Low-rate renewal theory and estimation

Georgios Fellouris

University of Illinois, Urbana-Champaign
Department of Statistics
725 South Wright Street,
Illini Hall, 116 Champaign, IL 61820 USA
e-mail: fellouri@illinois.edu

Abstract: Certain renewal theorems are extended to the case that the rate of the renewal process goes to 0 and, more generally, to the case that the drift of the random walk goes to infinity. These extensions are motivated by and applied to the problem of decentralized parameter estimation under severe communication constraints.

AMS 2000 subject classifications: Primary 62F12, 60G50; secondary 62F30, 60G40, 62N01.

Keywords and phrases: Anscombe’s theorem, Decentralized estimation, Efficient estimation, Low-rate, Random walk, Random sampling, Renewal theory, Overshoot.

1. Introduction

1.1. Low-rate renewal theory

Let \((\tau_n)_{n \in \mathbb{N}}\) be a random walk with finite, positive drift \(\delta\). That is, \((\tau_n - \tau_{n-1})_{n \in \mathbb{N}}\) are independent and identically distributed (i.i.d.) random variables with mean \(\delta := \mathbb{E}[\tau] \in (0, \infty)\), where \(\tau := \tau_1\) and \(\tau_0 := 0\). When \(\mathbb{P}(\tau > 0) = 1\), \((\tau_n)\) can be thought of as a sequence of random times at which certain events occur. Then, \(N(t) := \max\{n \in \mathbb{N} : \tau_n \leq t\}\) counts the number of events that have occurred up to time \(t\) and the “empirical rate”, \(N(t)/t\), converges in \(\mathcal{L}^r\) to the “theoretical rate”, \(1/\delta\), as \(t \to \infty\), i.e.,

\[
\frac{N(t)}{t} \frac{1}{\delta} \overset{\mathcal{L}^r}{\to} 1 \quad \text{as} \quad t \to \infty
\]

(1.1)

for any \(r > 0\). When \(\mathbb{P}(\tau > 0) < 1\), (1.1) is no longer valid, but it can be recovered if we replace \(N(t)\) by the first-passage time \(\nu(t) := \inf\{n \in \mathbb{N} : \tau_n > t\}\). Specifically, if \(\mathbb{E}[(\tau^-)^r] < \infty\) for some \(r \geq 1\), then

\[
\frac{\nu(t)}{t} \frac{1}{\delta} \overset{\mathcal{L}^r}{\to} 1 \quad \text{as} \quad t \to \infty.
\]

(1.2)
Here, and in what follows, \( \tau^+ := \max\{\tau, 0\} \) and \( \tau^- := \max\{-\tau, 0\} \) represent the positive and negative part of \( \tau \), respectively.

These asymptotic results are well known and it would be fair to say that they constitute the cornerstones of renewal theory; (1.1) goes back to Doob [10] for \( r = 1 \) and Hatori [16] for \( r > 1 \); (1.2) is due to Chow and Robbins [8] for \( r = 1 \), Chow [9] for \( r = 2 \) and Gut [15] for arbitrary \( r \geq 1 \). For more details on renewal theory we refer to Asmussen [2] and Gut [15].

Our first contribution in the present work is that we provide sufficient conditions on the moments of \( \tau \) so that these classical, asymptotic results remain valid when \( \delta \) is not fixed, but \( \delta \to \infty \) so that \( \delta = o(t) \). Specifically, in this asymptotic setup we show that (1.2) is preserved when

\[
E[(\tau^+)^{r+1}] = O(\delta^{r+1}) \quad \text{and} \quad E[|\tau - \delta|^r \vee 2] = O(\delta^r \vee 2),
\]

where \( r \vee 2 := \max\{r, 2\} \). When, in particular, \( P(\tau > 0) = 1 \), this condition reduces to \( E[\tau^{r+1}] = \mathcal{O}(\delta^{r+1}) \) and it guarantees that (1.1) is preserved as \( \delta \to \infty \) so that \( \delta = o(t) \).

In addition to these extensions, we also establish a version of Anscombe’s theorem [1] in the same, “low-rate” spirit. Thus, let \((Y_t)_{t \in \mathbb{N}}\) be an arbitrary sequence of random variables that converges in distribution as \( t \to \infty \) to some random variable \( Y_\infty \), i.e., \( Y_t \xrightarrow{D} Y_\infty \) as \( t \to \infty \). Suppose that \((Y_t)_{t \in \mathbb{N}}\) is observed only at a sequence, \((\tau_n)_{n \in \mathbb{N}}\), of random times that form a renewal process with rate \( 1/\delta \). If \( Y \) is uniformly continuous in probability, from classical renewal theory and Anscombe’s theorem [1] it follows that convergence is preserved when \( t \) is replaced by the last sampling time, \( \tau(t) := \tau_{N(t)} \), i.e.,

\[
Y_{\tau(t)} \xrightarrow{D} Y_\infty \quad \text{as} \quad t \to \infty
\]

for any fixed \( \delta \). Is this also the case when \( \delta \to \infty \) so that \( \delta = o(t) \)? We give a positive answer to this question under the assumption that \( \text{Var}[\tau] = \mathcal{O}(\delta^2) \).

The proofs of the above results are based on properties of stopped random walks, for which we refer to Gut [15], and Lorden’s [17] inequalities on the excess over a boundary and the age of a renewal process. Moreover, they provide alternative proofs of the classical, fixed-rate results under additional moment assumptions on \( \tau \).

1.2. Motivation

When \( P(\tau > 0) < 1 \), the asymptotic setup “\( \delta, t \to \infty \) so that \( \delta = o(t) \)” implies that the random walk \((\tau_n)\) has “large” drift \( \delta \) and crosses a positive level \( t \) that is
much larger than $\delta$. When $P(\tau > 0) = 1$, it implies that the renewal process $(\tau_n)$ is observed in an interval $[0, t]$ and the average time, $\delta$, between two consecutive events is “large”, yet much smaller than the length of observation window, $t$.

We are not aware of similar low-rate/large-drift extensions of (1.1)-(1.2) and (1.4). On the contrary, extensions and ramifications of the above results in the literature often require random walks with small drifts and renewal processes with high rates. For example, when $(\tau_n)$ is a random walk and the distribution of $\tau$ can be embedded in an exponential family, Siegmund [23] computed the limiting values of the moments of the overshoot $\tau_\nu(t) - t$ as $\delta \to 0$ and $t \to \infty$ so that $\delta t = O(1)$. For similar “corrected diffusion approximations” that require random walks with small drifts we refer to Chang[6], Chang and Peres[7], Blanchet and Glynn [5].

When $(\tau_n)$ is a renewal process that corresponds to the sampling times of some stochastic process, as for example in Mykland and Aït-Sahalia [3] or Rosenbaum and Tankov [19], limit theorems are also obtained as $\delta \to 0$, that is as the sampling rate goes to infinity. While such a high-frequency setup is natural in certain applications, such as mathematical finance (see, e.g., Zhang et. al [30], Florescu et. al [12]), and commonplace in the statistical inference of stochastic processes (see, e.g., Aït-Sahalia and Jacod [4], Tudor and Viens [26]), in other application areas it is often desirable to have low-frequency sampling.

This is, for example, the case when the process of interest is being monitored at some remote location by a network of (low-cost, battery-operated) sensors, which transmit their data to a fusion center subject to bandwidth and energy constraints [13]. This is often the situation in environmental monitoring, intelligent transportation, space exploration. In such applications, infrequent communication leads to significant energy gains, since it implies that for long time-periods the sensors can only sense the environment, without transmitting data to the fusion center [20]. On the other hand, limited bandwidth requires that each sensor transmit a low-bit message whenever it communicates with the fusion center [27]. Here, we consider a parameter estimation problem under such communication constraints, with the goal of demonstrating the usefulness of the aforementioned extensions of renewal theory.

### 1.3. Decentralized parameter estimation

Consider $K$ dispersed sensors that take repeated measurements of some quantity $\mu$ subject to measurement error. Thus, we assume that each sensor $k$ takes a sequence of i.i.d. observations $(X_k^t)_{t \in \mathbb{N}}$ with expectation $E[X_k^t] = \mu$, standard
deviation $\text{sd}[X_k^t] = \sigma_k$ and that observations are independent across sensors. The sensors transmit their data to a fusion center, which is responsible for estimating $\mu$. However, due to bandwidth constraints, they are allowed to transmit only low-bit messages. The problem is to decide (i) what messages each sensor transmits to the fusion center and (ii) how the fusion center utilizes these messages in order to estimate $\mu$.

This problem has been considered extensively in the engineering literature under the assumption that $\sigma_1, \ldots, \sigma_K$ are known and often assuming that $X_1^k, \ldots, X_K^k$ are identically distributed and/or that their densities have bounded support (see, e.g., Ribeiro and Giannakis [21], Xiao and Luo [28], Msechu and Giannakis [18] and the references therein). However, the proposed estimators in all these papers do not attain the asymptotic distribution of the optimal (in a mean-square sense) linear estimator that would be employed by the fusion center in the case that it had full access to the sensor observations. Indeed, if each sensor $k$ transmitted its exact observation $X_k^t$ at any time $t$, the fusion center could estimate $\mu$ at time $t$ with

$$
\hat{\mu}_t := \sum_{k=1}^{K} w_k \mu_k^t, \quad \mu_k^t := S_k^t / t, \quad S_k^t := X_1^k + \ldots + X_k^t,
$$

where $w_k$ is proportional to $(\sigma_k)^{-2}$ and $\sum_{k=1}^{K} w_k = 1$, and from the Central Limit Theorem we would have

$$
\frac{\hat{\mu}_t - \mu}{\sqrt{\sigma^2 \sigma / \sqrt{t}}} \rightarrow \mathcal{N}(0, 1) \quad \text{as} \quad t \rightarrow \infty,
$$

where $\sigma^2 := \sum_{k=1}^{K} \sigma_k^2 / K$. It has recently been shown that it is possible to obtain an estimator that attains this asymptotic distribution, even though it requires minimal transmission activity. Indeed, suppose that each sensor $k$ communicates with the fusion center whenever $(S_k^t)_{t \in \mathbb{N}}$ has changed by $\Delta_k > 0$ since the previous communication time. Then, the sequence of communication times of sensor $k$ is $\tau_k^t := \inf\{t \geq \tau_{k-1}^t : |S_k^t - S_k^{\tau_{k-1}^t}| \geq \Delta_k\}$ and $N_k(t) := \max\{n \in \mathbb{N} : \tau_k^n \leq t\}$ is the number of its transmitted messages up to time $t$. At time $\tau_k^n$, sensor $k$ transmits (with only 1 bit) the value of the indicator $z_k^n := 1(S_k^{\tau_k^n} - S_k^{\tau_{k-1}^n} \geq \Delta_k)$, informing the fusion center whether the upper or the lower threshold was crossed, but not about the size of the overshoot, $\eta_k^n := |S_k^{\tau_k^n} - S_k^{\tau_{k-1}^n}| - \Delta_k$. Based on this information, the fusion center can estimate $\mu$ at any time $t$ with

$$
\hat{\mu}_t := \sum_{k=1}^{K} w_k \hat{\mu}_k^t, \quad \hat{\mu}_k^t := \Delta_k \sum_{n=1}^{N_k^t} (2z_k^n - 1).
$$
It can be shown, as in Fellouris [11] (Section 4.2), that $\hat{\mu}_t$ preserves (1.6) when each $(X^k_t)_{t\in\mathbb{N}}$ results from high-frequency sampling of an underlying Brownian motion. More generally, if each sensor $k$ can transmit $d_k \geq 1$ bits whenever it communicates with the fusion center, then at each time $\tau^k_n$ it can transmit information not only about the sign, but also about the size of the overshoot, $\eta^k_n$. In this case, the resulting estimator will preserve (1.6) as long as the number of bits per transmission, $d_k$, goes to infinity as $t \to \infty$ for every $1 \leq k \leq K$ (see Yilmaz and Wang [31] for the case of Gaussian observations).

The above results, however, do not answer the question whether it is possible to construct an estimator at the fusion center that requires transmission of 1-bit messages from the sensors and, nevertheless, achieves the asymptotic distribution (1.6) for any distribution the sensor observations, $X^k_1$, $1 \leq k \leq K$, may follow. In this article, we show that this is indeed possible when sensor $k$ transmits the 1-bit message $z^k_n$ at each time $\tau^k_n$, as long as the fusion center estimator is modified as follows:

$$
\hat{\mu}_t := \sum_{k=1}^K w_k \hat{\mu}^k_t, \quad \hat{\mu}^k_t := \frac{\Delta^k}{\tau^k(t)} \sum_{n=1}^{N^k_t} (2z^k_n - 1)
$$

(1.8)

where $\tau^k(t) := \tau^k_{N^k(t)}$ is the last communication time from sensor $k$. More specifically, we show that (1.6) remains valid when $\mu_t$ is replaced by $\hat{\mu}_t$, as long as $\Delta^k \to \infty$ so that $\sqrt{t} \ll \Delta^k \ll t$ for every $1 \leq k \leq K$. In other words, the communication rate from the sensors to the fusion center must be sufficiently low for $\hat{\mu}_t$ to be efficient, not only from a practical, but also from a statistical point of view.

Note also that $\hat{\mu}_t$ does not require knowledge of the distribution of $X^k_1$, $1 \leq k \leq K$. When, however, these distributions are known and satisfy certain integrability conditions, we can construct a modification of $\hat{\mu}_t$ that preserves asymptotic distribution (1.6) when $\sqrt{t} \ll \Delta^k \ll t$ for every $1 \leq k \leq K$ (that is, for even higher rates of communication), whereas we can also construct a modification of $\hat{\mu}_t$ that, contrary to $\hat{\mu}_t$, does attain (1.6), as long as $\sqrt{t} \ll \Delta^k \ll \sqrt{t}$ for every $1 \leq k \leq K$. Finally, we consider the estimation of $\sigma$, when $\sigma_k$, $1 \leq k \leq K$ are equal to one another, but unknown. We show that it is also possible to estimate $\sigma$ at the fusion center with infrequent transmission of 1-bit messages from the sensors, which allows one to obtain asymptotic confidence intervals for the estimators of $\mu$ that satisfy (1.6).

The proofs of these properties rely on classical, renewal-theoretic results, such as the asymptotic behavior of the overshoots, as well as on the low-rate extensions of (1.1), (1.2) and (1.4), which we described in Subsection 1.1.
1.4. Summary

The rest of the paper is organized as follows: in Section 2, we obtain general, low-rate extensions of classical, renewal-theoretic results. More specifically, in Subsection 2.1 we establish (1.2) for an arbitrary random walk whose drift \( \delta \) goes to infinity so that \( \delta = o(t) \). In Subsection 2.2 we establish (1.1) for an arbitrary renewal process whose rate \( 1/\delta \) goes to 0 so that \( \delta = o(t) \). Moreover, we illustrate these two theorems in the case that \( \tau \) is stochastically less variable than an exponential random variable and in the case that \( \tau \) is the first hitting time of a spectrally negative Lévy process. In Subsection 2.3, we establish a low-rate version of Anscombe’s theorem.

In Section 3, we consider the problem of estimating the drift \( \mu \) of a random walk, \((S_t)_{t \in \mathbb{N}}\), where \( \mathbb{E}[S_t] = \mu t \) and \( \text{Var}[S_t] = \sigma^2 t \). In Subsection 3.1 we assume that \((S_t)_{t \in \mathbb{N}}\) is observed at an arbitrary renewal process, \((\tau_n)_{n \in \mathbb{N}}\). We show that the asymptotic distribution of the sample mean, \( S_t/t \), is preserved when we replace the current time \( t \) with the last sampling time, \( \tau(t) \), even if \( \delta \to \infty \) so that \( \delta = o(t) \), as long as \( \text{Var}[\tau] = \mathcal{O}(\delta^2) \). In Subsection 3.2, we focus on the case that \((S_t)\) is being sampled whenever it changes by \( \Delta > 0 \) since the previous sampling instance, but the size of the overshoot is not known. We show that, even in this setup, the asymptotic distribution of \( S_t/t \) can be preserved, as long as the sampling rate is sufficiently low, in the sense that \( \Delta \to \infty \) so that \( \sqrt{t} << \Delta << t \). In Subsection 3.3 we construct improved estimators using classical results from renewal theory that allow us to approximate the unobserved overshoots, as long as we know the distribution of the random walk and it satisfies certain assumptions. In Subsection 3.4, we compare numerically the above estimators and illustrate the advantages of low-rate sampling. In Subsection 3.5 we consider the estimation of \( \sigma \).

Section 3 focuses on the one-dimensional \((K = 1)\) version of the estimation problem that we described in Subsection 1.3. We do so mainly in order to lighten the notation. In Section 4, we formulate these results in a multi-dimensional setup. We conclude in Section 5.

2. Low-rate renewal theory

2.1. The case of random walks

Let \((\tau_n)_{n \in \mathbb{N}}\) be a random walk with positive drift \( \delta \), i.e., \((\tau_n - \tau_{n-1})_{n \in \mathbb{N}}\) is a sequence of i.i.d. random variables, not necessarily positive, with mean \( \delta := \mathbb{E}[\tau] \in (0, \infty) \) and finite variance \( \text{Var}[\tau] < \infty \), where \( \tau := \tau_1, \tau_0 := 0 \). For any
For $t > 0$, the first-passage time $\nu(t) := \inf\{n \in \mathbb{N} : \tau_n > t\}$ is a stopping time with respect to the filtration generated by $(\tau_n)_{n \in \mathbb{N}}$, therefore,

$$\tau_{\nu(t)} - \delta \nu(t) = \sum_{n=1}^{\nu(t)} (\tau_n - \tau_{n-1} - \delta)$$

is a stopped random walk with drift zero and from Wald’s identities we have

$$E[\tau_{\nu(t)} - \delta \nu(t)] = 0, \quad \text{Var}[\tau_{\nu(t)} - \delta \nu(t)] = E[\nu(t)] \text{Var}[\tau]. \quad (2.1)$$

Moreover, from Lorden’s [17] bound on the excess over a boundary we have

$$\sup_{t \geq 0} E[\tau_{\nu(t)} - t] \leq E[(\tau^+)^2] \delta. \quad (2.2)$$

Using (2.1)-(2.2), we can establish a low-rate extension of (1.2) when $r = 1$.

**Theorem 2.1.** If $\text{Var}[\tau] = \mathcal{O}(\delta^2)$ as $\delta \to \infty$, then

$$\frac{\delta \nu(t)}{t} \xrightarrow{t \to \infty} 1 \text{ as } \delta, t \to \infty \text{ so that } \delta = o(t).$$

When, in particular, $\text{Var}[\tau] = \mathcal{O}(\delta^q)$ as $\delta \to \infty$ for some $q \in [0, 2]$,

$$E\left[\left|\frac{\delta \nu(t)}{t} - 1\right|\right] = \mathcal{O}\left(\frac{\delta}{t}\right) + \mathcal{O}\left(\sqrt{\frac{\delta^{q-1}}{t}}\right). \quad (2.3)$$

**Proof.** It clearly suffices to prove (2.3). Let us first observe that

$$\left|\frac{\delta \nu(t)}{t} - 1\right| \leq \frac{|\delta \nu(t) - \tau_{\nu(t)}|}{t} + \frac{\tau_{\nu(t)} - t}{t}. \quad (2.4)$$

From (2.2) and the assumption that $E[(\tau^+)^2] = \mathcal{O}(\delta^2)$ we have

$$\frac{E[\tau_{\nu(t)} - t]}{\delta t} \leq \frac{E[(\tau^+)^2]}{\delta t} = \mathcal{O}(\delta^2) \delta t = \mathcal{O}\left(\frac{\delta}{t}\right). \quad (2.5)$$

Moreover, starting with an application of Hölder’s inequality, we have

$$\frac{E[|\delta \nu(t) - \tau_{\nu(t)}|]}{t} \leq \sqrt{E[(\delta \nu(t) - \tau_{\nu(t)})^2]} = \sqrt{E[\nu(t)|\text{Var}[\tau]]} = \sqrt{\frac{E[\tau_{\nu(t)}], \text{Var}[\tau]}{\delta t}}$$

$$= \sqrt{(1 + \mathcal{O}\left(\frac{\delta}{t}\right)) \frac{\mathcal{O}(\delta^q)}{\delta t}} = \mathcal{O}\left(\frac{\delta^{q-1}}{t}\right). \quad (2.6)$$

The first two equalities follow from Wald’s identities (2.1), whereas the third equality follows from (2.5) and the assumption that $\text{Var}[\tau] = \mathcal{O}(\delta^q)$. Now, taking expectations in (2.4) and applying (2.5)-(2.6) completes the proof. □
Remarks: (1) The proof of Theorem 2.1 provides an alternative way to prove (1.2) (for $r = 1$) when $\delta$ is fixed, under the additional assumption that $\text{Var}[\tau] < \infty$.

(2) The speed of convergence in $\delta \nu(t)/t \overset{L^1}{\to} 1$ is determined by $q$, which describes the growth of $\text{Var}[\tau]$ as $\delta \to \infty$. More specifically, the right-hand side in (2.3) is $O(\sqrt{\delta/t})$ when $q = 2$ and $O(\delta/t)$ when $q = 1$.

In order to extend (1.2) for arbitrary $r > 1$, in addition to (2.1)-(2.2) we need the stopped versions of the Marcinkiewicz-Zygmund inequalities (see [15], p. 22), according to which

$$E[(\tau - \delta)\nu(t)] \leq c_r E[\nu(t)^{r/2}] E[|\tau - \delta|^r],$$

for any $r \geq 2$, as well as the following generalization of (2.2),

$$\sup_{t \geq 0} E[(\tau - t)^r] \leq c_r \frac{E[(\tau^+)^{r+1}]}{\delta},$$

which is also due to Lorden [17] and is valid for any $r \geq 1$. Here, and in what follows, $c_r$ represents a generic, positive constant that depends on $r$.

Finally, we will need the following algebraic inequality:

$$|x + y|^r \leq 2^{r-1} (x^r + y^r), \quad x, y > 0.$$  \hfill (2.9)

**Theorem 2.2.** Let $r \geq 1$ and suppose that as $\delta \to \infty$

$$E[|\tau - \delta|^r \nu^2] = O(\delta^{r+2}) \quad \text{and} \quad E[(\tau^+)^{r+1}] = O(\delta^{r+1}).$$

Then

$$\frac{\delta \nu(t)}{t} \overset{L^r}{\to} 1 \quad \text{as} \quad \delta, t \to \infty \quad \text{so that} \quad \delta = o(t).$$

If, in particular, $E[|\tau - \delta|^r \nu^2] = \Theta(\delta^{r+2})$, where $q_{r+2} \in [0, r \lor 2]$, then

$$E\left[\left|\frac{\delta \nu(t)}{t} - 1\right|^r\right] = O\left(\frac{\delta}{t} \right)^r + O\left(\frac{\delta^{q_{r+2}}}{t^r}\right),$$

where $\alpha_r := 2(q_{r+2}/r \lor 2) - 1$. \hfill (2.10)

It is clear that Theorem 2.2 reduces to Theorem 2.1 when $r = 1$. Before we prove it, we need to state the following lemma, which implies that if condition (*) holds for some $r > 1$, then it also holds for any $s \in [1, r]$.

**Lemma 2.1.** (i) If $r \geq 1$, then

$$E[(\tau^+)^{r+1}] = O(\delta^{r+1}) \Rightarrow E[(\tau^+)^s] = O(\delta^s), \quad \forall s \in [1, r+1].$$
(ii) If \( r \geq 2 \), then
\[
E[|\tau - \delta^r|] = O(\delta^r) \Rightarrow E[|\tau - \delta^s|] = O(\delta^s), \quad \forall s \in [2, r].
\]

**Proof.** If \( E[(\tau^+)^r|t^r] = O(\delta^{r+1}) \) for some \( r \geq 1 \), then from Hölder’s inequality it follows that for any \( s \in [1, r + 1] \)
\[
E[(\tau^+)^s] \leq (E[(\tau^+)^{(r+1)}])^{s/(r+1)} = (O(\delta^{r+1}))^{s/(r+1)} = O(\delta^s).
\]

Similarly, if \( E[|\tau - \delta|] = O(\delta^r) \) for some \( r \geq 2 \), then
\[
E[|\tau - \delta^s|] \leq (E[|\tau - \delta|])^{s/r} = (O(\delta^r))^{s/r} = O(\delta^s).
\]

\[\square\]

**Proof of Theorem 2.2.** Let \( r \geq 1 \). By definition, \( \alpha_r \leq 1 \), thus, it suffices to prove (2.10). Applying the algebraic inequality (2.9) to (2.4) and taking expectations we have
\[
E\left[ \left| \frac{\delta \nu(t)}{t} - 1 \right|^r \right] \leq c_r \left( E\left[ |\delta \nu(t) - \tau_{\nu(t)}|^r \right] t^{-r} + E\left[ (\tau_{\nu(t)} - t)^r \right] \right). \tag{2.11}
\]

From (2.8) and the assumption \( E[(\tau^+)^{r+1}] = O(\delta^{r+1}) \) it follows that
\[
E[(\tau_{\nu(t)} - t)^r] \leq c_r E[(\tau^+)^{r+1}] \delta^{-r} = O\left( \frac{\delta^r}{t^r} \right). \tag{2.12}
\]

From Hölder’s inequality and (2.7) we have
\[
E[|\delta \nu(t) - \tau_{\nu(t)}|^r] \leq E[|\delta \nu(t) - \tau_{\nu(t)}|^{r\nu_2}]^{\frac{r}{r\nu_2}} \leq c_r \left( E[|\nu(t)|^{\nu_2}] E[|\tau - \delta|^{r\nu_2}] \right)^{\frac{r}{r\nu_2}} = c_r E\left[ |\nu(t)|^{\nu_2} \right]^{\frac{r}{r\nu_2}} E\left[ |\tau - \delta|^{r\nu_2} \right]^{\frac{r}{r\nu_2}},
\]
and, consequently,
\[
E\left[ \frac{E[|\delta \nu(t) - \tau_{\nu(t)}|^r]}{t^r} \right] \leq c_r E\left[ \left( \frac{\delta \nu(t)}{t} \right)^{\frac{r}{r\nu_2}} \right]^{\frac{r}{r\nu_2}} E\left[ |\tau - \delta|^{r\nu_2} \right]^{\frac{r}{r\nu_2}} \left( \frac{\delta}{t} \right)^\nu.
\]

Therefore, from the assumption \( E[|\tau - \delta|^{r\nu_2}] = O(\delta^{\alpha_r\nu_2}) \) and the definition of \( \alpha_r \) it follows that
\[
E\left[ \frac{E[|\delta \nu(t) - \tau_{\nu(t)}|^r]}{t^r} \right] \leq E\left[ \left( \frac{\delta \nu(t)}{t} \right)^{\frac{r}{r\nu_2}} \right]^{\frac{r}{r\nu_2}} O\left( \frac{\delta^{\alpha_r}}{t} \right)^\frac{\nu}{r}. \tag{2.13}
\]

From (2.11)-(2.13) we obtain
\[
E\left[ \left| \frac{\delta \nu(t)}{t} - 1 \right|^r \right] \leq O\left( \frac{\delta}{t} \right)^r + E\left[ \left( \frac{\delta \nu(t)}{t} \right)^{\frac{r}{r\nu_2}} \right]^{\frac{r}{r\nu_2}} O\left( \frac{\delta^{\alpha_r}}{t} \right)^\frac{\nu}{r}.
\]
and in order to complete the proof it suffices to show that for any \( r \geq 1 \),
\[
\mathbb{E} \left[ \left( \frac{\delta \nu(t)}{t} \right)^2 \right] = 1 + o(1) \quad \text{as} \quad \delta, t \to \infty \quad \text{so that} \quad \delta = o(t).
\]

More specifically, when \( r \in [1, 2] \), it suffices to show that if \( \text{Var}[\tau] = \Theta(\delta^q) \) for some \( q \in [0, 2] \) and \( \mathbb{E}[(\tau^+)^{r+1}] = O(\delta^{r+1}) \), then
\[
\mathbb{E} \left[ \frac{\delta \nu(t)}{t} \right] \to 1 \quad \text{as} \quad \delta, t \to \infty \quad \text{so that} \quad \delta = o(t).
\]

This follows directly from Theorem 2.1 and Lemma 2.1(i). Therefore, (2.10) holds for any \( 1 \leq r \leq 2 \). When \( r \in [2, 4] \), it suffices to show that if \( \mathbb{E}[(\tau^-)^{r+1}] = O(\delta^{r+1}) \) for some \( q \in [0, r] \) and \( \mathbb{E}[(\tau^+)^{r+1}] = O(\delta^{r+1}) \), then
\[
\mathbb{E} \left[ \left( \frac{\delta \nu(t)}{t} \right)^2 \right] \to 1 \quad \text{as} \quad \delta, t \to \infty \quad \text{so that} \quad \delta = o(t).
\]

This follows from Lemma 2.1 and the fact that (2.10) holds for \( r/2 \in (1, 2) \). This proves (2.10) for \( r \in [2, 4] \) and the proof is similar when \( r \in [4, 8] \), etc. \( \square \)

**Remarks:**

(1) The proof of Theorem 2.2 provides an alternative way to prove (1.2) in the case that \( \delta \) is fixed, under the additional condition that \( \mathbb{E}[(\tau^+)^{r+1}] < \infty \).

(2) The rate of convergence in \( \delta \nu(t)/t \overset{L^2}{\rightarrow} 1 \) is determined by \( q \in [0, r] \), which describes the growth of the \( r \)-th central moment of \( \tau \) as \( \delta \to \infty \). Specifically, the right-hand side in (2.10) is equal to \( O(\sqrt{\delta^r/t^r}) \) when \( q = (r \vee 2)/2 \).

(3) There is a clear dichotomy in the conditions required for \( \delta \nu(t)/t \overset{L^2}{\rightarrow} 1 \) to hold as \( \delta, t \to \infty \) so that \( \delta = o(t) \). Indeed, when \( 1 \leq r \leq 2 \), the second central moment of \( \tau \) must be finite and of order \( O(\delta^2) \); when \( r \geq 2 \), the \( r \)-th central moment of \( \tau \) must be finite and of order \( O(\delta^{r/2}) \). Compare, in particular, Theorem 2.1 with the following corollary of Theorem 2.2.

**Corollary 2.1.** If \( \text{Var}[\tau] = O(\delta^q) \) and \( \mathbb{E}[(\tau^+)^3] = O(\delta^3) \) as \( \delta \to \infty \), then
\[
\frac{\delta \nu(t)}{t} \overset{L^2}{\rightarrow} 1 \quad \text{as} \quad \delta, t \to \infty \quad \text{so that} \quad \delta = o(t).
\]

When, in particular, \( \text{Var}[\tau] = O(\delta^q) \) for some \( q \in [0, 2] \), then as \( \delta, t \to \infty \) so that \( \delta = o(t) \):
\[
\mathbb{E} \left[ \left( \frac{\delta \nu(t)}{t} - 1 \right)^2 \right] = O\left( \frac{\delta}{t} \right)^2 + O\left( \frac{\delta^{q-1}}{t} \right).
\]

**Proof.** If follows by setting \( r = 2 \) in Theorem 2.2. \( \square \)
2.2. The case of renewal processes

We now extend (1.1) to the case that \( \delta, t \to \infty \) so that \( \delta = o(t) \).

**Theorem 2.3.** Let \( r \geq 1 \) and suppose that \( P(\tau > 0) = 1 \). If \( E[\tau^{r+1}] = O(\delta^{r+1}) \) as \( \delta \to \infty \), then

\[
\frac{\delta N(t)}{t} \xrightarrow{c_r} 1 \quad \text{as} \quad \delta, t \to \infty \quad \text{such that} \quad \delta = o(t).
\]

When, in particular, \( E[|\tau - \delta|^{r \vee 2}] = O(\delta^{q r ^2}) \), where \( q r ^2 \in [0, r \vee 2] \), then

\[
E\left[ \left| \frac{\delta N(t)}{t} - 1 \right| \right] = O\left( \delta^r \right) + O\left( \frac{\delta^{qr^2}}{t} \right),
\]

(2.14)

where \( \alpha_r := 2 \left( q r^2 / r^2 \right) - 1 \).

**Proof.** It clearly suffices to show (2.14). Since \( \nu(t) = N(t) + 1 \) when \( P(\tau > 0) = 1 \), we have

\[
\left| \frac{\delta N(t)}{t} - 1 \right| \leq \left| \frac{\delta \nu(t)}{t} - 1 \right| + \frac{\delta}{t},
\]

thus, applying the algebraic inequality (2.9) and taking expectations we obtain

\[
E\left[ \left| \frac{\delta \nu(t)}{t} - 1 \right| \right] \leq c_r \left( E\left[ \left| \frac{\delta \nu(t)}{t} - 1 \right| \right] + \left( \frac{\delta}{t} \right)^r \right).
\]

Therefore, it suffices to show that condition (*) of Theorem 2.2 is implied by \( E[\tau^{r+1}] = O(\delta^{r+1}) \) when \( P(\tau > 0) = 1 \). Indeed, for any \( r \geq 1 \) we have

\[
E[|\tau - \delta|^{r \vee 2}] \leq E[\tau^{r \vee 2}] \leq E[\tau^{r+1}] \frac{\delta^{r+2}}{r+1} = O(\delta^{r+1}) \frac{\delta^{r+2}}{r+1} = O(\delta^{r \vee 2}),
\]

where the first inequality is due to \( P(\tau > 0) = 1 \) and the second one follows from Hölder’s inequality and the fact that \( r \vee 2 \leq r + 1 \).

We now illustrate the previous theorems for two classes of renewal processes.

2.2.1. Renewals stochastically less variable than the Poisson process

Let us start by showing that the condition of Theorem 2.3 is satisfied when \( \tau \) is exponentially distributed. Indeed, for any \( r \geq 1 \), if we denote by \( [r] := \min\{n \in \mathbb{N} : n \geq r\} \) the ceiling of \( r \), then

\[
E[\tau^{r+1}] \leq E[\tau^{[r+1]}] \frac{\delta^{r+1}}{\delta^{r+1}} \leq \delta^{r+1},
\]

where the first inequality follows from Hölder’s inequality and the second one holds because \( E[\tau^n] = \delta^n / n! \leq \delta^n \) for any \( n \in \mathbb{N} \).
More generally, suppose that \( \tau \) is not itself exponentially distributed, but is \textit{stochastically less variable} than an exponential random variable \( \tau' \) that has mean \( \delta \), i.e.,

\[
E[g(\tau)] \leq E[g(\tau')] \quad \forall \text{ convex and increasing function } g : (0, \infty) \to (0, \infty).
\]

Then, since \( x \to x^{r+1} \) is convex and increasing on \((0, \infty)\), it follows that \( E[\tau^{r+1}] \leq E[(\tau')^{r+1}] \leq \delta^{r+1} \), which proves that the condition of Theorem 2.3 is still satisfied when \( \tau \) is stochastically less variable than an exponential random variable. This is, for example, the case when \( \tau \) is \textit{new better than used in expectation}, i.e., \( E[\tau - t | \tau > t] \leq E[\tau] \) for every \( t > 0 \) (see, e.g., [22], pp. 435-437).

### 2.2.2. First-hitting times of spectrally negative Lévy process

Suppose that \( \tau = \inf \{ t : Y_t \geq \Delta \} \), where \( \Delta > 0 \) and \( (Y_t)_{t\geq0} \) is a spectrally negative Lévy process with positive drift. That is, \( (Y_t)_{t\geq0} \) is a stochastic process that is continuous in probability, has stationary and independent increments, does not have positive jumps and \( E[Y_t] = \mu t \) for some \( \mu > 0 \). Then, from Wald's identity it follows that \( \delta = E[\tau] = \Delta / \mu \) and if, additionally, \( E[(Y_t^-)^{r+1}] < \infty \), then from Theorem 4.2 in [14] we have

\[
E[\tau^{r+1}] = (\Delta / \mu)^{r+1}(1 + o(1)) \quad \text{as } \Delta \to \infty.
\]

Note that \( \delta \to \infty \) if either \( \Delta \to \infty \) or \( \mu \to 0 \). Thinking of \( (\tau_n) \) as sampling times of \( (Y_t) \), it is natural to consider \( \mu \) as fixed and to assume that the sampling period \( \delta \) is controlled by threshold \( \Delta \). This implies that \( E[\tau^{r+1}] = \Theta(\delta^{r+1}) \) as \( \delta \to \infty \) and proves that the condition of Theorem 2.3 is satisfied.

Furthermore, if \( \text{Var}[Y_t] = \sigma^2 t \) for some \( \sigma > 0 \), then \( \text{Var}[\tau] = (\sigma^2 / \mu^3) \Delta \), which implies that \( \text{Var}[\tau] = \Theta(\delta) \) as \( \delta \to \infty \). Therefore, from Theorem 2.3 we obtain

\[
E \left[ \left( \frac{\delta N(t)}{t} - 1 \right) \right] = O\left( \frac{\delta}{t} \right).
\]

**Remark:** Analogous results can be obtained when \( (Y_t)_{t\in\mathbb{N}} \) is a random walk and/or \( \tau = \inf \{ t : |Y_t| \geq \Delta \} \), a setup that we consider in detail in Subsection 3.2.

### 2.3. A low-rate version of Anscombe’s theorem

Let \( (Y_t)_{t\in\mathbb{N}} \) be a sequence of random variables that is uniformly continuous in probability and converges in distribution as \( t \to \infty \) to a random variable, \( Y_\infty \).
i.e., $Y_t \overset{D}{\to} Y_\infty$ as $t \to \infty$. Suppose that $(Y_t)_{t \in \mathbb{N}}$ can only be observed at a sequence of random times $(\tau_n)_{n \in \mathbb{N}}$ that form a renewal process. As before, we set $\tau := \tau_1$, $N(t) := \max\{n : \tau_n < t\}$ and $\tau(t) := \tau_{N(t)}$ is the most recent sampling time at time $t$. In the following theorem we show that this convergence remains valid when we replace $t$ with the most recent sampling time, $\tau(t)$, even if $\delta \to \infty$ so that $\delta = o(t)$, as long as the second moment of $\tau$ grows at most as the square of its mean, $\delta^2$.

**Theorem 2.4.** If $E[\tau^2] = O(\delta^2)$, or equivalently $\text{Var}[\tau] = O(\delta^2)$, then $\tau(t)/t \overset{L_1}{\to} 1$ and, consequently, $Y_{\tau(t)} \overset{D}{\to} Y_\infty$ as $\delta, t \to \infty$ so that $\delta = o(t)$.

**Proof.** From Lorden’s bound (see [17], p.526) on the expected value of the age of a renewal process and the assumption of the theorem we obtain

$$E[t - \tau(t)] \leq E[\tau^2]/\delta = O(\delta).$$

Therefore,

$$E\left[\frac{\tau(t)}{t} - 1\right] = \frac{E[t - \tau(t)]}{t} = O\left(\frac{\delta}{t}\right),$$

which proves that $\tau(t)/t \overset{L_1}{\to} 1$ (and, consequently, $Y_{\tau(t)} \overset{D}{\to} Y_\infty$, due to Anscombe’s theorem) as $\delta, t \to \infty$ so that $\delta = o(t)$. Finally, since $E[\tau^2] = \delta^2 + \text{Var}[\tau]$, it is clear that $E[\tau^2] = O(\delta^2)$ is equivalent to $\text{Var}[\tau] = O(\delta^2)$.

**Remarks:** (1) The condition of Theorem 2.4 is satisfied when $(\tau_n)$ is the Binomial process, i.e., sampling at any time $t \in \mathbb{N}$ has probability $1/\delta$ and is independent of the past. Indeed, in this case $\tau$ is geometrically distributed with mean $\delta$ and its variance is equal to $\text{Var}[\tau] = \delta^2(1 - \delta^{-1})$.

(2) Theorem 2.4 also applies when $(\tau_n)$ corresponds to the hitting times of random walk, a setup that we consider in detail in Subsection 3.2.

3. Low-rate estimation

In this section, $(X_t)_{t \in \mathbb{N}}$ is a sequence of i.i.d. random variables with unknown mean $\mu := E[X]$ and finite standard deviation $\sigma := SD[X]$, where $X := X_1$. From the Central Limit Theorem we know that the sample mean is an asymptotically normal estimator of $\mu$, i.e.,

$$\frac{\mu_t - \mu}{\sigma/\sqrt{t}} \to N(0, 1) \quad \text{as} \quad t \to \infty,$$

where $\mu_t := S_t/t$ and $S_t := X_1 + \ldots + X_t$. 
3.1. Sampling at an arbitrary renewal process

When the random walk \((S_t)_{t \in \mathbb{N}}\) is observed only at a sequence of random times \((\tau_n)_{n \in \mathbb{N}}\) that form a renewal process, two natural modifications of \(\mu_t\) are

\[
\bar{\mu}_t = \frac{S_{\tau(t)}}{\tau(t)} \quad \text{and} \quad \tilde{\mu}_t = \frac{S_{\tau(t)}}{t}.
\]

That is, \(\bar{\mu}_t\) is the sample mean evaluated at the most recent sampling time \(\tau(t) = \tau_N(t)\) and is well defined only when \(t \geq \tau_1\), in contrast to \(\tilde{\mu}_t\) that is well defined for any \(t > 0\). The following theorem describes the asymptotic behavior of these estimators when \(\delta, t \to \infty\) so that \(\delta = o(t)\).

**Theorem 3.1.** Suppose that \(\text{Var}[\tau] = O(\delta^2)\) as \(\delta \to \infty\).

(i) If \(\delta, t \to \infty\) so that \(\delta = o(t)\), then \(\bar{\mu}_t \xrightarrow{p} \mu\) and

\[
\frac{\bar{\mu}_t - \mu}{\sigma/\sqrt{t}} \xrightarrow{d} \mathcal{N}(0, 1).
\]

(ii) If \(\delta, t \to \infty\) so that \(\delta = o(t)\), then \(\tilde{\mu}_t \xrightarrow{p} \mu\). If also \(\delta = o(\sqrt{t})\), then

\[
\frac{\tilde{\mu}_t - \mu}{\sigma/\sqrt{t}} \xrightarrow{d} \mathcal{N}(0, 1).
\]

**Proof.** (i) From the Strong Law of Large Numbers we know that \(\mu_t \xrightarrow{a.s.} \mu\) as \(t \to \infty\). Theorem 2.4 implies that \(\tau(t)/t \xrightarrow{p} 1\) and, consequently, \(\tau(t) \xrightarrow{p} \infty\) and \(\bar{\mu}_t \xrightarrow{p} \mu\) as \(\delta, t \to \infty\) so that \(\delta = o(t)\) (see, e.g., [15], page 12). From Theorem 2.4 and (3.1) we obtain (3.3).

(ii) We observe that

\[
|\tilde{\mu}_t - \bar{\mu}_t| = |S_{\tau(t)}| \frac{t - \tau(t)}{t \tau(t)} = |\bar{\mu}_t| \left(1 - \frac{\tau(t)}{t}\right).
\]

Since \(\tau(t)/t \xrightarrow{p} 1\) and \(\bar{\mu}_t \xrightarrow{p} \mu\) as \(\delta, t \to \infty\) so that \(\delta = o(t)\), (3.5) clearly implies that \(\bar{\mu}_t \xrightarrow{p} \mu\) as \(\delta, t \to \infty\) so that \(\delta = o(t)\). Finally, in order to prove (3.4), it suffices to show that

\[
\sqrt{t}|\bar{\mu}_t - \tilde{\mu}_t| \xrightarrow{p} 0 \quad \text{as} \quad \delta, t \to \infty \quad \text{so that} \quad \delta = o(\sqrt{t}).
\]

This is indeed the case since from (i) we have that \(\bar{\mu}_t \xrightarrow{p} \mu\) when \(\delta = o(t)\), whereas from (2.15) it follows that \(t - \tau(t)/\sqrt{t} \xrightarrow{d} C\) when \(\delta = o(\sqrt{t})\).

**Remark:** When the sampling period \(\delta\) is fixed as \(t \to \infty\), it is clear that both \(\bar{\mu}_t\) and \(\tilde{\mu}_t\) preserve the asymptotic distribution (3.1). On the other hand, when
δ → ∞ the asymptotic behavior of the two estimators differs, since \( \tilde{\mu}_t \) requires that the sampling rate should not be too low (\( \delta \ll \sqrt{t} \)), a condition that is not necessary for \( \bar{\mu}_t \). Therefore, \( \tilde{\mu}_t \) and \( \hat{\mu}_t \) have similar behavior for small \( \delta \), but \( \bar{\mu}_t \) is more robust than \( \hat{\mu}_t \) under low-rate sampling, a conclusion that is also verified empirically in Figure 3.4.

### 3.2. Sampling at first hitting times

We now focus on the case that the random walk \((S_t)_{t \in \mathbb{N}}\) is being sampled whenever it changes by a fixed amount \( \Delta \) since the previous sampling instance, i.e.,

\[
\tau_n := \inf \{ t \geq \tau_{n-1} : S_t - S_{\tau_{n-1}} \geq \Delta \}, \quad n \in \mathbb{N},
\]  

(3.6)

where, for simplicity, we have assumed that \( \mu > 0 \). Then, the estimators \( \bar{\mu}_t, \tilde{\mu}_t \), defined in (3.2), take the form

\[
\bar{\mu}_t = \frac{1}{\tau(t)} \sum_{n=1}^{N(t)} \left[ \Delta + \eta_n \right] = \frac{N(t) \Delta}{\tau(t)} + \frac{1}{\tau(t)} \sum_{n=1}^{N(t)} \eta_n,
\]  

(3.7)

\[
\tilde{\mu}_t = \frac{1}{t} \sum_{n=1}^{N(t)} \left[ \Delta + \eta_n \right] = \frac{N(t) \Delta}{t} + \frac{1}{t} \sum_{n=1}^{N(t)} \eta_n,
\]  

(3.8)

where \( \eta_n := S_{\tau_n} - S_{\tau_{n-1}} - \Delta \). Since \( X \) has a finite second moment, the overshoots \((\eta_n)_{n \in \mathbb{N}}\) are i.i.d. with \( E[\eta] = O(1) \) as \( \Delta \to \infty \), where \( \eta := \eta_1 \) (see, e.g., Theorem 10.5 in [15]). Therefore, from Wald’s identity it follows that

\[
\delta = E[\tau] = \frac{E[S_{\tau}]}{\mu} = \frac{\Delta + E[\eta]}{\mu}
\]  

(3.9)

and, since we consider \( \mu \) to be fixed, the sampling period \( \delta \) is controlled by threshold \( \Delta \), in the sense that \( \delta = \Delta/\mu + O(1) = \Theta(\Delta) \) as \( \Delta \to \infty \). Moreover, since

\[
\text{Var}[\tau] = (\sigma^2/\mu^2) \Delta (1 + o(1)) \quad \text{as} \quad \Delta \to \infty,
\]  

(3.10)

(see, e.g., Theorem 9.1 in [15]), it is clear that \( \text{Var}[\tau] = \Theta(\delta) \) as \( \delta \to \infty \). Therefore, as \( \delta, t \to \infty \) so that \( \delta = o(t) \), from Theorems 2.2 and 2.3 we have

\[
E \left[ \left| \frac{\delta \nu(t)}{t} - 1 \right| \right] = O(\delta/t) \quad \text{and} \quad E \left[ \left| \frac{\delta N(t)}{t} - 1 \right| \right] = O(\delta/t),
\]  

(3.11)

from Theorem 2.4 we obtain

\[
\tau(t)/t \overset{P}{\to} 1,
\]  

(3.12)

and we also observe that the condition of Theorem 3.1 is satisfied. However, \( \tilde{\mu}_t \) and \( \hat{\mu}_t \) are not applicable when the overshoots \((\eta_n)_{n \in \mathbb{N}}\) are unobserved, i.e.,
when at each time $\tau_n$ we do not learn the excess of $S_{\tau_n} - S_{\tau_{n-1}}$ over $\Delta$. In this case, $\bar{\mu}_t$ and $\hat{\mu}_t$ reduce to

$$\hat{\mu}_t := \frac{N(t) \Delta}{\tau(t)} \quad \text{and} \quad \tilde{\mu}_t := \frac{N(t) \Delta}{t},$$

respectively. In the following theorem we show that both $\hat{\mu}_t$ and $\tilde{\mu}_t$ are consistent when $t$ and $\delta \to \infty$, but not for fixed $\delta$. More surprisingly, we also show that if the sampling rate is sufficiently low ($\sqrt{t} \ll \delta \ll t$), then $\hat{\mu}_t$ preserves the asymptotic distribution (3.1). On the other hand, $\tilde{\mu}_t$ fails to do so for any sampling rate.

**Theorem 3.2.** (i) If $\delta, t \to \infty$ so that $\delta = o(t)$, then

$$E[|\tilde{\mu}_t - \mu|] = O\left(\frac{\delta}{\sqrt{t}}\right) + E[\eta] \Theta\left(1\right),$$

and, consequently, $\sqrt{t} E[|\tilde{\mu}_t - \mu|] = O(1)$ when $\delta = \Theta(\sqrt{t})$.

(ii) If $\delta, t \to \infty$ so that $\delta = o(t)$, then $\hat{\mu}_t \xrightarrow{p} \mu$.

(iii) If $\delta, t \to \infty$ so that $\sqrt{t} \ll \delta \ll t$, i.e., $\delta = o(t)$ and $\sqrt{t} = o(\delta)$, then

$$\frac{\hat{\mu}_t - \mu}{\sigma/\sqrt{t}} \xrightarrow{d} N(0, 1).$$

**Proof.** (i) From the definition of $\bar{\mu}_t$ and (3.9) we have

$$|\bar{\mu}_t - \mu| = \left|\frac{\Delta N(t)}{t} - \mu\right| = \left|\delta \mu - E[\eta]\right| \frac{N(t)}{t} - \mu\right| \leq \mu \left|\frac{\delta N(t)}{t} - 1\right| + E[\eta] \frac{N(t)}{t}. \quad (3.14)$$

Taking expectations and applying (3.11) proves (i).

(ii) Due to Theorem 3.1(i), it suffices to show that

$$|\bar{\mu}_t - \mu| = \frac{1}{\tau(t)} \sum_{n=1}^{N(t)} \eta_n \xrightarrow{p} 0, \quad (3.15)$$

or equivalently, due to (3.12), that $(\sum_{n=1}^{N(t)} \eta_n)/t \xrightarrow{p} 0$ as $\delta, t \to \infty$ so that $\delta = o(t)$. Indeed,

$$\frac{1}{t} E\left[\sum_{n=1}^{N(t)} \eta_n\right] \leq \frac{1}{t} E\left[\sum_{n=1}^{\nu(t)} \eta_n\right] = \frac{E[\nu(t)] E[\eta]}{t} = O\left(\frac{1}{\delta}\right), \quad (3.16)$$

where the inequality is due to $\nu(t) = N(t) + 1$, the first equality follows from Wald’s first identity and the second one from (3.11) and the fact that $E[\eta] = O(1)$ as $\Delta \to \infty$. 

Due to Theorem 3.1(i), it clearly suffices to show that \[
\sqrt{t}|\hat{\mu}_t - \hat{\mu}_t| \xrightarrow{p} 0
\]
or equivalently, due to (3.12) and (3.15), that \[
\frac{\sum_{n=1}^{N(t)} \eta_n}{\sqrt{t}} \xrightarrow{p} 0
\] so that \[
\sqrt{t} \ll \delta < t,
\]
which follows directly from (3.16).

Remarks: (1) Theorem 3.2 shows that a sufficiently low sampling rate (or, equivalently, a sufficiently large threshold \(\Delta\)) is needed for \(\hat{\mu}_t\) to preserve the asymptotic distribution (3.11) and for \(\tilde{\mu}_t\) to be \(\sqrt{t}\)-consistent. This is quite intuitive, since small values of \(\Delta\) may lead to frequent sampling, but they also lead to fast accumulation of large unobserved overshoots, thus they intensify the related performance loss. On the contrary, large thresholds guarantee that the relative size of the overshoots will be small, mitigating in this way the corresponding performance loss.

(2) Theorem 3.2 and the discussion prior to it remain valid, with obvious modifications, when the sign of \(\mu\) is unknown, as long as (3.6) and (3.13) are replaced by
\[
\tau_n = \inf\{t \geq \tau_{n-1} : |S_t - S_{\tau_{n-1}}| \geq \Delta\}; \quad \tau_0 = 0
\]
and
\[
\hat{\mu}_t = \frac{1}{\tau(t)} \sum_{n=1}^{N(t)} (2z_n - 1)
\]
and
\[
\tilde{\mu}_t = \frac{1}{t} \sum_{n=1}^{N(t)} (2z_n - 1),
\]
respectively, where \(z_n := 1(S_{\tau_n} - S_{\tau_{n-1}} \geq \Delta)\).

### 3.3. Efficient estimation via overshoot correction

The estimators \(\hat{\mu}_t\) and \(\tilde{\mu}_t\) do not require knowledge of the distribution of \(X\). However, when this distribution is known (up to the unknown parameter \(\mu\)), it should be possible to improve \(\tilde{\mu}_t\) and \(\hat{\mu}_t\) by approximating, instead of ignoring, the unobserved overshoots. In order to achieve that, a first idea is to replace each \(\eta_n\) in (3.7) and (3.8) by its expectation, \(\mathbb{E}[\eta]\). However, since the latter is typically an intractable quantity, we could use the limiting average overshoot instead. Indeed, it is well known from classical renewal theory (see, e.g., [15], p. 105) that if \(X\) is non-lattice, then
\[
\lim_{\Delta \to \infty} \mathbb{E}[\eta] = \rho(\mu) := \frac{\mathbb{E}[S_{\tau_+}^2]}{2 \mathbb{E}[S_{\tau_+}]},
\]
where \(\tau_+ := \inf\{t \in \mathbb{N} : S_t > 0\}\) is the first ascending ladder time (and, consequently, \(S_{\tau_+}\) is the first ascending ladder height). Note that we have expressed the limiting average overshoot as a function of \(\mu\) in order to emphasize that...
it depends on the unknown parameter, $\mu$, thus, it cannot be used directly to approximate the unobserved overshoots. However, we can obtain a working approximation of $\rho(\mu)$ if we replace $\mu$ with an estimator that does not require knowledge of the distribution of $X$, such as $\hat{\mu}_t$ or $\tilde{\mu}_t$. Doing so, we obtain

$$
\frac{1}{t} \sum_{n=1}^{N(t)} \left[ \Delta + \rho(\hat{\mu}_t) \right] = N(t) \frac{\Delta + \rho(\hat{\mu}_t)}{t} = g(\hat{\mu}_t),
$$

(3.18)

$$
\frac{1}{\tau(t)} \sum_{n=1}^{N(t)} \left[ \Delta + \rho(\tilde{\mu}_t) \right] = N(t) \frac{\Delta + \rho(\tilde{\mu}_t)}{\tau(t)} = g(\tilde{\mu}_t),
$$

where $g(x) := x(1 + \rho(x)/\Delta)$, $x > 0$. Note that the factor in the parenthesis reflects the overshoot correction that is achieved by the suggested approximation.

In the next theorem we show that, under certain conditions on the distribution of $X$, $g(\hat{\mu}_t)$, unlike $\hat{\mu}_t$, can preserve the asymptotic distribution (3.1), whereas $g(\tilde{\mu}_t)$ attains it for a wider range of sampling rates than $\tilde{\mu}_t$.

**Theorem 3.3.** Suppose that $X$ is non-lattice and

(A1) $\mathbb{E}[|X|^{r+1}] < \infty$ for some $r \geq 2$,

(A2) $|\mathbb{E}[\eta] - \rho(\mu)| = O(\Delta^{-(r-1)})$ as $\Delta \to \infty$ for some $r \geq 2$,

(A3) $\mu \to \rho(\mu)$ is Lipschitz function.

Then, (i) as $\delta, t \to \infty$ so that $t^{1/4} << \delta << \sqrt{t}$

$$
g(\hat{\mu}_t) - \mu \frac{\sigma}{\sqrt{t}} \to \mathcal{N}(0,1)
$$

and (ii) as $\delta, t \to \infty$ so that $t^{1/4} << \delta << t$

$$
g(\tilde{\mu}_t) - \mu \frac{\sigma}{\sqrt{t}} \to \mathcal{N}(0,1).
$$

Before we prove this theorem, let us first comment on its assumptions. (A2) describes how fast the expected overshoot should converge to its limiting value as $\Delta \to \infty$. We will show in Subsection 3.3.1 that (A2) is implied by (A1) when $X > 0$ or by an exponentially decaying right tail of $X$. (A3) guarantees that if $\hat{\mu}_t$ (or $\tilde{\mu}_t$) is a “good” estimator of $\mu$, then so will $\rho(\hat{\mu}_t)$ (or $g(\tilde{\mu}_t)$) be for $\rho(\mu)$. Sufficient conditions for (A3) are presented in Subsection 3.3.2. Finally, (A1), which implies that $X$ must have at least a finite third moment, is needed for two reasons; it guarantees that $\mathbb{E}[\eta^2] = O(1)$ as $\Delta \to \infty$ and at the same time it allows us to apply Theorems 2.2 and 2.3 for $r = 2$ and obtain the asymptotic behavior of $\mathbb{E}[\nu^2(t)]$ and $\mathbb{E}[N^2(t)]$ as $\delta, t \to \infty$ so that $\delta = o(t)$. These two properties are summarized in the following lemma.
Lemma 3.1. (i) If $E[(X^+)^3] < \infty$, then $E[\eta^2] = \mathcal{O}(1)$ as $\Delta \to \infty$.
(ii) If $E[(X^-)^3] < \infty$, then as $\delta, t \to \infty$ so that $\delta = o(t)$

$$E\left[\left|\frac{\delta N(t)}{t} - 1\right|^2\right] = \mathcal{O}\left(\frac{\delta}{t}\right)^2 \quad \text{and} \quad E\left[\left|\frac{\delta \nu(t)}{t} - 1\right|^2\right] = \mathcal{O}\left(\frac{\delta}{t}\right)^2. \quad (3.19)$$

Proof. (i) is well-known from renewal theory (see, e.g., Theorem 10.9 in [15]).

Now, from Theorem 8.1 in [15] and the fact that $\delta = \Theta(\Delta)$ it follows that if $E[(X^-)^3] < \infty$, then $E[\tau^3] = (\Delta/\mu)^3(1 + o(1)) = \mathcal{O}(\delta^3)$. Moreover, from (3.10) we know that $\text{Var}[\tau] = \Theta(\delta)$, therefore, (ii) follows from Theorems 2.2 and 2.3.

Proof of Theorem 3.3. (i) From Theorem 3.1(ii) it is clear that it suffices to show that $\sqrt{t} |\bar{\mu}_t - g(\bar{\mu}_t)| \to 0$ as $\delta, t \to \infty$ so that $t^{1/4} << \delta << t$.

In particular, since

$$\sqrt{t} |\bar{\mu}_t - g(\bar{\mu}_t)| = \frac{1}{\sqrt{t}} \left| \sum_{n=1}^{N(t)} (\eta_n - \rho(\bar{\mu}_t)) \right| \leq \frac{1}{\sqrt{t}} \left| \sum_{n=1}^{N(t)} (\eta_n - \text{E}[\eta]) \right| + \frac{N(t)|\text{E}[\eta] - \rho(\mu)|}{\sqrt{t}} + \frac{N(t)|\rho(\bar{\mu}_t) - \rho(\mu)|}{\sqrt{t}}, \quad (3.20)$$

it suffices to show that each term in (3.20) converges to 0 in probability when $t^{1/4} << \delta << t$. We start with the first term and we observe that, since $\nu(t) = N(t) + 1$, using the triangle inequality we can write

$$\left| \sum_{n=1}^{N(t)} (\eta_n - \text{E}[\eta]) \right| \leq \left| \sum_{n=1}^{\nu(t)} (\eta_n - \text{E}[\eta]) \right| + |\eta_{\nu(t)} - \text{E}[\eta]|.$$

Taking expectations and applying the Cauchy-Schwartz inequality to both terms of the right-hand side we obtain

$$E\left[\left| \sum_{n=1}^{\nu(t)} (\eta_n - \text{E}[\eta]) \right|^2\right] \leq \sqrt{\text{Var}\left[\sum_{n=1}^{\nu(t)} (\eta_n - \text{E}[\eta])\right]} + \sqrt{E\left[(\eta_{\nu(t)} - \text{E}[\eta])^2\right]}$$

$$\leq \sqrt{\text{Var}\left[\sum_{n=1}^{\nu(t)} (\eta_n - \text{E}[\eta])\right]} + \sqrt{E\left[\left|\sum_{n=1}^{\nu(t)} (\eta_n - \text{E}[\eta])\right|^2\right]}$$

$$= \sqrt{\text{E}[\nu(t)]\text{Var}[\eta]} + \sqrt{\text{E}[\nu(t)]\text{Var}[\eta]},$$

where the second inequality is trivial and the equality follows from an application of Wald’s identities. But from (3.11) we have $E[\nu(t)] = \mathcal{O}(t/\delta)$ as $\delta, t \to \infty$ so
that \( \delta = o(t) \), whereas from Lemma 3.1(i) we have \( \text{Var}[\eta] = \mathcal{O}(1) \) as \( \Delta \to \infty \). Therefore,

\[
\frac{1}{\sqrt{t}} \mathbb{E} \left[ \sum_{n=1}^{\nu(t)} (\eta_n - \mathbb{E}[\eta]) \right] = \mathcal{O} \left( \frac{1}{\sqrt{\delta}} \right).
\] (3.21)

Regarding the second term in (3.20), from (A2) and the fact that \( \delta = \Theta(\Delta) \) as \( \Delta \to \infty \) it follows that \( |\mathbb{E}[\eta] - \rho(\mu)| = \mathcal{O}(\delta^{-(r-1)}) \) as \( \delta \to \infty \). Moreover, from (3.11) it follows that \( \mathbb{E}[N(t)] = \mathcal{O}(t/\delta) \) as \( \delta, t \to \infty \) so that \( \delta = o(t) \). Therefore,

\[
\frac{\mathbb{E}[N(t)] |\mathbb{E}[\eta] - \rho(\mu)|}{\sqrt{t}} = \mathcal{O}(t/\delta) \frac{\mathcal{O}(\delta^{-(r-1)})}{\sqrt{t}} = \mathcal{O} \left( \frac{\sqrt{t}}{\delta^r} \right).
\] (3.22)

Regarding the last term in (3.20), from the assumption that \( \mu \to \rho(\mu) \) is Lipschitz, there is a constant \( C > 0 \) so that

\[
|\rho(\mu_2) - \rho(\mu_1)| \leq C |\mu_2 - \mu_1|, \quad \forall \mu_1, \mu_2 > 0,
\]

and, as a result,

\[
\frac{N(t) |\rho(\hat{\mu}_t) - \rho(\mu)|}{\sqrt{t}} \leq C \frac{N(t) |\hat{\mu}_t - \mu|}{\sqrt{t}} \leq C \frac{N(t) |\delta N(t)/t - 1|}{\sqrt{t}} + C \mathbb{E}[\eta] \frac{N^2_t}{t^{3/2}}.
\]

where the second inequality follows from (3.14). Then, taking expectations and applying the Cauchy-Schwartz inequality in the first term we obtain

\[
\mathbb{E} \left[ \frac{N(t) |\rho(\hat{\mu}_t) - \rho(\mu)|}{\sqrt{t}} \right] \leq C \sqrt{\mathbb{E}[N^2_t]} \left[ \frac{\mathbb{E}[\delta N(t)/t]}{t} - 1 \right] + C \frac{\mathbb{E}[N^2_t]}{t^{3/2}}
\]

\[
= \mathcal{O} \left( \frac{1}{\sqrt{\delta}} \right) + \mathcal{O} \left( \frac{\sqrt{t}}{\delta^r} \right) = \mathcal{O} \left( \frac{\sqrt{t}}{\delta^2} \right).
\] (3.23)

The second equality follows from (3.19), which implies that \( \mathbb{E}[N^2_t] = \mathcal{O}(t^2/\delta^2) \).

From (3.20) - (3.23) we obtain

\[
\sqrt{t} |\bar{\mu}_t - g(\hat{\mu}_t)| = \mathcal{O} \left( \frac{1}{\sqrt{\delta}} \right) + \mathcal{O} \left( \frac{\sqrt{t}}{\delta^r} \right) + \mathcal{O} \left( \frac{\sqrt{t}}{\delta^2} \right),
\] (3.24)

which completes the proof, since the last term in the right hand side is the dominant one.

(ii) Due to Theorem 3.1(i), it suffices to show that \( \sqrt{t} |\bar{\mu}_t - g(\hat{\mu}_t)| \overset{p}{\to} 0 \) as
\( \delta, t \to \infty \) so that \( t^{1/4} \ll \delta \ll t \). Since

\[
\sqrt{t} |\mu_t - g(\hat{\mu}_t)| = \frac{\sqrt{t}}{\tau(t)} \left| \sum_{n=1}^{N(t)} (\eta_n - \rho(\hat{\mu}_t)) \right| = \frac{t}{\tau(t)} \sqrt{t} \sum_{n=1}^{N(t)} (\eta_n - \rho(\hat{\mu}_t)) \\
\leq \frac{t}{\tau(t)} \left[ \frac{1}{\sqrt{t}} \sum_{n=1}^{N(t)} (\eta_n - \rho(\hat{\mu}_t)) + \frac{N(t) |\rho(\hat{\mu}_t) - \rho(\hat{\mu}_t)|}{\sqrt{t}} \right], \quad (3.25)
\]

from (3.12) and (3.24) it is clear that it suffices to show that the second term in the parenthesis in (3.25) converges to 0 in probability as \( \delta, t \to \infty \) so that \( \delta = o(t) \). Indeed, due to the assumption that \( \mu \to \rho(\mu) \) is Lipschitz, we have

\[
\frac{N(t) |\rho(\hat{\mu}_t) - \rho(\hat{\mu}_t)|}{\sqrt{t}} \leq C \frac{N(t)}{\sqrt{t}} \frac{\Delta N(t)}{\tau(t)} \frac{t \tau(t)}{t} = C \frac{\Delta}{\delta} \tau(t) \left( \frac{\tau(t)}{t} \right)^2 \frac{t \tau(t)}{\delta \sqrt{t}}.
\]

From (3.19) and (2.15) it is clear that the right-hand side converges to 0 in probability as \( \delta, t \to \infty \) so that \( \delta = o(t) \), which completes the proof. \( \square \)

### 3.3.1. Sufficient conditions for (A2)

In order to find sufficient conditions for (A2), we appeal to Stone’s refinements of the renewal theorem \([24], [25]\). Thus, let \( U(\Delta) \) be the renewal function and \((\tau_n, +)_{n \in \mathbb{N}}\) the ascending ladder times of the random walk \((S_t)_{t \in \mathbb{N}}\), i.e.,

\[
U(\Delta) := \sum_{n=0}^{\infty} P(S_{\tau_n, +} \leq \Delta) \quad \text{and} \quad \tau_{n, +} := \inf \{ t \geq \tau_{n-1, +} : S_t > S_{\tau_{n-1, +}} \}; \quad \tau_{0, +} := 0.
\]

From \([24]\) we know that if \( X \) is strongly non-lattice, i.e.,

\[
\liminf_{|\theta| \to \infty} |1 - E[e^{i \theta X}]| > 0
\]

and (A1) holds, i.e., \( E[|X|^{2+r}] < \infty \) for some \( r \geq 2 \), then

\[
\left| U(\Delta) - \frac{\Delta}{\mu} - \frac{\rho(\mu)}{\mu} \right| = o\left( \frac{1}{\Delta^{r-1}} \right), \quad (3.26)
\]

Moreover, from \([25]\) it follows that if \( X \) is strongly non-lattice and it has an exponentially decaying right tail, in the sense that \( \mathbb{P}(X > x) = o(e^{-\theta_1 x}) \) as \( x \to \infty \) for some \( \theta_1 > 0 \), then

\[
\left| U(\Delta) - \frac{\Delta}{\mu} - \frac{\rho(\mu)}{\mu} \right| = o(e^{-\theta_2 \Delta}) \quad \text{as} \quad \Delta \to \infty \quad (3.27)
\]

for some \( \theta_2 > 0 \). Based on these results, we have the following lemma that provides sufficient conditions for (A2).
Lemma 3.2. (A2) is satisfied when $X$ is strongly non-lattice and one of the following holds

(i) $P(X > 0) = 1$ and $E[|X|^{2+r}] < \infty$ for some $r \geq 2$,

(ii) $P(X > x) = o(e^{-\theta_1 x})$ as $x \to \infty$ for some $\theta_1 > 0$.

Proof of Lemma 3.2. We observe that

$$|E[\eta] - \rho(\mu)| = |E[S_\tau - \Delta] - \rho(\mu)| = \mu \left| \frac{\Delta}{\mu} - \frac{\rho(\mu)}{\mu} \right|,$$

where the second equality follows from an application of Wald’s identity. When $P(X > 0) = 1$, then $E[\tau] = U(\Delta)$ and the Lemma follows from (3.26). When $X$ has an exponentially decaying right tail, using (3.27) and the following representation of the expected overshoot

$$E[\eta] = \int_0^\infty \int_0^\infty \left( 1 - P(S_\tau + > \Delta + y - x) \right) U(dx) dy,$$

we can show, working similarly to Chang [6], p. 723, that $|E[\eta] - \rho(\mu)| \rightarrow 0$ exponentially fast as $\Delta \rightarrow \infty$, which of course implies (A2).

3.3.2. Sufficient conditions for (A3)

When $P(X > 0) = 1$, from (3.17) it follows that

$$\rho(\mu) = \frac{E[X^2]}{2E[X]} = \frac{\mu^2 + \sigma^2}{2\mu}.$$

Therefore, for (A3) to hold, $\sigma^2/\mu$ has to be Lipschitz as a function of $\mu$. Clearly, this is not the case when $\sigma$ is independent of $\mu$, unless $\mu$ is restricted on a compact interval. However, when $\sigma^2 = c\mu^2$ for some $c > 0$, then $\rho(\mu) = [(1 + c)/2] \mu$ and (A3) is satisfied. For example, if $X$ follows the Gamma distribution with shape parameter $k > 0$ and rate parameter $\lambda > 0$, then $\sigma^2 = \mu^2/k$ and $\rho(\mu)$ is proportional to $\mu$ for any given $k$.

When $P(X > 0) < 1$, $\rho(\mu)$ does not typically admit a convenient closed-form expression in terms of $\mu$. An exception is the Gaussian distribution, in which case (3.17) takes the following form

$$\rho(\mu) = \frac{\mu^2 + \sigma^2}{2\mu} - \sigma \sum_{n=1}^{\infty} \frac{\phi(b_n) - b_n \Phi(-b_n)}{\sqrt{n}}, \quad b_n := \frac{\mu \sqrt{n}}{\sigma},$$

(see [29], p. 34), where $\Phi(\cdot)$ and $\phi(\cdot)$ are the c.d.f. and p.d.f. respectively of the standard normal distribution. In this case as well, $\sigma^2 = c\mu^2$ for some $c > 0$. 

\[\square\]
implies \( b_n = \sqrt{n/c} \) and, consequently, \( \rho(\mu) = w_c \mu \), where

\[
w_c := \frac{1 + c}{2} - \sum_{n=1}^{\infty} \phi(\sqrt{n/c}) - \sqrt{n/c} \Phi(-\sqrt{n/c})/\sqrt{n/c}.
\]

3.4. Summary and simulation experiments

In Table 1, we present the estimators of \( \mu \) that we have considered so far in the context of the sampling scheme (3.6). For each estimator, we report whether it requires knowledge of the overshoots (\( \eta_n \)) and/or the distribution of \( X \) and we present the sampling rates for which its performance is optimized.

In Figure 3.4 we plot the relative efficiency of each estimator as a function of the sampling period, \( \delta \), for a fixed horizon of observations, \( t = 300 \), when \( X \sim \mathcal{N}(\mu, c\mu^2) \) and \( c = \mu = 4 \). More specifically, we define the relative efficiency of \( \bar{\mu}_t \) (similarly for the other estimators) as the ratio of its mean square error over the mean square error of \( S_t/t \),

\[
RE(\bar{\mu}_t) := \frac{E[(\bar{\mu}_t - \mu)^2]}{E[(\mu_t - \mu)^2]} = \frac{E[(\bar{\mu}_t - \mu)^2]}{\sigma^2/t}
\]

and we compute it using simulation experiments.

The findings depicted in this figure verify our asymptotic results. First of all, we observe that \( \bar{\mu}_t, \hat{\mu}_t \) and \( g(\hat{\mu}_t) \) are more efficient than \( \tilde{\mu}_t, \hat{\mu}_t \) and \( g(\hat{\mu}_t) \), respectively, for any \( \delta \) and especially for large values of \( \delta \). Thus, it is always preferable to use \( \tau(t) \) in the denominator, instead of \( t \), especially when one is interested in large sampling periods.

Moreover, we see that for \( \bar{\mu}_t \) and \( \hat{\mu}_t \), the smaller the sampling period \( \delta \), the better the performance; for \( \hat{\mu}_t \) and \( g(\hat{\mu}_t) \), performance improves (and eventually remains flat) as \( \delta \) increases; for \( \bar{\mu}_t \) and \( g(\bar{\mu}_t) \), performance is optimized when \( \delta \) is in a particular range and deteriorates for both very small and very large values of \( \delta \). Therefore, when the random walk (\( S_t \)) is fully observed at the sampling times (\( \tau_n \)), it is preferable to have a high sampling rate, but when the overshoots (\( \eta_n \)) are unobserved and \( \tau(t) \) (resp. \( t \)) is used in the denominator, the sampling rate should be low (resp. moderate).

Finally, we observe that \( g(\tilde{\mu}_t) \) is more efficient than \( \tilde{\mu}_t \) uniformly over \( \delta \), whereas its performance attains its optimum over a much wider range for \( \delta \). On the other hand, \( g(\hat{\mu}_t) \) is more efficient than \( \hat{\mu}_t \) only for small \( \delta \). For very large values of \( \delta \), \( \hat{\mu}_t \) turns out to be more efficient even than \( \tilde{\mu}_t \)!

Therefore, when using \( \tau(t) \) (resp. \( t \)) in the denominator, approximating the overshoots is beneficial only for high sampling rates (resp. for any sampling rate and especially for low sampling rates).
Fig 1. Relative efficiency against sampling period for each of the estimators in Table 1 in a curved Gaussian distribution. The horizon of observations is \( t = 300 \), the true mean is \( \mu = 4 \) and the variance \( \sigma^2 = 4\mu^2 \). The computations are based on simulation experiments.

### Table 1

| Estimator | Formula | Distribution | Overshoots | Optimal Rate |
|-----------|---------|--------------|------------|--------------|
| \( \bar{\mu}_t \) | \( S_{\tau(t)}/\tau(t) \) | no | yes | \( \delta << t \) |
| \( \hat{\mu}_t \) | \( S_{\tau(t)}/t \) | no | yes | \( \delta << \sqrt{t} \) |
| \( \tilde{\mu}_t \) | \( N(t)\Delta/\tau(t) \) | no | no | \( \sqrt{t} << \delta << t \) |
| \( g(\bar{\mu}_t) \) | \( g(\bar{\mu}_t) \) | yes | no | \( \delta \sim \sqrt{t} \) |
| \( g(\hat{\mu}_t) \) | \( g(\hat{\mu}_t) \) | yes | no | \( \delta \sim \sqrt{t} \) |
| \( g(\tilde{\mu}_t) \) | \( g(\tilde{\mu}_t) \) | yes | no | \( \delta \sim \sqrt{t} \) |

3.5. Estimating the standard deviation

In order to attach an (asymptotic) standard error to the estimators we considered in this section when \( \sigma \) is unknown, we need a consistent estimator of \( \sigma \), which is not possible to obtain using only the sampling times \( (\tau_n) \). If, however, in addition to \( (\tau_n) \) we also observe the following stopping times:

\[
\theta_n := \inf\{t \geq \theta_{n-1} : Z_t - Z_{\theta_{n-1}} \geq \Gamma\}, \quad n \in \mathbb{N},
\]
where $Z_t := (X_1)^2 + \ldots + (X_t)^2$ and $\Gamma > 0$, then we can estimate $\sigma$ at some arbitrary time $t \geq \theta_1$ with

$$\hat{\sigma}_t := \sqrt{\frac{\Gamma M(t)}{\theta(t)}} - (\hat{\mu}_t)^2,$$

where $M(t) := \max \{n : \theta_n \leq t\}$ and $\theta(t) := \theta_{M(t)}$. The following theorem shows that $\hat{\sigma}_t$ is a consistent estimator of $\sigma$ under low-rate sampling and, consequently, it implies that Theorems 3.2(iii) and 3.3 remain valid if we replace $\sigma$ by $\hat{\sigma}_t$.

**Theorem 3.4.** If $E[|X|^3] < \infty$, then $\hat{\sigma}_t \overset{p}{\to} \sigma$ as $\Delta, \Gamma, t \to \infty$ so that $\Delta, \Gamma = o(t)$.

**Proof.** Due to Theorem 3.2(ii), it suffices to prove that

$$\frac{\Gamma M(t)}{\theta(t)} \overset{p}{\to} E[X^2] = \sigma^2 + \mu^2$$

as $\Gamma, t \to \infty$ so that $\Gamma = o(t)$. Under a third moment assumption, this can be done in exactly the same way as we showed that $\hat{\mu}_t \overset{p}{\to} \mu$ when $\Delta, t \to \infty$ so that $\Delta = o(t)$ in Theorem 3.2(ii).

4. Decentralized parameter estimation

We now apply the results of Section 3 to the decentralized parameter estimation that we described in Subsection 1.3 of the Introduction. Thus, we consider $K$ sensors, dispersed at various locations, so that the observations at each sensor $k$, $(X_k^t)_{t \in \mathbb{N}}$, are i.i.d. with unknown mean $\mu$ and standard deviation $\sigma_k$, $k = 1, \ldots, K$. All sensors communicate with a fusion center, whose goal is to estimate $\mu$. When $\sigma_1, \ldots, \sigma_k$ are known and each sensor transmits its exact observation $X_k^t$ at every time $t$, the fusion center can use the best linear estimator of $\mu$ at any time $t$, which is given by (1.5), and whose asymptotic distribution as $t \to \infty$ is given by (1.6), under the assumption of independence across sensors. However, due to bandwidth and energy constraints, the sensors are not, typically, able to transmit their complete observations to the fusion center. Instead, they should ideally transmit, infrequently, low-bit messages. Our goal in this section is to show that it is possible to construct estimators that preserve asymptotic distribution (1.6), even under such severe communication constraints. Indeed, assuming for simplicity that $\mu > 0$, each sensor $k$ needs to communicate with the fusion center at the following sequence of stopping times

$$\tau_k^n := \inf \{t \geq \tau_{n-1} : S_t^k - S_{\tau_{k,n-1}}^k \geq \Delta_k\}, \quad n \in \mathbb{N}; \quad \tau_0^k := 0,$$  \quad (4.1)
where $\Delta^k > 0$ is a fixed threshold. Then, two natural estimators of $\mu$ at some time $t$ are given by

$$
\hat{\mu}_t := \sum_{k=1}^{K} w_k \hat{\mu}_t^k, \quad \hat{\mu}_t^k := \Delta^k N^k_t / t
$$

(4.2)

$$
\hat{\mu}_t := \sum_{k=1}^{K} w_k \hat{\mu}_t^k, \quad \hat{\mu}_t^k := \Delta^k N^k_t / \tau^k(t),
$$

(4.3)

where $N^k_t = \max\{n : \tau_n < t\}$ is the number of messages transmitted by sensor $k$ up to time $t$ and $\tau^k(t) = \tau_{N^k(t)}$ the last communication time from sensor $k$. If, additionally, the form of the limiting average overshoot $\rho_k(\mu) := \lim_{\Delta^k \to \infty} \mathbb{E}[S_{\tau^k} - \Delta^k]$; $\tau^k := \tau_{1^k}$ is known for each $k$, two alternative estimators of $\mu$ are

$$
\sum_{k=1}^{K} w_k g_k(\hat{\mu}_t^k)
$$

and

$$
\sum_{k=1}^{K} w_k g_k(\hat{\mu}_t^k)
$$

where $\Delta := \max_{1 \leq k \leq K} \Delta^k$.

**Theorem 4.1.** (i) If $\Delta, t \to \infty$ so that $\sqrt{t} \ll \Delta$ and $\Delta \ll t$, then

$$
\frac{\hat{\mu}_t - \mu}{\sqrt{K \sigma / \sqrt{t}}} \to \mathcal{N}(0, 1).
$$

(ii) Suppose that the assumptions of Theorem 3.3 are satisfied by each $X^k_1$, $\rho_k(\cdot)$. If $\Delta, t \to \infty$ so that $\sqrt{t} \ll \Delta$ and $\Delta \ll t$, then

$$
\frac{\sum_{k=1}^{K} w_k g_k(\hat{\mu}_t^k) - \mu}{\sqrt{K \sigma / \sqrt{t}}} \to \mathcal{N}(0, 1).
$$

If, additionally, $\Delta \ll \sqrt{t}$, then

$$
\frac{\sum_{k=1}^{K} w_k g_k(\hat{\mu}_t^k) - \mu}{\sqrt{K \sigma / \sqrt{t}}} \to \mathcal{N}(0, 1).
$$

When $\sigma_k = \sigma$ for every $1 \leq k \leq K$, then $w_k = 1/K$ and Theorem 4.1 will remain valid if we replace $\sigma$ by a consistent estimator of it. Following Subsection 3.5, we can find such an estimator based on infrequent transmissions of 1-bit messages from all sensors. To this end, each sensor $k$ needs to communicate with the fusion center, in addition to (4.1), at the following sequence of stopping times:

$$
\theta_n^k := \inf\{t \geq \theta_{n-1}^k : Z_t^k - Z_{\theta_{n-1}^k}^k \geq \Gamma^k\}, \quad n \in \mathbb{N},
$$

where $\Gamma^k > 0$ is a fixed threshold.
where $Z_t^k := (X_t^{1^k})^2 + \ldots + (X_t^{k})^2$ and $\Gamma^k > 0$. Then,

$$\hat{\sigma}_t := \sqrt{\sum_{k=1}^K w_k \frac{\Gamma^k M_k^k}{\theta^k(t)} - \left(\tilde{\mu}_t\right)^2}$$

turns out to be a consistent estimator of $\sigma$, where $M^k(t) := \max\{n : \theta^k_n \leq t\}$ and $\theta^k(t) := \theta^k_{M^k(t)}$. This is the content of the following theorem, for which we set $\Gamma := \min_{1 \leq k \leq K} \Gamma^k$ and $\Gamma := \max_{1 \leq k \leq K} \Gamma^k$.

**Theorem 4.2.** If $E[|X_t^{k}|^3] < \infty$ and $\sigma^2 = \sigma^2$ for every $1 \leq k \leq K$, then $\hat{\sigma}_t \overset{p}{\to} \sigma$ as $\Gamma, t \to \infty$ so that $\overline{\Gamma}, \underline{\Gamma} = o(t)$.

**Proof.** It suffices to show that for every $1 \leq k \leq K$,

$$\frac{\Gamma^k M_k^k}{\theta^k(t)} \overset{p}{\to} E[(X_1^k)^2] = \sigma^2 + \mu^2$$

as $\Gamma^k, t \to \infty$ so that $\Gamma^k = o(t)$, which can be done as in Theorem 3.4. \hfill \Box

Finally, when the sign of $\mu$ is not known in advance, we need to replace the communication times (4.1) with $\tau^k_n := \inf\{t \geq \tau^k_{n-1} : |S_k^t - S_k^{\tau^k_{n-1}}| \geq \Delta^k\}$ and at each time $\tau^k_n$ sensor $k$ needs to transmit the 1-bit message $z^k_n = 1(S_k^\tau^k_n - S_k^{\tau^k_{n-1}} \geq \Delta^k)$. Then, the estimators (4.2) and (4.3) are generalized into (1.7) and (1.8), respectively, and Theorem 4.1(ii) remains valid, however, the extension of Theorem 4.1(ii) is not straightforward.

## 5. Conclusions

In the present work, we extended some fundamental renewal-theoretic results for renewal processes whose rates go to 0 and for random walks whose drifts go to infinity. We applied these extensions to a problem of parameter estimation subject to communication constraints, but we believe that they can be useful in a variety of setups where recurrent events occur infrequently. It remains an open problem to examine under what conditions, if any, other classical results from renewal theory remain valid in such a low-rate setup.

### References

[1] Anscombe, F.J. (1952) Large-sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* 48: 600-607.

[2] Asmussen, S. (2003) *Applied Probability and Queues*, 2nd ed. Springer-Verlag, New York.
[3] Aït-Sahalia, Y. and Mykland, P.A. (2004). Estimators of diffusions with randomly spaced discrete observations: A general theory, Ann. Stat. 32 2186-2222.

[4] Aït-Sahalia, Y. and Jacod, J. (2011). Testing whether jumps have finite or infinite activity. Ann. Stat. 39 1689–1719.

[5] Blanchet, J., and Glynn, P. (2006) Corrected diffusion approximations for the maximum of light-tailed random walk. Ann. Appl. Probab. 16 952–983.

[6] Chang, J. T. (1992). On moments of the first ladder height of random walks with small drift. Ann. Appl. Probab. 2 714-738.

[7] Chang, J. and Peres, Y. (1997) Ladder heights, Gaussian random walks and the Riemann zeta function. Ann. Prob. 25 787–802.

[8] Chow, Y.S. and Robbins, H. (1963) A renewal theorem for random variables which are dependent or nonidentically distributed. Ann. Math. Statist. 34 390–395.

[9] Chow, Y.S. (1966) On the moments of some one-sided stopping rules. Ann. Math. Statist. 37 382–387.

[10] Doob, J.L. (1948). Renewal theory from the point of view of the theory of probability. Trans. Amer. Math. Soc. 63 422–438.

[11] Fellouris, G. (2012). Asymptotically optimal parameter estimation under communication constraints. Ann. Stat. 40(4) 2239-2265.

[12] Florescu, I., Mariani, C. and Viens, F.G. (Eds.) Handbook of Modeling High Frequency Data in Finance. Publisher: World Scientific Publishing Co Pte Ltd.

[13] Foresti, G.L., Regazzoni, C.S. and P.K. Varshney, P.K. (Eds.) (2003). Multisensor surveillance systems: The fusion perspective, Kluwer Academic Publishers.

[14] Gut, A. (1974) On the moments and limit distributions of some first passage times. Ann. Probab. 2 277-308.

[15] Gut, A. (2008). Stopped Random Walks, Second Edition. Springer.

[16] Hatori, H. (1959) Some theorems in extender renewal theory I. Kodai Math. Sem. Rep 11, 139–146

[17] Lorden, G. (1970). On excess over the boundary, Ann. Math. Stat. 41(2) 520–527.

[18] Msechu, E.J. and Giannakis, G.B. (2012). Sensor-centric data reduction for estimation with WSNs via censoring and quantization. IEEE Trans. Sig. Proc., 60(1) 400–414.

[19] Rosenbaum, M. and Tankov, P. (2011). Asymptotic results for time-
changed Lévy processes sampled at hitting times, *Stochastic Processes and Their Applications* **121** 1607–1632.

[20] **Raghunathan, V., Schurgers, C., Park, S. and Srivastava, M.B.** (2002). Energy-aware wireless microsensor networks. *IEEE Sig. Proc. Mag.* **19**(2) 40–50.

[21] Ribeiro, A. and Giannakis, G.B. (2006) Bandwidth-Constrained Distributed Estimation for Wireless Sensor NetworksPart II: Unknown Probability Density Function. *IEEE Trans. Sig. Proc.*, **54**(7) 2784–2796.

[22] **Ross, S.M.** (1996). *Stochastic Processes, Second Edition*. Wiley.

[23] **Siegmund, D.** (1979). Corrected diffusion approximations in certain random walk problems. *Adv. in Applied Prob.* **11** 701–719.

[24] **Stone, C.J.** (1965). On characteristic functions and renewal theory. *Trans. Amer. Math. Soc.* **120** 327-342.

[25] **Stone, C.J.** (1965). On moment generating functions and renewal theory. *Ann. Math. Stat.* **36** 1298-1301.

[26] **Tudor, C. and Viens, F.G.** (2009). Variations and estimators for selfsimilarity parameter through Malliavin calculus. *Ann. Prob.** **37**(6) 2093–2134

[27] **Tsitsiklis, J.N.** (1990). Decentralized detection, *Advances in Statistical Signal Processing*, Greenwich, CT: JAI Press.

[28] **Xiao, J-J. and Luo, Z-Q.** (2005). Decentralized estimation in an inhomogeneous sensing environment, *IEEE Trans. Inform. Theory*, **51**(6), 2210–2219.

[29] **Woodroofe, M.** (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

[30] **Zhang, L., Mykland, P. and Aıt-Sahalia, Y.** (2005) A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* **100** 1394-1411.

[31] **Yilmaz, Y. and Wang, X.** (2012) Sequential decentralized parameter estimation under randomly observed Fisher information. *Preprint*