ \frac{\xi(x)e^{-\theta_a - ix}}{1 - e^{-\theta_a - ix}} + \frac{\xi(x) + \log[(a - \mu_0)(1 - e^{ix})]}{1 - e^{ix}} \right\} dx.$$

The right-hand sides of (2.7) and (2.8) can be calculated by numerical integration. For example, [Woodroffe \(1979\)](#) calculated the right-hand side of (2.7) for normal, gamma and chi-squared distributions, and [Tu and Siegmund \(1999\)](#) calculated the right-hand side of (2.8) for binomial distributions.

REMARK 2.4. For  $M_{n;t}$  defined in (1.1) and  $b > 0$ , [Dembo and Karlin \(1992\)](#) proposed the simple approximation to  $\mathbb{P}(M_{n;t} \geq b)$  given by  $1 - e^{-\lambda}$ , where [cf. [Theorem 2 of Dembo and Karlin \(1992\)](#)]

$$\lambda = (n - t + 1)\mathbb{P}(X_1 + \dots + X_t \geq b).$$

Similar approximations have also been considered for more complicated biological models. See [Chen and Karlin \(2000\)](#), [Karlin and Chen \(2000\)](#) and [Chen and Karlin \(2007\)](#). Such a simple approximation requires specific conditions on the relation of  $b$ ,  $t$  and  $n$  and does not hold when  $b$  is proportional to  $t$ , which leads to the “clumping” phenomenon; see, for example, Section 4.2 of [Arratia, Goldstein and Gordon \(1990\)](#) or the book by [Aldous \(1989\)](#). In applications, one must judge whether the appropriate scaling relations hold for specific values of  $t$  and  $b$ . In this regard, it is interesting to note that our approximation becomes the Dembo–Karlin approximation when the scaling relations of [Dembo and Karlin \(1992\)](#) are satisfied.

Next, we compute the limiting probability  $1 - e^{-\lambda}$  in (2.4) explicitly for normal and Bernoulli random variables. We show that the limiting probability is close to the true probability by using simulation and known results.

EXAMPLE 2.1. Suppose  $X_1 \sim N(0, 1)$ . We have that  $X_1^a \sim N(a, 1)$ ,  $D_1 \sim N(a, 2)$  and in the definition of  $\lambda$  in (2.5),

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}(e^{-\theta_a D_k^+}) &= 2 \sum_{k=1}^{\infty} \frac{1}{k} \Phi(-a\sqrt{k/2}) \\ &= -\log[(a - \mu_0)^2 v(\sqrt{2}a)], \end{aligned}$$

where  $\Phi$  is the standard normal distribution function and the function  $v(x)$  was defined in (4.38) of [Siegmund \(1985\)](#). It was shown there that  $v(x) = \exp(-cx) + o(x^2)$  as  $x \rightarrow 0$  for  $c \approx 0.583$ , while  $x^2 v(x)/2 \rightarrow 1$  as  $x \rightarrow \infty$ . [Siegmund and Yakir \(2007\)](#) indicate that a very simple and good approximation is

$$v(x) \approx [(2/x)\{\Phi(x/2) - 1/2\}]/\{(x/2)\Phi(x/2) + \phi(x/2)\},$$

where  $\phi$  is the standard normal density function. Table 1 presents a numerical study with different values of  $n$ ,  $t$  and  $a$ . The limiting probability  $1 - e^{-\lambda}$  is denoted by  $p_1$ . The values of  $p_2$  are simulated with 10,000 repetitions each. We can see from the table that  $p_1$  is very close to the true probability.

TABLE 1  
Example 2.1

$n$	$t$	$a$	$p_1$	$p_2$
1000	50	0.2	0.9315	0.9594
1000	50	0.4	0.2429	0.2624
1000	50	0.5	0.0331	0.0334
2000	50	0.5	0.0668	0.0672

EXAMPLE 2.2. Let  $\{X_1, \dots, X_n\}$  be a sequence of independent Bernoulli random variables with  $\mathbb{P}(X_i = 1) = \mu_0$  for all  $i$  where  $0 < \mu_0 < 1$ . The distribution of  $X_1$  can be imbedded in an exponential family of probability measures  $\{F_\theta : \theta \in \mathbb{R}\}$  where  $F_\theta$  is defined as in (2.1) with

$$(2.9) \quad \Psi(\theta) = \log\left(1 + \frac{\mu_0 e^\theta}{1 - \mu_0}\right) + \log(1 - \mu_0).$$

For  $0 < p < 1$ , define

$$(2.10) \quad \theta_p = \log\left(\frac{p}{1 - p}\right) - \log\left(\frac{\mu_0}{1 - \mu_0}\right).$$

It is straightforward to check that

$$\Psi'(\theta_p) = p, \quad \theta_{\mu_0} = 0, \quad \Psi(\theta_{\mu_0}) = 0.$$

Let  $1 > a > \mu_0$  (see Corollary 2.1 for the relatively easier case where  $a = 1$ ). From (2.6) and (2.8), we have

$$(2.11) \quad \begin{aligned} \lambda = & \frac{(n - t + 1)e^{-(a\theta_a - \Psi(\theta_a))t} e^{-\theta_a(\lceil at \rceil - at)}}{(1 - e^{-\theta_a})[2a(1 - a)\pi t]^{1/2}} \times (a - \mu_0) \\ & \times \exp\left(-\frac{1}{2\pi} \int_0^{2\pi} \left\{ \frac{\xi(x)e^{-\theta_a - ix}}{1 - e^{-\theta_a - ix}} \right. \right. \\ & \left. \left. + \frac{\xi(x) + \log[(a - \mu_0)(1 - e^{ix})]}{1 - e^{ix}} \right\} dx \right), \end{aligned}$$

where  $\theta_a$  and  $\Psi(\theta_a)$  are defined in (2.10) and (2.9),

$$g(x) = a(1 - \mu_0)e^{ix} + (1 - a)\mu_0e^{-ix} + a\mu_0 + (1 - a)(1 - \mu_0)$$

and  $\xi(x) = \log\{1/[1 - g(x)]\}$ .

Let  $t < n$ , and let  $M_{n;t}$  be defined as in (1.1). The bound (2.4) suggests the following approximation to  $\mathbb{P}(M_{n;t} \geq at)$ :

$$(2.12) \quad \mathbb{P}(M_{n;t} \geq at) \approx 1 - e^{-\lambda},$$

where  $\lambda$  is defined in (2.11). Table 2 presents a numerical study with different values of  $n, t$  and  $a$ . The probability  $p_1$  is calculated by the right-hand side of (2.12).

TABLE 2  
Example 2.2

$n$	$t$	$\mu_0$	$a$	$p_1$	$p_2$
7680	30	0.1	11/30	0.14097	0.14021
7680	30	0.1	0.4	0.029614	0.029387
15360	30	0.1	0.4	0.058458	0.058003

The values of  $p_2$  are found in Table 1 of Haiman (2007) and are shown there to be very accurate. We can see from the table that  $p_1$  is very close to the true probability. The derivation in Haiman (2007) uses the distribution function of

$$Z_k := \max\{T_1, \dots, T_{kt+1}\} \quad \text{for } k = 1 \text{ and } 2,$$

where  $T_\alpha = X_\alpha + \dots + X_{\alpha+t-1}$ . However, the distribution functions of  $Z_k$  for  $k = 1$  and  $2$  are only known in limited cases. See Haiman (2000) for another example on Poisson processes.

REMARK 2.5. From the proof of Theorem 2.1, the  $\lambda$  in Example 2.2 can be reduced to

$$\lambda = \frac{(n - t + 1)e^{-(a\theta_a - \Psi(\theta_a))t} e^{-\theta_a(\lceil at \rceil - at)}}{[2a(1 - a)\pi t]^{1/2}} \times (a - \mu_0).$$

The reason is that the intermediate quantity  $\lambda_2$  in (3.39) for Bernoulli random variables can be expressed as

$$\lambda_2 = (n - t + 1)\mathbb{P}(T_1 = \lceil at \rceil)\mathbb{P}(D_i > 0, i \geq 1)$$

and

$$\begin{aligned} \mathbb{P}(D_i > 0, i \geq 1) &= \mathbb{P}(D_1 > 0) \times \mathbb{P}(D_i \geq 0, i \geq 1) \\ &= a(1 - \mu_0) \times \left[ 1 - \frac{\mu_0(1 - a)}{a(1 - \mu_0)} \right] \\ &= a - \mu_0, \end{aligned}$$

where the second equation is from a known result for the first visit to  $-1$  of a Bernoulli random walk starting from  $0$  [see, e.g., page 272 of Feller (1968)].

The following corollary considers the case that  $X_i$  is integer-valued and  $a$  is the largest value  $X_i$  can take. The proof of it, which is deferred to Section 3, is simpler than the proof of Theorem 2.1 and the convergence rate we obtain is faster.

COROLLARY 2.1. *Let  $\{X_1, \dots, X_n\}$  be independent, identically distributed integer-valued random variables. For integers  $t < n$ , define  $M_{n;t}$  as in (1.1). Suppose  $a = \sup\{x : p_x := \mathbb{P}(X_1 = x) > 0\}$  is finite. We have, with constants  $C$  and  $c$  depending only on  $p_a$ ,*

$$(2.13) \quad \left| \mathbb{P}(M_{n;t} \geq at) - (1 - e^{-\lambda}) \right| \leq C(\lambda \wedge 1)e^{-ct},$$

where

$$\lambda = (n - t)p_a^t(1 - p_a) + p_a^t.$$

2.2. *Scan statistics with varying window size.* Next, we study the maximum log likelihood ratio statistic (1.2). Suppose in (1.3),  $f_0(x) = dF_{\theta_0}(x)$  and  $f_1(x) = dF_{\theta_1}(x)$  where  $\{F_\theta : \theta \in \Theta\}$  is an exponential family as in (2.1) and  $\theta_0 < \theta_1$ . Then we have

$$S_i = \sum_{k=1}^i \log[f_1(X_k)/f_0(X_k)]$$

$$= \sum_{k=1}^i (\theta_1 - \theta_0) \left( X_k - \frac{\Psi(\theta_1) - \Psi(\theta_0)}{\theta_1 - \theta_0} \right).$$

By appropriate change of parameters and a slight abuse of notation, studying (1.2) is equivalent to studying the following problem.

Let  $\{X_1, \dots, X_n\}$  be independent, identically distributed random variables with distribution function  $F$  that can be imbedded in an exponential family, as in (2.1). Let  $\mathbb{E}X_1 = \mu_0 < 0$ . Let  $S_0 = 0$  and  $S_i = \sum_{j=1}^i X_j$  for  $1 \leq i \leq n$ . We are interested in the distribution of  $\max_{0 \leq i < j \leq n} (S_j - S_i)$ . Statistics of this form have been widely studied in the context of CUSUM tests. Its limiting distribution was derived by Iglehart (1972), who observed that it can be interpreted as the maximum waiting time of the first  $n$  customers in a single server queue. Genomic applications are discussed by Karlin, Dembo and Kawabata (1990).

Suppose there exists a positive  $\theta_1 \in \Theta$  such that

$$(2.14) \quad \Psi'(\theta_1) = \mu_1, \quad \Psi(\theta_1) = 0.$$

For  $b > 0$ , in the following theorem we give an approximation to

$$(2.15) \quad p_{n,b} := \mathbb{P}\left(\max_{0 \leq i < j \leq n} (S_j - S_i) \geq b\right)$$

with an explicit error bound. We again consider two cases:

Case 1: The distribution  $F_{\theta_1}$  satisfies  $\int_{-\infty}^{\infty} |\varphi_{\theta_1}(t)| dt < \infty$ .

Case 2:  $X_1$  is an integer-valued random variable not concentrated on the set  $\{jd, -\infty < j < \infty\}$  for any  $d > 1$ .

In the following, let  $\mathbb{P}_\theta(\cdot)$  [ $\mathbb{E}_\theta(\cdot)$  resp.] denote the probability (expectation resp.) under which  $\{X_1, X_2, \dots\}$  are independent, identically distributed as  $F_\theta$ .

**THEOREM 2.2.** *Let  $h(b) > 0$  be any function such that*

$$h(b) \rightarrow \infty, \quad h(b) = O(b^{1/2}) \quad \text{as } b \rightarrow \infty.$$

*Suppose  $n - b/\mu_1 > b^{1/2}h(b)$ . Under the above setting, we have, for some constants  $c, C$  only depending on the exponential family  $F_\theta$  and  $\theta_1$ ,*

$$(2.16) \quad |p_{n,b} - (1 - e^{-\lambda})| \leq C\lambda \left\{ \left(1 + \frac{b/h^2(b)}{n - b/\mu_1}\right) e^{-ch^2(b)} + \frac{b^{1/2}h(b)}{n - b/\mu_1} \right\},$$



where for case 1,

$$\lambda = \left( n - \frac{b}{\mu_1} \right) \frac{e^{-\theta_1 b}}{\theta_1 \mu_1} \exp \left( -2 \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}_{\theta_1} e^{-\theta_1 S_k^+} \right),$$

and for case 2 and integers  $b$ ,

$$\lambda = \left( n - \frac{b}{\mu_1} \right) \frac{e^{-\theta_1 b}}{(1 - e^{-\theta_1}) \mu_1} \exp \left( -2 \sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}_{\theta_1} e^{-\theta_1 S_k^+} \right).$$

REMARK 2.6. We refer to Remark 2.3 for the numerical calculation of  $\lambda$ . By choosing  $h(b) = b^{1/2}$ , we get

$$|p_{n,b} - (1 - e^{-\lambda})| \leq C\lambda \left\{ e^{-cb} + \frac{b}{n} \right\}$$

from (2.16). By choosing  $h(b) = C(\log b)^{1/2}$  with large enough  $C$ , we can see that the relative error in the Poisson approximation goes to zero under the conditions

$$b \rightarrow \infty, \quad (b \log b)^{1/2} \ll n - b/\mu_1 = O(e^{\theta_1 b}),$$

where  $n - b/\mu_1 = O(e^{\theta_1 b})$  ensures that  $\lambda$  is bounded. For the smaller range (in which case  $\lambda \rightarrow 0$ )

$$b \rightarrow \infty, \quad \delta b \leq n - b/\mu_1 = o(e^{(1/2)\theta_1 b})$$

for some  $\delta > 0$ , Theorem 2 of Siegmund (1988) obtained more accurate estimates by a technique different from ours.

As in the case of Theorem 2.1, in some simple cases there is also the possibility here to evaluate  $\lambda$  by direct argument, and hence avoid the need for the numerical calculations of the general theory. Suppose the  $X_j$  are integer valued with either the maximum of the support equal to 1 or the minimum of the support equal to  $-1$ . Two interesting examples mentioned explicitly in Karlin, Dembo and Kawabata (1990) involve these possibilities. For example, assume that  $X_i$  equals  $k \geq 0$  with probability  $p_k$  and the negative value  $-k$  with probability  $q_k$ . Let  $G(z) = \sum_0^\infty p_k z^k + \sum_1^\infty q_k z^{-k}$ , and let  $z_0$  denote the unique root  $> 1$  of  $G(z) = 1$ . For the case  $p_k = 0$  for  $k > 1$ , using the notation  $Q(z) = \sum_k q_k z^k$ , one can show for large values of  $n$  and  $b$  that  $\lambda \sim n z_0^{-b} \{ [Q(1) - Q(z_0^{-1})] - (1 - z_0^{-1}) z_0^{-1} Q'(z_0^{-1}) \}$ . For the case  $q_k = 0$  for  $k > 1$ ,  $\lambda \sim n z_0^{-b} (1 - z_0^{-1}) |G'(1)|^2 / G'(z_0)$ . In particular if  $q_1 = q$  and  $p_1 = p$ , where  $p + q = 1$ , both these results specialize to  $\lambda \sim n(p/q)^b (q - p)^2 / q$ . These results differ from those given in Karlin, Dembo and Kawabata (1990), and hence produce numerical results slightly different from those cited there.

**3. Proofs.** Before proving our main theorems, we first introduce our main tool: Stein’s method. Stein’s method was first introduced by Stein (1972) and further developed in Stein (1986) for normal approximation. Chen (1975) developed Stein’s method for Poisson approximation, which has been widely applied especially in computational biology after the work by Arratia, Goldstein and Gordon (1990). We refer to Barbour and Chen (2005) for an introduction to Stein’s method.

The following theorem provides a useful upper bound on the total variation distance between the distribution of a sum of locally dependent Bernoulli random variables and a Poisson distribution. The total variation distance between two distributions is defined as

$$d_{TV}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{A \subset \mathbb{R}} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

**THEOREM 3.1 [Arratia, Goldstein and Gordon (1990)].** *Let  $W = \sum_{\alpha \in A} Y_\alpha$  be a sum of Bernoulli random variables where  $A$  is the index set and  $\mathbb{P}(Y_\alpha = 1) = 1 - \mathbb{P}(Y_\alpha = 0) = p_\alpha$ . Let  $\lambda = \sum_{\alpha \in A} p_\alpha$ , and let  $\text{Poi}(\lambda)$  denote the Poisson distribution with mean  $\lambda$ . Then*

$$(3.1) \quad d_{TV}(\mathcal{L}(W), \text{Poi}(\lambda)) \leq \left(1 \wedge \frac{1}{\lambda}\right)(b_1 + b_2 + b_3),$$

where for each  $\alpha$  and  $B_\alpha$  such that  $\alpha \in B_\alpha \subset A$ ,

$$(3.2) \quad \begin{aligned} b_1 &:= \sum_{\alpha \in A} \sum_{\beta \in B_\alpha} p_\alpha p_\beta, \\ b_2 &:= \sum_{\alpha \in A} \sum_{\alpha \neq \beta \in B_\alpha} \mathbb{E}(Y_\alpha Y_\beta), \\ b_3 &:= \sum_{\alpha \in A} \mathbb{E}|\mathbb{E}(Y_\alpha - p_\alpha | \sigma(Y_\beta : \beta \notin B_\alpha))|. \end{aligned}$$

**REMARK 3.1.** If  $B_\alpha$  is chosen such that  $X_\alpha$  is independent of  $\{X_\beta : \beta \notin B_\alpha\}$ , then  $b_3$  in (3.1) equals 0. Roughly speaking, in order for  $b_1$  and  $b_2$  to be small, the size of  $B_\alpha$  has to be small and  $\mathbb{E}(Y_\beta | Y_\alpha = 1) = o(1)$  for  $\alpha \neq \beta \in B_\alpha$ .

**3.1. Proof of Theorem 2.1.** In this subsection, let  $C$  and  $c$  denote positive constants depending only on the exponential family (2.1),  $\mu_0$  and  $a$ . They may represent different values in different expressions. The lemmas used in the proof of Theorem 2.1 will be stated and proved after the proof of the theorem.

**PROOF OF THEOREM 2.1.** By the union bound and Lemma 3.1, we have

$$(3.3) \quad \begin{aligned} \mathbb{P}(M_{n;t} \geq at) &\leq (n - t + 1)\mathbb{P}(X_1 + \dots + X_t \geq at) \\ &\sim (n - t + 1)e^{-(a\theta_a - \Psi(\theta_a))t} / t^{1/2}, \end{aligned}$$

where  $x \sim y$  means that  $x/y$  is bounded away from zero and infinity. On the other hand, by the definition of  $\lambda$  in (2.5) and (2.6), we have

$$(3.4) \quad \lambda \sim (n - t + 1)e^{-(a\theta_a - \Psi(\theta_a))t} / t^{1/2}.$$

From (3.3) and (3.4), if  $t$  or  $n - t$  is bounded, then the bound (2.4) holds true by choosing  $C$  in (2.4) to be large enough. Therefore, in the sequel, we can assume  $t$  and  $n - t$  to be larger than any given constant.

Let  $\delta$  be a positive number such that

$$(3.5) \quad 0 < \delta < 1 \wedge (a - \mu_0)/4 \quad \text{and} \quad \Psi(\theta_a) - (\mu_0 + \delta)\theta_a > 0.$$

The second inequality above is possible because of the strict convexity of  $\Psi$ . Let

$$(3.6) \quad m = \lfloor C(\log t \wedge \log(n - t)) \rfloor,$$

where the constant  $C$  will be chosen later in (3.17). By Lemma 3.2, we can find  $\theta_{a_1} \in \Theta$  such that  $\theta_a < \theta_{a_1}$ ,  $\theta_{a_1} \leq \theta_{a'}$  for case 1 and for  $t$  and  $n - t$  larger than some unspecified constant, the following bound holds uniformly in  $a_2 \in [a, a_1]$ :

$$(3.7) \quad d_{TV}(\mathcal{L}(\tilde{X}_i^{a_2} : 1 \leq i \leq m), \mathcal{L}(X_i^{a_2} : 1 \leq i \leq m)) \leq C \frac{m}{t},$$

where  $\{\tilde{X}_1^{a_2}, \dots, \tilde{X}_m^{a_2}\}$  is distributed as the conditional distribution of  $\{X_1, \dots, X_m\}$  given  $X_1 + \dots + X_t = a_2 t$ , and  $\{X_1^{a_2}, \dots, X_m^{a_2}\}$  are independent, identically distributed as  $F_{\theta_{a_2}}$ . In the following, we fix such an  $a_1$  and assume  $t$  and  $n - t$  to be large enough so that the bound (3.7) holds and

$$(3.8) \quad m < (a_1 - a)t/\delta.$$

We embed the sequence  $\{X_1, \dots, X_n\}$  into an infinite i.i.d. sequence

$$\{\dots, X_{-1}, X_0, X_1, \dots\}.$$

For each integer  $\alpha$ , let

$$(3.9) \quad T_\alpha = X_\alpha + \dots + X_{\alpha+t-1}, \quad \tilde{Y}_\alpha = I(T_\alpha \geq at).$$

To avoid the clumping of 1's in the sequence  $(\tilde{Y}_\alpha)$  which makes a Poisson approximation invalid, we define

$$(3.10) \quad Y_\alpha = I(T_\alpha \geq at, T_{\alpha-1} < at, \dots, T_{\alpha-m} < at).$$

Let

$$(3.11) \quad W = \sum_{\alpha=1}^{n-t+1} Y_\alpha, \quad \lambda_1 = \mathbb{E}W = (n - t + 1)\mathbb{E}Y_1.$$

In the following, we first bound  $|\mathbb{P}(M_{n;t} \geq at) - \mathbb{P}(W \geq 1)|$ , then bound the total variation distance between the distribution of  $W$  and  $\text{Poi}(\lambda_1)$ , finally we bound  $|\lambda_1 - \lambda|$ .

First, since  $\{M_{n;t} \geq at\} \setminus \{W \geq 1\} \subset \bigcup_{\alpha=1}^m \{T_\alpha \geq at\}$ , we have

$$(3.12) \quad 0 \leq \mathbb{P}(M_{n;t} \geq at) - \mathbb{P}(W \geq 1) \leq m\mathbb{P}(T_1 \geq at).$$

Next, by applying Theorem 3.1, we prove in Lemma 3.3 that

$$(3.13) \quad \begin{aligned} & |\mathbb{P}(W \geq 1) - (1 - e^{-\lambda_1})| \\ & \leq C \left(1 \wedge \frac{1}{\lambda_1}\right) (n - t + 1) \mathbb{P}(T_1 \geq at) [t\mathbb{P}(T_1 \geq at) + e^{-cm}], \end{aligned}$$

where the constant  $c$  does not depend on the choice of the constant  $C$  in (3.6), as can be seen from the proof of Lemma 3.3. Since  $\lambda_1$  does not have an explicit expression, our final goal is to show that  $\lambda_1$  is close to  $\lambda$ , which can be calculated explicitly as discussed in Remark 2.3. For this purpose, we first introduce an intermediate quantity  $\lambda_2$  defined as

$$(3.14) \quad \lambda_2 = (n - t + 1) \int_{at}^\infty \mathbb{P}(D_i > s - at, i \geq 1) d\mathbb{P}(T_1 \leq s).$$

Lemma 3.4 shows that

$$(3.15) \quad |\lambda_1 - \lambda_2| \leq C(n - t) \left[\frac{m^2}{t} + e^{-cm}\right] \mathbb{P}(T_1 \geq at),$$

and Lemma 3.5 shows that

$$(3.16) \quad \lambda_2 = \lambda \left(1 + O\left(\frac{(\log t)^2}{t}\right)\right).$$

Again, from the proof of Lemma 3.4, the constant  $c$  in (3.15) does not depend on the choice of the constant  $C$  in (3.6). Let the constant  $C$  in (3.6) be chosen such that

$$(3.17) \quad e^{-cm} = O\left(\frac{1}{t} \vee \frac{1}{n - t}\right)$$

for the constants  $c$  in (3.13) and (3.15). By Lemma 3.1 and (3.4), we have

$$(3.18) \quad \lambda \sim (n - t) \mathbb{P}(T_1 \geq at).$$

This, together with (3.15), (3.16) and (3.6), implies

$$(3.19) \quad |\lambda_1 - \lambda| \leq C\lambda \left[\frac{(\log t)^2}{t} + \frac{1}{n - t}\right] \quad \text{and} \quad \lambda_1 \sim \lambda.$$

By (3.12) and (3.13), we have

$$\begin{aligned} & |\mathbb{P}(M_{n;t} \geq at) - (1 - e^{-\lambda_1})| \\ & \leq C \left(1 \wedge \frac{1}{\lambda_1}\right) (n - t) \mathbb{P}(T_1 \geq at) \left[t\mathbb{P}(T_1 \geq at) + e^{-cm} + \frac{m}{n - t}\right], \end{aligned}$$

where the constant  $c$  is the same as that in (3.13). By (3.18), (3.19), (3.3), (3.17) and (3.6), this is further bounded by

$$(3.20) \quad |\mathbb{P}(M_{n;t} \geq at) - (1 - e^{-\lambda_1})| \leq C(\lambda \wedge 1) \left[ \frac{1}{t} + \frac{\log t \wedge \log(n-t)}{n-t} \right].$$

The bound (2.4) is proved by using (3.19) and (3.20) for the cases  $\lambda \leq 1$  and  $\lambda > 1$  separately and using  $|e^{-\lambda} - e^{-\lambda_1}| \leq |\lambda - \lambda_1|e^{-(\lambda \wedge \lambda_1)}$ .  $\square$

The following lemmas have been used in the above proof.

LEMMA 3.1 [Theorem 1 and Theorem 6 of Petrov (1965)]. *Under the setting of Theorem 2.1, we have*

$$\mathbb{P}(X_1 + \dots + X_t \geq at) \sim e^{-(a\theta_a - \Psi(\theta_a))t} / t^{1/2}.$$

LEMMA 3.2. *Under the setting of Theorem 2.1, there exists  $\theta_{a_1} \in \Theta$  such that  $\theta_a < \theta_{a_1}$ ,  $\theta_{a_1} \leq \theta_{a'}$  for case 1 and the bound (3.7) holds uniformly in  $a_2 \in [a, a_1]$  and in  $m$  and  $t$  such that  $m, t$  and  $t/m$  are larger than some unspecified constant.*

PROOF. For case 1, by (2.2), we have  $|\varphi_{\theta_a}(x)| < 1$  for  $x \neq 0$  and  $|\varphi_{\theta_a}(x)| \rightarrow 0$  as  $|x| \rightarrow \infty$ . Therefore, there exists  $M > 0$  such that  $|\varphi_{\theta_a}(x)| < 1/2$  for  $|x| > M$ . By the dominated convergence theorem,

$$|\varphi_{\theta_{a+h}}(x) - \varphi_{\theta_a}(x)| \rightarrow 0 \quad \text{as } h \rightarrow 0^+ \text{ uniformly in } x.$$

This, together with the continuity of the function  $\varphi_{\theta}(\cdot)$ , implies that there exists  $a_1 \leq a'$  such that

$$(3.21) \quad \sup_{\theta_a \leq \theta \leq \theta_{a_1}} \sup_{|x| > \varepsilon} |\varphi_{\theta}(x)| < 1 \quad \text{for all } \varepsilon > 0.$$

We now show that with such choice of  $a_1$ , (3.7) is satisfied. We follow the proof of Theorem 1.6 of Diaconis and Freedman (1988). Since only the range of parameters  $[a, a_1]$  enters into considerations, we do not need their Condition 1.1. By (2.2) and (3.21), their Conditions 1.2–1.4 are satisfied for the range of parameters  $[a, a_1]$ . Following their proof of Theorem 1.6, (3.7) holds uniformly in  $a_2 \in [a, a_1]$  and in  $m$  and  $t$  such that  $m, t$  and  $t/m$  are larger than some unspecified constant. The bound (3.7) for case 2 can be proved by rewriting the proof (e.g., changing density functions to probability mass functions) for case 1. As mentioned in Diaconis and Freedman (1988), the proof for the discrete case is a little easier. Therefore, we omit the details here.  $\square$

LEMMA 3.3. *Under the setting of Theorem 2.1, we have*

$$\begin{aligned} & |\mathbb{P}(W \geq 1) - (1 - e^{-\lambda_1})| \\ & \leq C \left( 1 \wedge \frac{1}{\lambda_1} \right) (n - t + 1) \mathbb{P}(T_1 \geq at) [t \mathbb{P}(T_1 \geq at) + e^{-cm}], \end{aligned}$$

where  $W$  and  $\lambda_1$  are defined in (3.11),  $T_1$  is defined in (3.9), and  $m$  is defined in (3.6).

PROOF. We apply Theorem 3.1 to bound the total variation distance between the distribution of  $W$  and  $\text{Poi}(\lambda_1)$ . For each  $1 \leq \alpha \leq n - t + 1$ , define  $B_\alpha = \{1 \leq \beta \leq n - t + 1 : |\alpha - \beta| < t + m\}$ . By definition of  $B_\alpha$ ,  $Y_\alpha$  in (3.10) is independent of  $\{Y_\beta : \beta \notin B_\alpha\}$ . Therefore,  $b_3$  in (3.2) equals zero. Since  $|B_\alpha| < 2(t + m)$ ,

$$b_1 = \sum_{1 \leq \alpha \leq n-t+1} \sum_{\beta \in B_\alpha} \mathbb{E}Y_\alpha \mathbb{E}Y_\beta < 2(t + m)\lambda_1 \mathbb{E}Y_1.$$

By the definition of  $Y_\alpha$ , for  $1 \leq |\beta - \alpha| \leq m$ ,  $\mathbb{E}Y_\alpha Y_\beta = 0$ , and for  $m < |\beta - \alpha| < t + m$ ,  $\mathbb{E}Y_\alpha Y_\beta \leq \mathbb{E}\tilde{Y}_{\alpha \wedge \beta} \tilde{Y}_{\alpha \vee \beta}$  where  $\tilde{Y}_\alpha$  is defined in (3.9). Therefore, by symmetry,

$$b_2 = \sum_{1 \leq \alpha \leq n-t+1} \sum_{\alpha \neq \beta \in B_\alpha} \mathbb{E}Y_\alpha Y_\beta < 2(n - t + 1)\mathbb{E}\tilde{Y}_1 \sum_{\beta=m+2}^{m+t} \mathbb{P}(T_\beta \geq at | T_1 \geq at).$$

For  $\beta \geq t + 1$ , by independence,

$$\mathbb{P}(T_\beta \geq at | T_1 \geq at) = \mathbb{P}(T_1 \geq at).$$

Let  $\delta$  be the positive number defined above (3.5) such that (3.5) is satisfied, and let  $a_1$  be as in Lemma 3.2. We observe that for  $m + 2 \leq \beta \leq t$ ,  $T_\beta \geq at$  and  $X_{t+1} + \dots + X_{t+\beta-1} \leq (\mu_0 + \delta)(\beta - 1)$  together imply  $X_\beta + \dots + X_t \geq at - (\mu_0 + \delta)(\beta - 1)$ . Therefore,

$$\begin{aligned} & \sum_{\beta=m+2}^t \mathbb{P}(T_\beta \geq at | T_1 \geq at) \\ & \leq \sum_{\beta=m+2}^t \{ \mathbb{P}(X_{t+1} + \dots + X_{t+\beta-1} > (\mu_0 + \delta)(\beta - 1)) \\ & \quad + \mathbb{P}(X_\beta + \dots + X_t \geq at - (\mu_0 + \delta)(\beta - 1) | T_1 \geq at) \}. \end{aligned}$$

For the first term, we have

$$\begin{aligned} & \sum_{\beta=m+2}^t \mathbb{P}(X_{t+1} + \dots + X_{t+\beta-1} > (\mu_0 + \delta)(\beta - 1)) \\ (3.22) \quad & \leq \sum_{\beta=m+2}^t e^{-[\theta_{\mu_0+\delta}(\mu_0+\delta) - \Psi(\theta_{\mu_0+\delta})](\beta-1)} \\ & \leq \frac{e^{-[\theta_{\mu_0+\delta}(\mu_0+2\delta) - \Psi(\theta_{\mu_0+\delta})]m}}{1 - e^{-[\theta_{\mu_0+\delta}(\mu_0+2\delta) - \Psi(\theta_{\mu_0+\delta})]}}. \end{aligned}$$

By the bound on  $V$  on page 613 of [Komlós and Tusnády \(1975\)](#) and recalling that we have chosen  $\delta$  such that  $\Psi(\theta_a) - (\mu_0 + \delta)\theta_a > 0$ , we have

$$\begin{aligned}
 & \sum_{\beta=m+2}^t \mathbb{P}(X_\beta + \dots + X_t \geq at - (\mu_0 + \delta)(\beta - 1) | T_1 \geq at) \\
 (3.23) \quad & \leq C \sum_{\beta=m+2}^t e^{-[\Psi(\theta_a) - (\mu_0 + \delta)\theta_a](\beta - 1)} \sqrt{\frac{t}{t - \beta + 1}} \\
 & \leq C \frac{e^{-[\Psi(\theta_a) - (\mu_0 + \delta)\theta_a]m}}{(1 - e^{-[\Psi(\theta_a) - (\mu_0 + \delta)\theta_a]})}.
 \end{aligned}$$

Therefore,

$$b_2 \leq C(n - t + 1)\mathbb{P}(T_1 \geq at)[m\mathbb{P}(T_1 \geq at) + e^{-cm}].$$

Lemma 3.3 is then followed by Theorem 3.1 and the above bounds on  $b_1$  and  $b_2$ . □

LEMMA 3.4. *Under the setting of Theorem 2.1, we have*

$$|\lambda_1 - \lambda_2| \leq C(n - t) \left[ \frac{m^2}{t} + e^{-cm} \right] \mathbb{P}(T_1 \geq at),$$

where  $\lambda_1$  and  $\lambda_2$  are defined in (3.11) and (3.14),  $m$  is defined in (3.6) and satisfies (3.8), and  $T_1$  is defined in (3.9).

PROOF. By symmetry, we can write

$$\begin{aligned}
 & \mathbb{E}Y_1 = \mathbb{I}(T_1 \geq at, T_2 < at, \dots, T_{m+1} < at) \\
 & = \mathbb{E}\tilde{Y}_1(1 - \tilde{Y}_2) \dots (1 - \tilde{Y}_{m+1}) \\
 (3.24) \quad & \leq \int_{at}^{at+m\delta} \mathbb{E}[(1 - \tilde{Y}_2) \dots (1 - \tilde{Y}_{m+1}) | T_1 = s] d\mathbb{P}(T_1 \leq s) \\
 & \quad + \mathbb{P}(T_1 > at + m\delta),
 \end{aligned}$$

where  $\tilde{Y}_\alpha$  is defined in (3.9) and  $\delta$  is the positive number defined above (3.5) such that (3.5) is satisfied. Observe that  $T_1 = s$  and  $T_{i+1} < at$  imply  $T_1 - T_{i+1} = S_i - (S_{i+t} - S_t) > s - at$  where  $S_i = \sum_{j=1}^i X_j$ . Therefore, given  $T_1 = s$ ,  $(1 - \tilde{Y}_2) \dots (1 - \tilde{Y}_{m+1})$  is the indicator of the event that  $\{\tilde{S}_i^{s/t} - S_i > s - at, 1 \leq i \leq m\}$  where  $\tilde{S}_i^{s/t}$  is independent of  $S_i$ ,

$$\tilde{S}_i^{s/t} = \sum_{j=1}^i \tilde{X}_j^{s/t} \quad \text{and} \quad \mathcal{L}(\tilde{X}_i^{s/t} : 1 \leq i \leq m) = \mathcal{L}(X_i : 1 \leq i \leq m | S_t = s).$$

Note that the assumption  $m < (a_1 - a)t/\delta$  in (3.8) implies  $a \leq s/t \leq a_1$  for  $s \in [at, at + m\delta]$ .

By the definition of total variation distance, for  $at \leq s \leq at + m\delta$  and  $m \geq \nu$ ,

$$\begin{aligned}
 & d_{TV}(\mathcal{L}(X_i^{s/t} : 1 \leq i \leq m), \mathcal{L}(X_i^a : 1 \leq i \leq m)) \\
 (3.25) \quad & \leq \mathbb{E}_{\theta_{s/t}} I(S_m > t/m) + \mathbb{E}_{\theta_a} I(S_m > t/m) \\
 & \quad + \mathbb{E}_{\theta_a} |e^{(\theta_{s/t} - \theta_a)S_m - m(\Psi(\theta_{s/t}) - \Psi(\theta_a))} - 1| I(S_m \leq t/m).
 \end{aligned}$$

For  $s/t \in [a, a_1]$  and  $s/t - a \leq m\delta/t$ , we have

$$\begin{aligned}
 (3.26) \quad & |\theta_{s/t} - \theta_a| \leq \sup_{\theta_a \leq \theta \leq \theta_{a_1}} \frac{1}{\Psi''(\theta)}(s/t - a) \leq \sup_{\theta_a \leq \theta \leq \theta_{a_1}} \frac{m\delta}{\Psi''(\theta)t}, \\
 & |\Psi(\theta_{s/t}) - \Psi(\theta_a)| \leq \sup_{\theta_a \leq \theta \leq \theta_{a_1}} |\Psi'(\theta)|(s/t - a) \leq \sup_{\theta_a \leq \theta \leq \theta_{a_1}} |\Psi'(\theta)| \frac{m\delta}{t}.
 \end{aligned}$$

This implies that if  $a \leq s/t \leq a + m\delta/t$ ,  $S_m \leq t/m$  and  $m \leq \sqrt{t}$ , then

$$(3.27) \quad (\theta_{s/t} - \theta_a)S_m - m(\Psi(\theta_{s/t}) - \Psi(\theta_a)) \leq C.$$

By (3.25), Markov’s inequality, (3.27) and the fact that  $|e^t - 1| \leq Ct$  for bounded  $t$ , we have, for  $at \leq s \leq at + m\delta$ ,

$$\begin{aligned}
 (3.28) \quad & d_{TV}(\mathcal{L}(X_i^{s/t} : 1 \leq i \leq m), \mathcal{L}(X_i^a : 1 \leq i \leq m)) \\
 & \leq \frac{m}{t} (\mathbb{E}_{\theta_{s/t}} |S_m| + \mathbb{E}_{\theta_a} |S_m|) \\
 & \quad + C\mathbb{E}_{\theta_a} |(\theta_{s/t} - \theta_a)S_m - m(\Psi(\theta_{s/t}) - \Psi(\theta_a))| \\
 & \leq Cm^2/t,
 \end{aligned}$$

where in the last inequality we used  $s/t \in [a, a_1]$ ,  $\mathbb{E}_\theta |S_m| \leq Cm$  for  $\theta \in [\theta_a, \theta_{a_1}]$  and (3.26). By (3.24) and the argument just below it, (3.7) and (3.28), we have

$$\begin{aligned}
 (3.29) \quad \mathbb{E}Y_1 & \leq \int_{at}^{at+m\delta} \left[ \mathbb{P}(D_i > s - at, 1 \leq i \leq m) + C \frac{m^2}{t} \right] d\mathbb{P}(T_1 \leq s) \\
 & \quad + \mathbb{P}(T_1 > at + m\delta).
 \end{aligned}$$

By (3.5),  $at \leq s \leq at + m\delta$  implies

$$(3.30) \quad s - at - m(a - \mu_0)/2 < m(\mu_0 - a)/4.$$

From the FKG inequality [cf. (1.7) of Karlin and Rinott (1980)] and the fact that  $I(D_i > s - at, 1 \leq i \leq m)$  and  $I(D_i > s - at, m_1 \geq i > m)$  are both increasing



functions of  $\{X_1^a - X_1, \dots, X_{m_1}^a - X_{m_1}\}$  for  $m_1 > m$ , we have

$$\begin{aligned}
 & \mathbb{P}(D_i > s - at, m_1 \geq i \geq 1) \\
 &= \mathbb{P}(D_i > s - at, 1 \leq i \leq m) \mathbb{P}(D_i > s - at, m_1 \geq i > m | \\
 & \quad D_i > s - at, 1 \leq i \leq m) \\
 (3.31) \quad & \geq \mathbb{P}(D_i > s - at, 1 \leq i \leq m) \mathbb{P}(D_i > s - at, m_1 \geq i > m) \\
 & \geq \mathbb{P}(D_i > s - at, 1 \leq i \leq m) \mathbb{P}(D_i > s - at, i > m, D_m \geq m(a - \mu_0)/2) \\
 & \geq \mathbb{P}(D_i > s - at, 1 \leq i \leq m) \\
 & \quad \times \left\{ 1 - \mathbb{P}(D_m < m(a - \mu_0)/2) - \sum_{i=1}^{\infty} \mathbb{P}(D_i < m(\mu_0 - a)/4) \right\},
 \end{aligned}$$

where the last inequality in (3.31) follows from (3.30). Letting  $m_1 \rightarrow \infty$  in (3.31), we have

$$\begin{aligned}
 & \mathbb{P}(D_i > s - at, i \geq 1) \\
 (3.32) \quad & \geq \mathbb{P}(D_i > s - at, 1 \leq i \leq m) \\
 & \quad \times \left\{ 1 - \mathbb{P}(D_m < m(a - \mu_0)/2) - \sum_{i=1}^{\infty} \mathbb{P}(D_i < m(\mu_0 - a)/4) \right\}.
 \end{aligned}$$

For  $0 < r \leq \theta_a$ , we have

$$\mathbb{E} \exp(-r D_i) = \exp\{-i[\Psi(\theta_a) - \Psi(r) - \Psi(\theta_a - r)]\}.$$

By Taylor's expansion,

$$\Psi(\theta_a) - \Psi(\theta_a - r) = ra - \frac{r^2}{2} \Psi''(\theta_a - r_1), \quad -\Psi(r) = -r\mu_0 - \frac{r^2}{2} \Psi''(r_2),$$

where  $0 \leq r_1, r_2 \leq r$ . Therefore,

$$\begin{aligned}
 & \mathbb{P}(D_m < m(a - \mu_0)/2) \\
 (3.33) \quad & \leq \exp\left\{-m\left[\Psi(\theta_a) - \Psi(r) - \Psi(\theta_a - r) - \frac{(a - \mu_0)r}{2}\right]\right\} \\
 & = \exp\left\{-m\left[\frac{r}{2}(a - \mu_0) - \frac{r^2}{2}(\Psi''(\theta_a - r_1) + \Psi''(r_2))\right]\right\}.
 \end{aligned}$$

Let

$$c_1 = \frac{a - \mu_0}{4\theta_a} \vee \max_{\theta \in \Theta: 0 \leq \theta \leq \theta_a} \Psi''(\theta).$$

Choosing  $r = (a - \mu_0)/(4c_1)$  in (3.33), we have

$$(3.34) \quad \mathbb{P}(D_m < m(a - \mu_0)/2) \leq \exp\left\{-\frac{(a - \mu_0)^2}{16c_1} m\right\}.$$

Similarly,

$$(3.35) \quad \mathbb{P}(D_i < m(\mu_0 - a)/4) \leq \exp\left\{-\frac{(a - \mu_0)^2}{16c_1}i - \frac{(a - \mu_0)^2}{16c_1}m\right\}.$$

Applying (3.34) and (3.35) in (3.32), we obtain

$$\begin{aligned} &\mathbb{P}(D_i > s - at, 1 \leq i \leq m) \\ &\leq \mathbb{P}(D_i > s - at, i \geq 1) \\ &\quad + 2 \exp\left\{-\frac{(a - \mu_0)^2}{16c_1}m\right\} / \left(1 - \exp\left\{-\frac{(a - \mu_0)^2}{16c_1}\right\}\right). \end{aligned}$$

Therefore, by (3.29),

$$(3.36) \quad \begin{aligned} &\mathbb{E}Y_1 - \lambda_2/(n - t + 1) \\ &\leq \left\{ \frac{2 \exp\{-m(a - \mu_0)^2/(16c_1)\}}{1 - \exp\{-(a - \mu_0)^2/(16c_1)\}} \right. \\ &\quad \left. + C \frac{m^2}{t} + \frac{\mathbb{P}(T_1 > at + m\delta)}{\mathbb{P}(T_1 \geq at)} \right\} \mathbb{P}(T_1 \geq at). \end{aligned}$$

From the corollary on page 611 of Komlós and Tusnády (1975), and recalling that in proving (2.4), we can only consider those  $t$  larger than any given constant, we have

$$(3.37) \quad \mathbb{P}(T_1 > at + m\delta | T_1 \geq at) \leq Ce^{-\theta_a m\delta}.$$

After proving a similar and easier lower bound of  $\mathbb{E}Y_1$ , we obtain Lemma 3.4.  $\square$

LEMMA 3.5. *Under the setting of Theorem 2.1, we have*

$$\lambda_2 = \lambda \left( 1 + O\left(\frac{(\log t)^2}{t}\right) \right),$$

where  $\lambda_2$  is define in (3.14).

PROOF. We first consider case 1 of Theorem 2.1. By the proof of Theorem 2.7 of Woodroffe (1982), we have for  $x \geq 0$ ,

$$(3.38) \quad \mathbb{P}(D_i > x, i \geq 1) = \frac{\mathbb{P}(D_{\tau_+} > x)}{\mathbb{E}\tau_+},$$

where  $\tau_+ = \inf\{i \geq 1, D_i > 0\}$ . Let  $x_0 = \log t/\theta_a$ . By change of variable and (2.1), we have

$$(3.39) \quad \begin{aligned} \lambda_2 &= (n - t + 1) \int_0^\infty \mathbb{P}(D_i > x, i \geq 1) d\mathbb{P}(T_1 \leq at + x) \\ &= (n - t + 1)e^{-(a\theta_a - \Psi(\theta_a))t} \end{aligned}$$

$$\begin{aligned} &\times \int_0^{x_0} \mathbb{P}(D_i > x, i \geq 1) e^{-\theta_a x} d\mathbb{P}_{\theta_a}(T_1 \leq at + x) \\ &+ O((n - t + 1)\mathbb{P}(T_1 > at + x_0)), \end{aligned}$$

where  $T_\alpha$  is defined in (3.9). By the local central limit theorem [cf. Feller (1971)], uniformly for  $0 \leq x \leq x_0$ ,

$$(3.40) \quad d\mathbb{P}_{\theta_a}(T_1 \leq at + x) = \frac{1}{\sigma_a(2\pi t)^{1/2}} + O\left(\frac{(\log t)^2}{t^{3/2}}\right).$$

From (3.37) and Lemma 3.1, we have

$$(3.41) \quad \begin{aligned} \mathbb{P}(T_1 > at + x_0) &= \mathbb{P}(T_1 > at + x_0 | T_1 \geq at) \mathbb{P}(T_1 \geq at) \\ &\leq C e^{-\theta_a x_0} \frac{e^{-(a\theta_a - \Psi(\theta_a))t}}{\sqrt{t}} \leq C \frac{e^{-(a\theta_a - \Psi(\theta_a))t}}{\sqrt{t}} \frac{1}{t}. \end{aligned}$$

Applying (3.38), (3.40) and (3.41) in (3.39), we obtain

$$\begin{aligned} \lambda_2 &= \frac{(n - t + 1)e^{-(a\theta_a - \Psi(\theta_a))t}}{(\mathbb{E}\tau_+) \sigma_a(2\pi t)^{1/2}} \\ &\times \int_0^{x_0} \mathbb{P}(D_{\tau_+} > x) e^{-\theta_a x} \left(1 + O\left(\frac{(\log t)^2}{t}\right)\right) dx \\ &+ O((n - t + 1)\mathbb{P}(T_1 > at + x_0)) \\ &= \frac{(n - t + 1)e^{-(a\theta_a - \Psi(\theta_a))t}}{(\mathbb{E}\tau_+) \sigma_a(2\pi t)^{1/2}} \\ &\times \left[ \int_0^\infty \mathbb{P}(D_{\tau_+} > x) e^{-\theta_a x} \left(1 + O\left(\frac{(\log t)^2}{t}\right)\right) dx + O\left(\frac{1}{t}\right) \right]. \end{aligned}$$

By the integration by parts formula, we have

$$(3.42) \quad \begin{aligned} &\frac{1}{\mathbb{E}\tau_+} \int_0^\infty \mathbb{P}(D_{\tau_+} > x) e^{-\theta_a x} dx \\ &= \frac{1}{\theta_a \mathbb{E}\tau_+} [1 - \mathbb{E}e^{-\theta_a D_{\tau_+}}] \\ &= \frac{1}{\theta_a \mathbb{E}\tau_+} \exp\left[-\sum_{k=1}^\infty k^{-1} \mathbb{E}(e^{-\theta_a D_k}, D_k > 0)\right] \\ &= \frac{1}{\theta_a} \exp\left[-\sum_{k=1}^\infty k^{-1} \mathbb{E}(e^{-\theta_a D_k^+})\right], \end{aligned}$$

where we used the first equality in the proof of Corollaries 2.7 and 2.4 of Woodroffe (1982). Therefore,

$$\lambda_2 = \frac{(n - t + 1)e^{-(a\theta_a - \Psi(\theta_a))t}}{\theta_a \sigma_a (2\pi t)^{1/2}} \exp \left[ - \sum_{k=1}^{\infty} k^{-1} \mathbb{E}(e^{-\theta_a D_k^+}) \right] \left( 1 + O \left( \frac{(\log t)^2}{t} \right) \right).$$

Next, we consider case 2 of Theorem 2.1. The calculation of  $\lambda_2$  is similar to case 1 except that we have, for integers  $0 \leq k \leq x_0$ ,

$$\mathbb{P}_{\theta_a}(T_1 = \lceil at \rceil + k) = \frac{1}{\sigma_a (2\pi t)^{1/2}} + O \left( \frac{(\log t)^2}{t^{3/2}} \right)$$

and

$$\begin{aligned} & \sum_{k=0}^{\infty} \mathbb{P}(D_{\tau_+} > \lceil at \rceil - at + k) e^{-\theta_a(\lceil at \rceil - at + k)} \\ (3.43) \quad & = e^{-\theta_a(\lceil at \rceil - at)} \sum_{k=0}^{\infty} \mathbb{P}(D_{\tau_+} > k) e^{-\theta_a k} \\ & = \frac{e^{-\theta_a(\lceil at \rceil - at)}}{1 - e^{-\theta_a}} [1 - \mathbb{E}e^{-\theta_a D_{\tau_+}}]. \end{aligned}$$

Therefore, for the arithmetic case,

$$\begin{aligned} \lambda_2 & = \frac{(n - t + 1)e^{-(a\theta_a - \Psi(\theta_a))t} e^{-\theta_a(\lceil at \rceil - at)}}{(1 - e^{-\theta_a})\sigma_a (2\pi t)^{1/2}} \\ & \quad \times \exp \left[ - \sum_{k=1}^{\infty} k^{-1} \mathbb{E}(e^{-\theta_a D_k^+}) \right] \left( 1 + O \left( \frac{(\log t)^2}{t} \right) \right). \end{aligned} \quad \square$$

3.2. Proof of Corollary 2.1. Following the proof of Theorem 2.1, let

$$Y_1 = \mathbf{I}(X_1 = \dots = X_t = a)$$

and for  $2 \leq \alpha \leq n - t + 1$ ,

$$Y_\alpha = \mathbf{I}(X_{\alpha-1} < a, X_\alpha = \dots = X_{\alpha+t-1} = a).$$

Then with  $W = \sum_{\alpha=1}^{n-t+1} Y_\alpha$ ,

$$\mathbb{E}W = p_a^t + (n - t)p_a^t(1 - p_a) = \lambda.$$

Instead of (3.12), we have  $\mathbb{P}(M_{n;t} \geq at) = \mathbb{P}(W \geq 1)$ , and instead of (3.13), we have

$$|\mathbb{P}(W \geq 1) - (1 - e^{-\lambda})| \leq \left( 1 \wedge \frac{1}{\lambda} \right) (n - t + 1)(2t + 1)p_a^{2t}.$$

This proves the bound (2.13).

3.3. *Proof of Theorem 2.2.* In this subsection, let  $C$  and  $c$  denote positive constants depending on the exponential family  $F_\theta$  and  $\theta_1$ , and may represent different values in different expressions. The lemmas used in the proof of Theorem 2.2 will be stated and proved after the proof.

PROOF OF THEOREM 2.2. Recall  $S_i = \sum_{k=1}^i X_k$ . Define  $\tau_+ = \inf\{n \geq 1 : S_n > 0\}$  and

$$(3.44) \quad \tau_b := \inf\{n \geq 1 : S_n \geq b\}, \quad T_b := \inf\{n \geq 1 : S_n \notin [0, b)\}.$$

If  $b$  is bounded, then by choosing  $C$  to be large enough in (2.16), and observing that

$$(3.45) \quad \lambda \sim \left(n - \frac{b}{\mu_1}\right) e^{-\theta_1 b},$$

we have  $C\lambda b^{1/2}h(b)/(n - b/\mu_1) \geq 1$  and (2.16) is trivial. Therefore, in the following we can assume  $b$  is larger than any given constant. Moreover, since we assume  $h(b) = O(b^{1/2})$  in the theorem, by choosing  $C$  to be large enough and  $c$  to be small enough in (2.16), we only need to consider the case where  $h(b)/b^{1/2}$  is smaller than any given positive constant. In particular, we can assume

$$(3.46) \quad \frac{h(b)}{b^{1/2}} \leq \min\left\{\frac{2}{\mu_1}, \frac{2(\theta'_1 - \theta_1) \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)}{\mu_1^2}\right\}$$

for some  $\theta'_1 \in \Theta$  and  $\theta_1 < \theta'_1 < 2\theta_1$ .

We embed the sequence  $\{X_1, \dots, X_n\}$  into an infinite i.i.d. sequence

$$\{\dots, X_{-1}, X_0, X_1, \dots\}.$$

For a positive integer  $m$ , let  $\omega_m^+$  be the  $m$ -shifted sample path of  $\omega := \{X_1, \dots, X_n\}$ , so  $S_i(\omega_m^+) = S_{m+i}(\omega) - S_m(\omega)$ ,  $T_b(\omega_m^+) = \inf\{n \geq 1 : S_n(\omega_m^+) \notin [0, b)\}$ , and  $\tau_b(\omega_m^+)$ ,  $\tau_+(\omega_m^+)$  are defined similarly. Let  $t = \lceil \frac{b}{\mu_1} + b^{1/2}h(b) \rceil$  and  $m = \lfloor ch^2(b) \rfloor$  such that  $m < t$ . For  $1 \leq \alpha \leq n - t$ , let

$$(3.47) \quad Y_\alpha = \mathbb{I}(S_\alpha < S_{\alpha-\beta}, \forall 1 \leq \beta \leq m; T_b(\omega_\alpha^+) \leq t, S_{T_b(\omega_\alpha^+)} \geq b).$$

That is,  $Y_\alpha$  is the indicator of the event that the sequence  $\{S_i\}$  reaches a local minimum at  $\alpha$  and the  $\alpha$ -shifted sequence  $\{S_i(\omega_\alpha^+)\}$  exits the interval  $[0, b)$  within time  $t$  and the first exiting position is  $b$ . Let

$$(3.48) \quad W = \sum_{\alpha=1}^{n-t} Y_\alpha.$$

In the following, we first compare  $p_{n,b}$  with  $\mathbb{P}(W \geq 1)$ . Then we approximate the distribution of  $W$  by the Poisson distribution with mean  $\mathbb{E}(W)$ . Finally, we calculate approximately  $\mathbb{E}(W)$ .

First, from the definition of  $W$ , we have  $p_{n,b} \geq \mathbb{P}(W \geq 1)$  and with  $t_1 = \lfloor b/\mu_1 - b^{1/2}h(b) \rfloor$ ,

$$\begin{aligned} & \left\{ \max_{0 \leq i < j \leq n} (S_j - S_i) \geq b \right\} \setminus \{W \geq 1\} \\ & \subset \left( \bigcup_{k=0}^{n-t-1} \{S_{T_b}(\omega_k^+) \geq b, T_b(\omega_k^+) > t\} \right) \\ & \quad \cup \left( \bigcup_{k \in [0,m] \cup (n-t, n-t_1)} \{S_{T_b}(\omega_k^+) \geq b, T_b(\omega_k^+) \leq t\} \right) \\ & \quad \cup \left( \bigcup_{n-t_1 \leq i < j \leq n} \{S_j - S_i \geq b\} \right). \end{aligned}$$

By symmetry,

$$\begin{aligned} & p_{n,b} - \mathbb{P}(W \geq 1) \\ (3.49) \quad & \leq (n-t)\mathbb{P}(S_{T_b} \geq b, T_b > t) + (m + 2b^{1/2}h(b) + 2)\mathbb{P}(S_{T_b} \geq b, T_b \leq t) \\ & \quad + \mathbb{E} \mathbb{I} \left( \bigcup_{0 \leq i < j \leq t_1} \{S_j - S_i \geq b\} \right). \end{aligned}$$

By (3.60) and Lemma 3.7, we have

$$(3.50) \quad \mathbb{P}(S_{T_b} \geq b) = \mathbb{E}_{\theta_1} [e^{-\theta_1 S_{T_b}} \mathbb{I}(S_{T_b} \geq b)] \leq e^{-\theta_1 b}$$

and

$$\begin{aligned} (3.51) \quad & \mathbb{P}(S_{T_b} \geq b, T_b > t) = \mathbb{E}_{\theta_1} [e^{-\theta_1 S_{T_b}} \mathbb{I}(S_{T_b} \geq b, T_b > t)] \\ & \leq e^{-\theta_1 b} \mathbb{P}_{\theta_1}(T_b > t) \leq C e^{-\theta_1 b - ch^2(b)}. \end{aligned}$$

Along with Lemma 3.9, we have

$$\begin{aligned} (3.52) \quad & p_{n,b} - \mathbb{P}(W \geq 1) \\ & \leq C(n - b/\mu_1) \\ & \quad \times e^{-\theta_1 b} \left\{ e^{-ch^2(b)} + \frac{m + b^{1/2}h(b)}{n - b/\mu_1} + \frac{b/h^2(b)}{n - b/\mu_1} e^{-ch^2(b)} \right\}. \end{aligned}$$

Next, we use Theorem 3.1 to obtain a bound on the total variation distance between the distribution of  $W$  and  $\text{Poi}(\lambda_1)$  with

$$(3.53) \quad \lambda_1 := \mathbb{E}(W) = (n-t)\mathbb{E}Y_\alpha.$$

For each  $1 \leq \alpha \leq n-t$ , let  $B_\alpha = \{1 \leq \beta \leq n-t : |\beta - \alpha| \leq t+m\}$ . In applying Theorem 3.1, by our definition of  $B_\alpha$ ,  $b_3 = 0$ . From  $|B_\alpha| \leq 2(t+m) + 1$  and (3.50),

we have

$$(3.54) \quad \begin{aligned} b_1 < [2(t+m) + 1]\lambda_1 \mathbb{E}Y_\alpha \leq C(n-t)(t+m)\mathbb{P}^2(S_{T_b} \geq b) \\ \leq C(n-t)(t+m)e^{-2\theta_1 b}. \end{aligned}$$

Let

$$\tilde{Y}_\alpha = \mathbb{I}(T_b(\omega_\alpha^+) \leq t, S_{T_b}(\omega_\alpha^+) \geq b).$$

We have for  $b_2$  in (3.2)

$$\begin{aligned} b_2 &\leq \sum_{\alpha=1}^{n-t} \sum_{\alpha \neq \beta \in B_\alpha} \mathbb{E}(Y_\alpha Y_\beta) \\ &\leq 2 \sum_{\beta=1}^{n-t} \left[ \sum_{\beta-t-m \leq \alpha < \beta-m} \mathbb{E}Y_\beta \tilde{Y}_\alpha + \sum_{\beta-m \leq \alpha \leq \beta-1} \mathbb{E}Y_\beta \tilde{Y}_\alpha \right]. \end{aligned}$$

For  $\beta - t - m \leq \alpha < \beta - m$ , because  $S_\beta < S_\alpha$  implies  $T_b(\omega_\alpha^+) \leq \beta - \alpha$ , we have

$$\begin{aligned} \mathbb{E}Y_\beta \tilde{Y}_\alpha &= \mathbb{E}Y_\beta \tilde{Y}_\alpha [\mathbb{I}(S_\beta \geq S_\alpha) + \mathbb{I}(S_\beta < S_\alpha)] \\ &\leq \mathbb{E}\mathbb{I}(S_\beta - S_\alpha \geq 0) \tilde{Y}_\beta + \mathbb{E}\mathbb{I}(S_{T_b}(\omega_\alpha^+) \geq b, T_b(\omega_\alpha^+) \leq \beta - \alpha) \tilde{Y}_\beta. \end{aligned}$$

By independence and symmetry, we have

$$\sum_{\beta-t-m \leq \alpha < \beta-m} \mathbb{E}Y_\beta \tilde{Y}_\alpha \leq \mathbb{E}\tilde{Y}_1 \sum_{i=m}^{t+m} [\mathbb{P}(S_i \geq 0) + \mathbb{P}(S_{T_b} \geq b, T_b \leq t+m)].$$

For  $\beta - m \leq \alpha \leq \beta - 1$ , because  $Y_\beta = 1$  implies  $S_\alpha > S_\beta$ , which in turn implies  $T_b(\omega_\alpha^+) \leq \beta - \alpha$ , we have

$$\begin{aligned} \sum_{\beta-m \leq \alpha \leq \beta-1} \mathbb{E}Y_\beta \tilde{Y}_\alpha &\leq \sum_{\beta-m \leq \alpha \leq \beta-1} \mathbb{E}\tilde{Y}_\beta \mathbb{I}(S_{T_b}(\omega_\alpha^+) \geq b, T_b(\omega_\alpha^+) \leq \beta - \alpha) \\ &\leq \mathbb{E}\tilde{Y}_1 \sum_{i=1}^m \mathbb{P}(S_{T_b} \geq b, T_b \leq i). \end{aligned}$$

Therefore, by Lemma 3.8 and (3.50),

$$(3.55) \quad \begin{aligned} b_2 &\leq 2(n-t)\mathbb{E}\tilde{Y}_1 \left[ \sum_{i=m}^{t+m} (\mathbb{P}(S_i \geq 0) + \mathbb{P}(S_{T_b} \geq b, T_b \leq t+m)) \right. \\ &\quad \left. + \sum_{i=1}^m \mathbb{P}(S_{T_b} \geq b, T_b \leq i) \right] \\ &\leq 2(n-t)e^{-\theta_1 b} [C e^{-cm} + (t+m)e^{-\theta_1 b}]. \end{aligned}$$

From (3.1), (3.54) and (3.55), we obtain

$$(3.56) \quad |\mathbb{P}(W \geq 1) - (1 - e^{-\lambda_1})| \leq C(n - t)e^{-\theta_1 b}[(t + m)e^{-\theta_1 b} + e^{-cm}].$$

Since  $\lambda_1$  does not have an explicit expression, our final goal is to show that  $\lambda_1$  is close to  $\lambda$ . For this purpose, we first introduce an intermediate quantity  $\lambda_2$  defined as

$$(3.57) \quad \lambda_2 = (n - t)\mathbb{P}(\tau_0 = \infty)\mathbb{P}(S_{T_b} \geq b).$$

Recall

$$\begin{aligned} \lambda_1 &= (n - t)\mathbb{E}Y_\alpha \\ &= (n - t)\mathbb{P}(S_{\alpha-\beta} - S_\alpha > 0, \forall 1 \leq \beta \leq m)\mathbb{P}(T_b(\omega_\alpha^+) \leq t, S_{T_b}(\omega_\alpha^+) \geq b) \\ &= (n - t)\mathbb{P}(\tau_0 > m)\mathbb{P}(T_b \leq t, S_{T_b} \geq b). \end{aligned}$$

From the upper and lower bounds of their difference

$$\begin{aligned} \lambda_2 - \lambda_1 &\leq (n - t)\mathbb{P}(T_b > t, S_{T_b} \geq b), \\ \lambda_1 - \lambda_2 &\leq (n - t)\mathbb{P}(S_{T_b} \geq b)\mathbb{P}(m < \tau_0 < \infty), \end{aligned}$$

we have

$$(3.58) \quad \begin{aligned} |\lambda_1 - \lambda_2| &\leq C(n - t)e^{-\theta_1 b - ch^2(b)} + (n - t)e^{-\theta_1 b} \sum_{i=m}^{\infty} \mathbb{P}(S_i \geq 0) \\ &\leq C(n - t)e^{-\theta_1 b} [e^{-ch^2(b)} + e^{-cm}], \end{aligned}$$

where we used (3.51), (3.50) and Lemma 3.7.

Finally, in Lemma 3.11, we will show that

$$(3.59) \quad \lambda_2 = \left[ 1 + O\left(\frac{b^{1/2}h(b)}{n - b/\mu_1}\right) \right] \lambda + (n - t)e^{-\theta_1 b} o(e^{-cb}).$$

Theorem 2.2 is proved by combining (3.52), (3.56), (3.58) and (3.59) and using (3.45).  $\square$

The following lemmas have been used in the above proof.

**LEMMA 3.6.** *Let  $\{X_1, \dots, X_n\}$  be independent, identically distributed random variables with distribution function  $F$  that can be imbedded in an exponential family, as in (2.1). Let  $\mathbb{E}X_1 = \mu_0 < 0$ . Let  $S_0 = 0$  and  $S_i = \sum_{k=1}^i X_k$  for  $1 \leq i \leq n$ . Suppose there exist  $\theta_1 > 0$  such that  $\Psi(\theta_1) = 0$ . Let  $\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}$ , and let  $T$  be a stopping time with respect to  $\{\mathcal{F}_n\}$ . Then we have*

$$(3.60) \quad \mathbb{P}(G \cap \{T < \infty\}) = \mathbb{E}_{\theta_1} [e^{-\theta_1 S_T} \mathbf{I}(G \cap \{T < \infty\})]$$

for any  $G \in \mathcal{F}_T$ .



PROOF. Equation (3.60) follows by a direct application of Wald’s likelihood ratio identity [cf. Theorem 1.1 of Woodroofe (1982)] to the sequence  $\{X_1, X_2, \dots\}$ .  $\square$

LEMMA 3.7. Under the setting of Theorem 2.2, let  $t = \lceil \frac{b}{\mu_1} + b^{1/2}h(b) \rceil$ . We have

$$\mathbb{P}_{\theta_1}(T_b > t) \leq C e^{-ch^2(b)},$$

where  $T_b$  is defined in (3.44).

PROOF. Let

$$r = \frac{\mu_1^2}{2 \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)} h(b)/b^{1/2}.$$

By (3.46), we have  $r < \theta_1$  and

$$(3.61) \quad \mu_1 r - \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)r^2/2 \geq \mu_1 r/2.$$

Form  $\{T_b > t\} \subset \{S_t \leq b\}$  and Markov’s inequality, we have

$$\mathbb{P}_{\theta_1}(T_b > t) \leq \mathbb{P}_{\theta_1}(S_t \leq b) \leq e^{rb} \mathbb{E}_{\theta_1} e^{-rS_t}.$$

This is further bounded by

$$\begin{aligned} &\mathbb{P}_{\theta_1}(T_b > t) \\ &\leq \exp\{rb - [\Psi(\theta_1) - \Psi(\theta_1 - r)]t\} \quad (\text{by direct computation}) \\ &\leq \exp\left\{rb - \left[\mu_1 r - \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)r^2/2\right]t\right\} \quad (\text{by Taylor’s expansion}) \\ &\leq \exp\left\{\frac{\sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)r^2b}{2\mu_1} - \left[\mu_1 r - \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)r^2/2\right]b^{1/2}h(b)\right\} \\ &\hspace{15em} (\text{from the definition of } t) \\ &\leq \exp\left\{\frac{\sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)r^2b}{2\mu_1} - \frac{\mu_1 r}{2}b^{1/2}h(b)\right\} \quad [\text{from (3.61)}] \\ &\leq \exp\left\{-\frac{\mu_1^3}{8 \sup_{0 \leq \theta \leq \theta_1} \Psi''(\theta)}h^2(b)\right\} \quad (\text{from the definition of } r). \end{aligned}$$

This proves Lemma 3.7.  $\square$

LEMMA 3.8. Under the setting of Theorem 2.1, for positive integers  $m$ , we have

$$\sum_{i=m}^{\infty} \mathbb{P}(S_i \geq 0) \leq C e^{-cm}.$$

PROOF. Lemma 3.8 follows from

$$\mathbb{P}(S_i \geq 0) \leq \mathbb{E}e^{\theta^* S_i} = e^{\Psi(\theta^*)i},$$

where  $0 < \theta^* < \theta_1$  and  $\Psi(\theta^*) < 0$ .  $\square$

LEMMA 3.9. Under the setting of Theorem 2.2, let  $t_1 = \lfloor \frac{b}{\mu_1} - b^{1/2}h(b) \rfloor$ . We have

$$\mathbb{E}\mathbb{I}\left(\bigcup_{0 \leq i < j \leq t_1} \{S_j - S_i \geq b\}\right) \leq C e^{-\theta_1 b} \frac{b}{h^2(b)} e^{-ch^2(b)}.$$

PROOF. We only need to consider the case when  $t_1 > 0$ . Let

$$r = \frac{\mu_1^2}{2 \sup_{\theta_1 \leq \theta \leq \theta'_1} \Psi''(\theta)} h(b)/b^{1/2},$$

where  $\theta'_1$  is defined just below (3.46). By (3.46),  $\theta_1 + r \leq \theta'_1 \in \Theta$ . By (2.1), we have

$$\begin{aligned} \mathbb{P}_{\theta_1}(S_j \geq b) &= \mathbb{E}_{\theta_1+r}\{e^{-rS_j+j[\Psi(\theta_1+r)-\Psi(\theta_1)]}I(S_j \geq b)\} \\ &\leq \exp\{j[\Psi(\theta_1+r) - \Psi(\theta_1)] - rb\}. \end{aligned}$$

Thus,

$$\begin{aligned} (3.62) \quad &\sum_{j=1}^i \mathbb{P}_{\theta_1}(S_j \geq b) \\ &\leq \frac{1}{1 - e^{\Psi(\theta_1) - \Psi(\theta_1+r)}} \exp\{i[\Psi(\theta_1+r) - \Psi(\theta_1)] - rb\}. \end{aligned}$$

By the union bound, (3.60) and (3.62), we have

$$\begin{aligned} &\mathbb{E}\mathbb{I}\left(\bigcup_{0 \leq i < j \leq t_1} \{S_j - S_i \geq b\}\right) \\ &\leq e^{-\theta_1 b} \sum_{i=1}^{t_1} \sum_{j=1}^i \mathbb{P}_{\theta_1}(S_j \geq b) \\ &\leq \frac{e^{-\theta_1 b}}{1 - e^{\Psi(\theta_1) - \Psi(\theta_1+r)}} \sum_{i=1}^{t_1} \exp\{i[\Psi(\theta_1+r) - \Psi(\theta_1)] - rb\} \\ &\leq e^{-\theta_1 b} \left(\frac{1}{1 - e^{\Psi(\theta_1) - \Psi(\theta_1+r)}}\right)^2 \exp\{t_1[\Psi(\theta_1+r) - \Psi(\theta_1)] - rb\}. \end{aligned}$$

By the definition of  $r$  and  $t_1$ , the inequality  $(1/(1 - e^{-x}))^2 \leq Cx^{-2}$  for bounded  $x$ , and Taylor's expansion, the above bound can be further bounded as

$$\begin{aligned} & \mathbb{E} \mathbf{I} \left( \bigcup_{0 \leq i < j \leq t_1} \{S_j - S_i \geq b\} \right) \\ & \leq C e^{-\theta_1 b} \frac{b}{h^2(b)} \exp \{t_1 [\Psi(\theta_1 + r) - \Psi(\theta_1)] - rb\} \\ & \leq C e^{-\theta_1 b} \frac{b}{h^2(b)} \\ & \quad \times \exp \left\{ \left( \frac{b}{\mu_1} - b^{1/2} h(b) \right) \left( r\mu_1 + \frac{\sup_{\theta_1 \leq \theta \leq \theta'_1} \Psi''(\theta)}{2} r^2 \right) - rb \right\} \\ & \leq C e^{-\theta_1 b} \frac{b}{h^2(b)} \exp \left\{ -b^{1/2} h(b) r\mu_1 + \frac{\sup_{\theta_1 \leq \theta \leq \theta'_1} \Psi''(\theta) b}{2\mu_1} r^2 \right\} \\ & \leq C e^{-\theta_1 b} \frac{b}{h^2(b)} e^{-ch^2(b)}. \end{aligned}$$

This proves Lemma 3.9.  $\square$

The next lemma will be used in proving Lemma 3.11.

LEMMA 3.10. *Under the setting of Theorem 2.2, if  $\int_{-\infty}^{\infty} |\varphi_{\theta_1}(t)| dt < \infty$  where  $\varphi_{\theta_1}(t) = \mathbb{E}_{\theta_1} e^{itX_1}$ , then  $S_{\tau_+}$  under  $F_{\theta_1}$  has bounded density and is strongly nonarithmetic in the sense that*

$$\liminf_{|\lambda| \rightarrow \infty} |1 - \phi_{\theta_1}(\lambda)| > 0 \quad \text{where } \phi_{\theta_1}(\lambda) = \mathbb{E}_{\theta_1} e^{i\lambda S_{\tau_+}},$$

where  $\tau_+$  is defined just above (3.44).

PROOF. The condition  $\int_{-\infty}^{\infty} |\varphi_{\theta_1}(t)| dt < \infty$  implies that under  $F_{\theta_1}$ ,  $X_1$  is strongly nonarithmetic. By (8.42) of Siegmund (1985) with  $s = 1$ , the distribution of  $S_{\tau_+}$  under  $F_{\theta_1}$  is also strongly nonarithmetic. The condition  $\int_{-\infty}^{\infty} |\varphi_{\theta_1}(t)| dt < \infty$  also implies that the density of  $X_1$  under  $F_{\theta_1}$  is bounded by a constant  $M$ . Therefore,

$$\begin{aligned} & \mathbb{P}_{\theta_1}(S_{\tau_+} \in [x, x + dx]) \\ & \leq \mathbb{E}_{\theta_1} \sum_{n=0}^{\infty} \mathbf{I}(S_1, \dots, S_n \leq 0, S_{n+1} \in [x, x + dx]) \\ & \leq \mathbb{E}_{\theta_1} \sum_{n=0}^{\infty} \mathbf{I}(S_n \leq 0, S_{n+1} \in [x, x + dx]) \end{aligned}$$

$$\begin{aligned}
 &= \sum_{n=0}^{\infty} \int_{-\infty}^0 \mathbb{P}_{\theta_1}(S_n = dt) \mathbb{P}_{\theta_1}(X_1 \in [x - t, x + dx - t]) \\
 &\leq M dx \sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_n \leq 0) \leq C dx,
 \end{aligned}$$

where in the last inequality we used

$$(3.63) \quad \mathbb{P}_{\theta_1}(S_n \leq 0) \leq e^{\Psi(\theta_1 - \theta^*)n}$$

for  $0 < \theta^* < \theta_1$  such that  $\Psi(\theta_1 - \theta^*) < 0$ . This proves that  $S_{\tau_+}$  under  $F_{\theta_1}$  has bounded density.  $\square$

LEMMA 3.11. *Under the setting of Theorem 2.2, (3.59) holds.*

PROOF. From (3.57) and (3.60), we have

$$(3.64) \quad \lambda_2 = (n - t)e^{-\theta_1 b} \mathbb{P}(\tau_0 = \infty) \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{T_b} - b)}, S_{T_b} \geq b).$$

Since

$$\begin{aligned}
 \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)}) &= \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)}, S_{T_b} \geq b) \\
 &\quad + \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)}, S_{T_b} < 0),
 \end{aligned}$$

we have

$$\begin{aligned}
 &\mathbb{E}_{\theta_1}(e^{-\theta_1(S_{T_b} - b)}, S_{T_b} \geq b) \\
 (3.65) \quad &= \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)}) - \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)}, S_{T_b} < 0) \\
 &= \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)}) - \mathbb{E}_{\theta_1}\{\mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)} | S_{T_b}), S_{T_b} < 0\}.
 \end{aligned}$$

We first consider case 1. Let  $\tau_+^{(0)} = 0$ , and let  $\tau_+^{(k)}$  be defined recursively as  $\tau_+^{(k+1)} = \inf\{n > \tau_+^{(k)} : S_n > S_{\tau_+^{(k)}}\}$ . Define  $U(x) = \sum_{k=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(k)}} \leq x)$ . Observe that  $\{S_{\tau_+^{(k+1)}} - S_{\tau_+^{(k)}}, k = 0, 1, \dots\}$  are i.i.d. with the same distribution as  $S_{\tau_+}$ . By Lemma 3.10 and (2) of Stone (1965), we have

$$(3.66) \quad U(x) = \frac{x}{\mathbb{E}_{\theta_1} S_{\tau_+}} + \frac{\mathbb{E}_{\theta_1}(S_{\tau_+}^2)}{2(\mathbb{E}_{\theta_1} S_{\tau_+})^2} + o(e^{-cx}) \quad \text{as } x \rightarrow \infty.$$

Following the proof of Corollary 8.33 of Siegmund (1985), we have for  $x \geq 0$ ,

$$\begin{aligned}
 &\mathbb{P}_{\theta_1}(S_{\tau_b} - b > x) \\
 (3.67) \quad &= \sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(n)}} < b, S_{\tau_+^{(n+1)}} > b + x)
 \end{aligned}$$

$$\begin{aligned}
 &= \left( \int_{(0, b/2]} + \int_{(b/2, b)} \right) U(dt) \mathbb{P}_{\theta_1}(S_{\tau_+} > b + x - t) \\
 &= O(\mathbb{P}_{\theta_1}(S_{\tau_+} > b/2)U(b/2)) + \int_{(b/2, b)} U(dt) \mathbb{P}_{\theta_1}(S_{\tau_+} > b + x - t).
 \end{aligned}$$

For  $x > 0$ ,

$$\begin{aligned}
 \mathbb{P}_{\theta_1}(S_{\tau_+} > x) &= \mathbb{E}_{\theta_1} \left[ \sum_{i=0}^{\infty} \mathbf{I}(S_0, \dots, S_i \leq 0, X_{i+1} > x - S_i) \right] \\
 &\leq \sum_{i=0}^{\infty} \mathbb{P}_{\theta_1}(S_i \leq 0, X_{i+1} > x) \\
 &= \sum_{i=0}^{\infty} \mathbb{P}_{\theta_1}(S_i \leq 0) \mathbb{P}_{\theta_1}(X_{i+1} > x) \leq C \mathbb{P}_{\theta_1}(X_1 > x),
 \end{aligned}$$

where we used (3.63). Therefore, the right tail probability of  $S_{\tau_+}$  under  $F_{\theta_1}$  decays exponentially. From this fact and (3.66), the first term on the right-hand side of (3.67) is bounded by  $o(e^{-cb})$ . Let  $j = \lceil e^{cb} \rceil$  with small enough  $c$ , and let  $\Delta = \frac{b}{2j}$ . Let

$$A = \sum_{k=1}^j [U(b - (k - 1)\Delta) - U(b - k\Delta)] \mathbb{P}_{\theta_1}(S_{\tau_+} > x + k\Delta).$$

We have

$$\int_{(b/2, b)} U(dt) \mathbb{P}_{\theta_1}(S_{\tau_+} > b + x - t) \geq A$$

and by (3.66) and the fact that  $S_{\tau_+}$  under  $F_{\theta_1}$  has bounded density (cf. Lemma 3.10),

$$\begin{aligned}
 &\int_{(b/2, b)} U(dt) \mathbb{P}_{\theta_1}(S_{\tau_+} > b + x - t) - A \\
 &\leq \sum_{k=1}^j [U(b - (k - 1)\Delta) - U(b - k\Delta)] \mathbb{P}_{\theta_1}(S_{\tau_+} \in [x + (k - 1)\Delta, x + k\Delta]) \\
 &= o(e^{-cb}).
 \end{aligned}$$

From (3.66),

$$(3.68) \quad A = \sum_{k=1}^j \frac{\Delta}{\mathbb{E}_{\theta_1} S_{\tau_+}} \mathbb{P}_{\theta_1}(S_{\tau_+} > x + k\Delta) + O(je^{-cb})$$

with the same constant  $c$  as in (3.66). By choosing  $c$  in the definition of  $j$  to be small enough, the second term on the right-hand side of (3.68) is of smaller order

of  $e^{-cb}$ . Using the fact that  $S_{\tau_+}$  under  $F_{\theta_1}$  has bounded density and an exponential tail, we have

$$\begin{aligned} & \sum_{k=1}^j \frac{\Delta}{\mathbb{E}_{\theta_1} S_{\tau_+}} \mathbb{P}_{\theta_1}(S_{\tau_+} > x + k\Delta) \\ &= \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_x^\infty \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}). \end{aligned}$$

Therefore,

$$\begin{aligned} A &= \sum_{k=1}^j \left[ \frac{\Delta}{\mathbb{E}_{\theta_1} S_{\tau_+}} + o(e^{-cb}) \right] \mathbb{P}_{\theta_1}(S_{\tau_+} > x + k\Delta) \\ &= \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_x^\infty \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}). \end{aligned}$$

By (3.67) and the above argument, we have

$$\begin{aligned} (3.69) \quad & \mathbb{P}_{\theta_1}(S_{\tau_b} - b > x) \\ &= \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_x^\infty \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}). \end{aligned}$$

Using the integration by parts formula, (3.69) and (3.42), we obtain

$$\begin{aligned} (3.70) \quad & \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)}) \\ &= 1 - \theta_1 \int_0^\infty \mathbb{P}_{\theta_1}(S_{\tau_b} - b > x) e^{-\theta_1 x} dx \\ &= 1 - \frac{\theta_1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_0^\infty \int_x^\infty \mathbb{P}_{\theta_1}(S_{\tau_+} > y) e^{-\theta_1 x} dy dx + o(e^{-cb}) \\ &= 1 + \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \int_0^\infty (e^{-\theta_1 y} - 1) \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}) \\ &= \frac{1}{\mu_1 \mathbb{E}_{\theta_1} \tau_+} \int_0^\infty e^{-\theta_1 y} \mathbb{P}_{\theta_1}(S_{\tau_+} > y) dy + o(e^{-cb}) \\ &= \frac{1}{\theta_1 \mu_1} \exp\left(-\sum_{k=1}^\infty \frac{1}{k} \mathbb{E}_{\theta_1} e^{-\theta_1 S_k^+}\right) + o(e^{-cb}). \end{aligned}$$

Choosing  $\theta^*$  such that  $0 < \theta^* < \theta_1$  and  $\Psi(\theta_1 - \theta^*) < 0$ , we have

$$\begin{aligned} (3.71) \quad & 0 \leq \mathbb{P}_{\theta_1}(S_{T_b} \geq b) - \mathbb{P}_{\theta_1}(\tau_- = \infty) \\ & \leq \sum_{i=1}^\infty \mathbb{P}_{\theta_1}(S_i < -b) \leq e^{-\theta^* b} \sum_{i=1}^\infty e^{\Psi(\theta_1 - \theta^*) i} = o(e^{-cb}). \end{aligned}$$

From (3.65), (3.70) and (3.71), we have, with  $\tau_- := \inf\{n : S_n < 0\}$ ,

$$\begin{aligned} &\mathbb{E}_{\theta_1}(e^{-\theta_1(S_{T_b} - b)}, S_{T_b} \geq b) \\ &= \frac{1}{\theta_1 \mu_1} \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+}\right) \mathbb{P}_{\theta_1}(S_{T_b} \geq b) + o(e^{-cb}) \\ &= \frac{1}{\theta_1 \mu_1} \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+}\right) \mathbb{P}_{\theta_1}(\tau_- = \infty) + o(e^{-cb}) \end{aligned}$$

as  $b \rightarrow \infty$ .

From Lemma 3.6 and Corollary 2.4 of Woodroffe (1982), we have

$$\begin{aligned} &\mathbb{P}(\tau_0 = \infty) \mathbb{P}_{\theta_1}(\tau_- = \infty) \\ &= \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k} [\mathbb{P}(S_k \geq 0) + \mathbb{P}_{\theta_1}(S_k < 0)]\right\} \\ (3.72) \quad &= \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k} [\mathbb{P}_{\theta_1}(e^{-\theta_1 S_k}, S_k \geq 0) + \mathbb{P}_{\theta_1}(S_k < 0)]\right\} \\ &= \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k} \mathbb{E}_{\theta_1} e^{-\theta_1 S_k^+}\right\}. \end{aligned}$$

From (3.64) and (3.72), we have

$$\begin{aligned} \lambda_2 &= (n - t)e^{-\theta_1 b} \left\{ \frac{1}{\theta_1 \mu_1} \exp\left(-2 \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+}\right) + o(e^{-cb}) \right\} \\ &= \left[ 1 + O\left(\frac{b^{1/2} h(b)}{n - b/\mu_1}\right) \right] \lambda + (n - t)e^{-\theta_1 b} o(e^{-cb}). \end{aligned}$$

Next, we consider case 2. By a similar and simpler argument as for (3.69), we obtain, for integers  $k \geq 0$ ,

$$\begin{aligned} &\mathbb{P}_{\theta_1}(S_{\tau_b} - b = k) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(n)}} < b, S_{\tau_+^{(n+1)}} = b + k) \\ &= \sum_{m=1}^{b-1} \left[ \sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(n)}} = m) \right] \mathbb{P}_{\theta_1}(S_{\tau_+} = b + k - m) \\ &= O\left(\sum_{m=1}^{\lfloor b/2 \rfloor} \sum_{n=0}^{\infty} \mathbb{P}_{\theta_1}(S_{\tau_+^{(n)}} = m) \mathbb{P}_{\theta_1}(S_{\tau_+} \geq \lfloor b/2 \rfloor)\right) \end{aligned}$$

$$\begin{aligned}
 &+ \sum_{m=\lfloor b/2 \rfloor + 1}^{b-1} \mathbb{P}_{\theta_1}(S_{\tau_+} = b + k - m) \left( \frac{1}{\mathbb{E}_{\theta_1}(S_{\tau_+})} + o(e^{-cb}) \right) \\
 &= \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \mathbb{P}_{\theta_1}(S_{\tau_+} > k) + o(e^{-cb}).
 \end{aligned}$$

By the above equality and (3.43), we have

$$\begin{aligned}
 \mathbb{E}_{\theta_1}(e^{-\theta_1(S_{\tau_b} - b)}) &= \sum_{k=0}^{\infty} e^{-\theta_1 k} \frac{1}{\mathbb{E}_{\theta_1} S_{\tau_+}} \mathbb{P}_{\theta_1}(S_{\tau_+} > k) + o(e^{-cb}) \\
 &= \frac{1}{\mu_1 \mathbb{E}_{\theta_1} \tau_+ (1 - e^{-\theta_1})} [1 - \mathbb{E}_{\theta_1} e^{-\theta_1 S_{\tau_+}}] + o(e^{-cb}) \\
 &= \frac{1}{\mu_1 (1 - e^{-\theta_1})} \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+}\right) + o(e^{-cb}).
 \end{aligned}$$

Similar calculation as for case 1 yields

$$\begin{aligned}
 \lambda_2 &= (n - t) e^{-\theta_1 b} \left\{ \frac{1}{(1 - e^{-\theta_1}) \mu_1} \exp\left(-2 \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_{\theta_1} e^{-\theta_1 S_n^+}\right) + o(e^{-cb}) \right\} \\
 &= \left[ 1 + O\left(\frac{b^{1/2} h(b)}{n - b/\mu_1}\right) \right] \lambda + (n - t) e^{-\theta_1 b} o(e^{-cb}).
 \end{aligned}$$

The lemma is now proved.  $\square$

**4. Discussion.** The arguments we used to prove Theorems 2.1 and 2.2 may be useful in proving rates of convergence for tail probabilities of other test statistics for detecting local signals in sequences of independent random variables. Two for which some new techniques will be needed are the generalized likelihood ratio statistic and the Levin and Kline statistic [Levin and Kline (1985)].

For example, let  $\{X_1, \dots, X_n\}$  be independent random variables from the exponential family (2.1). Consider the testing problem at the beginning of the Introduction. If the mean of  $X_1$  is known and without loss of generality equal to 0, the generalized likelihood ratio statistic is  $\max_{1 \leq i < j \leq n} \sup_{\theta} [\theta(S_j - S_i) - (j - i)\Psi(\theta)]$ , where we have assumed without loss of generality that  $\Psi(0) = 0 = \Psi'(0)$ . Siegmund and Venkatraman (1995) derived an asymptotic approximation for the tail probability of this statistic in the normal case, while Siegmund and Yakir (2000) obtained similar results for a general exponential family. The bounds in (24) and (25) of Siegmund and Yakir (2000) suggest that the contribution to the maximum from  $i$  and  $j$  such that  $j - i$  is large can be neglected. However, how to define a local indicator function as in (3.47) which avoids “clumping” of 1’s and to evaluate approximately its expectation remains an open question.



If the mean of  $X_1$  is unknown, the statistic is more complicated; and its tail probability should be evaluated conditionally, given the value of  $S_n$ , which is a sufficient statistic for the unknown value of  $\theta$  under the null hypothesis of no change-point. The random variables  $\{X_1, \dots, X_n\}$  given  $S_n$  are globally dependent. In applying Theorem 3.1 to a sum of Bernoulli random variables  $W$  defined similarly as in (3.48), the error term  $b_3$  is no longer zero, although we believe it is small. Moreover, it is more challenging to derive a bound as in (3.58) conditionally.

Besides the distribution of the scan statistic  $M_{n;t}$  in (1.1), one may also be interested in the distribution of

$$N^+(b) := \sum_{i=1}^{n-t+1} I(X_i + \dots + X_{i+t-1} \geq b).$$

In fact, the distribution of  $M_{n;t}$  can be deduced from that of  $N^+(b)$  by the relation

$$\{M_{n;t} \geq b\} = \{N^+(b) \geq 1\}.$$

Theorem 2 of Dembo and Karlin (1992) gives a Poisson approximation result for  $N^+(b)$ . However, as discussed in Remark 2.4, their approximation may not be adequate because of the “clumping” phenomenon. A more suitable choice of the limiting distribution is a compound Poisson distribution. Stein’s method has been used to prove error bounds for the compound Poisson approximation for sums of Bernoulli random variables; see, for example, Barbour, Chen and Loh (1992). By combining Stein’s method with the analysis in this paper, one may be able to prove a compound Poisson approximation result for  $N^+(b)$ .

**Acknowledgments.** We would like to thank the Associate Editor and the referees for their helpful comments and suggestions which have significantly improved the presentation of this paper.

## REFERENCES

- ALDOUS, D. (1989). *Probability Approximations Via the Poisson Clumping Heuristic*. Applied Mathematical Sciences **77**. Springer, New York. [MR0969362](#)
- ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1990). Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5** 403–434. [MR1092983](#)
- ARRATIA, R., GORDON, L. and WATERMAN, M. S. (1990). The Erdős–Rényi law in distribution, for coin tossing and sequence matching. *Ann. Statist.* **18** 539–570. [MR1056326](#)
- BARBOUR, A. D. and CHEN, L. H. Y. (2005). The permutation distribution of matrix correlation statistics. In *Stein’s Method and Applications*. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. **5** 223–245. Singapore Univ. Press, Singapore. [MR2205339](#)
- BARBOUR, A. D., CHEN, L. H. Y. and LOH, W.-L. (1992). Compound Poisson approximation for nonnegative random variables via Stein’s method. *Ann. Probab.* **20** 1843–1866. [MR1188044](#)
- CHAN, H. P. and ZHANG, N. R. (2007). Scan statistics with weighted observations. *J. Amer. Statist. Assoc.* **102** 595–602. [MR2370856](#)
- CHEN, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545. [MR0428387](#)

- CHEN, C. and KARLIN, S. (2000). Poisson approximations for conditional  $r$ -scan lengths of multiple renewal processes and application to marker arrays in biomolecular sequences. *J. Appl. Probab.* **37** 865–880. [MR1782459](#)
- CHEN, C. and KARLIN, S. (2007).  $r$ -scan statistics of a Poisson process with events transformed by duplications, deletions, and displacements. *Adv. in Appl. Probab.* **39** 799–825. [MR2357382](#)
- DEMBO, A. and KARLIN, S. (1992). Poisson approximations for  $r$ -scan processes. *Ann. Appl. Probab.* **2** 329–357. [MR1161058](#)
- DIACONIS, P. and FREEDMAN, D. A. (1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. *J. Theoret. Probab.* **1** 381–410. [MR0958245](#)
- FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications. Vol. I*, 3rd ed. Wiley, New York. [MR0228020](#)
- FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II*, 2nd ed. Wiley, New York. [MR0270403](#)
- GLAZ, J., NAUS, J. and WALLENSTEIN, S. (2001). *Scan Statistics*. Springer, New York. [MR1869112](#)
- HAIMAN, G. (2000). Estimating the distributions of scan statistics with high precision. *Extremes* **3** 349–361. [MR1870463](#)
- HAIMAN, G. (2007). Estimating the distribution of one-dimensional discrete scan statistics viewed as extremes of 1-dependent stationary sequences. *J. Statist. Plann. Inference* **137** 821–828. [MR2301718](#)
- IGLEHART, D. L. (1972). Extreme values in the  $GI/G/1$  queue. *Ann. Math. Stat.* **43** 627–635. [MR0305498](#)
- KARLIN, S. and CHEN, C. (2000).  $r$ -scan statistics of a marker array in multiple sequences derived from a common progenitor. *Ann. Appl. Probab.* **10** 709–725. [MR1789977](#)
- KARLIN, S., DEMBO, A. and KAWABATA, T. (1990). Statistical composition of high-scoring segments from molecular sequences. *Ann. Statist.* **18** 571–581. [MR1056327](#)
- KARLIN, S. and RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivariate Anal.* **10** 467–498. [MR0599685](#)
- KOMLÓS, J. and TUSNÁDY, G. (1975). On sequences of “pure heads.” *Ann. Probab.* **3** 608–617. [MR0438440](#)
- LEVIN, B. and KLINE, J. (1985). The cusum test of homogeneity with an application in spontaneous abortion epidemiology. *Stat. Med.* **4** 469–488.
- PETROV, V. V. (1965). On the probabilities of large deviations for sums of independent random variables. *Teor. Veroyatnost. i Primenen* **10** 310–322. [MR0185645](#)
- SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York. [MR0799155](#)
- SIEGMUND, D. (1988). Approximate tail probabilities for the maxima of some random fields. *Ann. Probab.* **16** 487–501. [MR0929059](#)
- SIEGMUND, D. and VENKATRAMAN, E. S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.* **23** 255–271. [MR1331667](#)
- SIEGMUND, D. and YAKIR, B. (2000). Tail probabilities for the null distribution of scanning statistics. *Bernoulli* **6** 191–213. [MR1748719](#)
- SIEGMUND, D. and YAKIR, B. (2007). *The Statistics of Gene Mapping*. Springer, New York. [MR2301277](#)
- STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory* 583–602. Univ. California Press, Berkeley, CA. [MR0402873](#)
- STEIN, C. (1986). *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes—Monograph Series 7*. IMS, Hayward, CA. [MR0882007](#)

- STONE, C. (1965). On moment generating functions and renewal theory. *Ann. Math. Stat.* **36** 1298–1301. [MR0179857](#)
- TU, I.-P. and SIEGMUND, D. (1999). The maximum of a function of a Markov chain and application to linkage analysis. *Adv. in Appl. Probab.* **31** 510–531. [MR1724565](#)
- WOODROOFE, M. (1979). Repeated likelihood ratio tests. *Biometrika* **66** 453–463. [MR0556732](#)
- WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **39**. SIAM, Philadelphia, PA. [MR0660065](#)
- YAO, Q. W. (1993). Tests for change-points with epidemic alternatives. *Biometrika* **80** 179–191. [MR1225223](#)

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY  
NATIONAL UNIVERSITY OF SINGAPORE  
6 SCIENCE DRIVE 2  
SINGAPORE 117546  
REPUBLIC OF SINGAPORE  
E-MAIL: [stafx@nus.edu.sg](mailto:stafx@nus.edu.sg)

DEPARTMENT OF STATISTICS  
STANFORD UNIVERSITY  
SEQUOIA HALL  
390 SERRA MALL  
STANFORD, CALIFORNIA 94305-4065  
USA  
E-MAIL: [siegmund@stanford.edu](mailto:siegmund@stanford.edu)