Statistics & Decisions 26, 263–274 (2008) / **DOI** 10.1524/stnd.2008.1016

Estimation of search tree size and approximate counting: A likelihood approach

Florian Dennert, Rudolf Grübel

Received: December 12, 2008; Accepted: March 23, 2009

Summary: We consider the problem of estimating the size of a random digital search tree on the basis of the maximal node depth observed along a specific path. We show that the maximum likelihood estimator exists and we investigate its properties. A similar problem arises in the context of approximate counting. In both cases a simple pure birth process plays a central role. We also construct confidence bounds.

1 Introduction

Let $(X_n)_{n\in\mathbb{N}_0}$ be a pure birth process with start at 0 and

$$P(X_{n+1} = k+1 \mid X_n = k) = 2^{-k} \quad \text{for all } k \in \mathbb{N}_0. \tag{1.1}$$

This simple structure arises in connection with search trees and approximate counting. In the first application X_n is the depth of the first free node along a specific path of a random binary tree generated by the digital search tree (DST) algorithm from uniform random input of size n; see [Ma92] and [SF96] and the references given there for a detailed description of the algorithm and its properties. From the large number of papers dealing with the DST algorithm [Lo87] and [DG07] are the most relevant for our present purposes. In the second application we have a counter that is incremented at random: If its current value is k then, independently on each arrival, it is increased by 1 with probability 2^{-k}, and X_n is the value of the counter after n arrivals. Again, there is a considerable amount of related literature: The basic idea appears in [Mo78]. A first detailed analysis was given in [Fl85]; extensions and related results have been discussed in [KP91], [Pr94], and elsewhere.

In the present paper we regard n as an unknown parameter. We consider statistical inference for n, where the data consists of the current value k of the process. In terms of the above applications we are therefore interested in estimation of the size of the tree on the basis of the observed level of the external node along a specific path (or, equivalently, its depth along this path), or the estimation of the total number of events on the basis

AMS 2000 subject classification: Primary: 62F10; Secondary: 60J10, 62F25, 68Q25
Key words and phrases: Birth process, confidence intervals, digital search tree algorithm, limit distribution, maximum likelihood estimation

of the current value of the counter. We suggest to use the likelihood principle. In the estimation context this leads to the maximum likelihood estimator, which we discuss in the next section. Section 3 compares the maximum likelihood estimator with a well-known unbiased estimator for n. In the last section we consider confidence intervals for n. Throughout, the approach that we suggested in [DG07], which is to consider the birth process as a nonhomogeneous renewal process, turns out to be useful.

2 The maximum likelihood estimator

In the context of point estimation the likelihood principle leads us to estimate the parameter n by the value $\hat{n}(k)$ that maximizes the likelihood function. In the discrete case this is the probability of the observed value k as a function of n,

$$L(n|k) = P(X_n = k).$$

It is enough to consider $k \in \mathbb{N}$ and $n \geq k$. Numerical evaluation of L can be based on the simple recursion

$$L(n|k) = 2^{-k+1} L(n-1|k-1) + (1-2^{-k}) L(n-1|k), \tag{2.1}$$

which is a direct consequence of (1.1). The recursion starts with

$$L(n|1) = 2^{1-n} \quad \text{for all } n \in \mathbb{N}. \tag{2.2}$$

We mention in passing that these values are all binary rationals so that they can in principle be evaluated without errors by representing the numbers involved as finite sequences of 0's and 1's.

As in [DG07] a key to our analysis is the interpretation of $(X_n)_{n\in\mathbb{N}_0}$ as the counting process associated with a sequence of independent random variables $(Y_k)_{k\in\mathbb{N}}$, where Y_k has a geometric distribution with parameter 2^{1-k} (for $k = 1$ we take this to be the distribution concentrated on the single value 1). This is possible for every pure birth process: Only transitions $k \to k+1$ are allowed, so the Markov chain is specified by its holding times Y_k in state $k-1$. Let $S_k := \sum_{l=1}^{k} Y_j$. We then have

$$P(X_n \geq k) = P(S_k \leq n) \quad \text{for all } n, k \in \mathbb{N}, \tag{2.3}$$

an argument also known as renewal inversion.

Lemma 2.1 *With the notation introduced above,*

$$L(n|k) = 2^k \, P(S_{k+1} = n+1) \quad \textit{for all } n, k \in \mathbb{N}. \tag{2.4}$$

Proof: We have

$$\begin{aligned} P(X_n = k) &= 2^k \, P(X_{n+1} = k+1, X_n = k) \\ &= 2^k \, P(X_n < k+1 \leq X_{n+1}) \\ &= 2^k \left(P(X_{n+1} \geq k+1) - P(X_n \geq k+1) \right) \\ &= 2^k \left(P(S_{k+1} \leq n+1) - P(S_{k+1} \leq n) \right) \\ &= 2^k \, P(S_{k+1} = n+1). \end{aligned}$$

□

Hence, apart from a multiplicative factor depending on k, the likelihood function is equal to the probability mass function of S_{k+1}, where S_k is the entry time into state k. From this it follows immediately that a maximizing value exists; our first theorem shows that it is also unique.

Theorem 2.2 *For each $k \in \mathbb{N}$ there is a unique $\hat{n}(k) \in \mathbb{N}$ such that*

$$L(\hat{n}(k)|k) \geq L(n|k) \quad \textit{for all } n \in \mathbb{N}.$$

Proof: For $k = 1$ this is immediate from (2.2). We now assume that $k > 1$.

The law of S_{k+1} is the convolution of geometric distributions and hence strongly unimodal; see Chapter 4 in [DJ88]. As a consequence, $n \to P(X_n = k)$ is weakly increasing for $n = 1, \ldots, n_0$ and weakly decreasing on $n \geq n_0$, where n_0 depends on k. Now suppose that the maximum likelihood estimate is not unique. Then in view of the unimodality we would have $P(S_k = n) = P(S_k = n + 1)$ for some $k \in \mathbb{N}$, $n \geq k$. Using

$$M(n, k) := \{(j_2, \ldots, j_k) \in \mathbb{N}^{k-1} : 1 + j_1 + \cdots + j_k = n\}$$

and $Y_1 \equiv 1$ we obtain

$$\begin{aligned} P(S_k = n) &= \sum_{(j_2,\ldots,j_k)\in M(n,k)} P(Y_2 = j_2, \ldots, Y_k = j_k) \\ &= \sum_{(j_2,\ldots,j_k)\in M(n,k)} \prod_{m=2}^{k} 2^{1-m}(1 - 2^{1-m})^{j_m - 1} \\ &= 2^{-k(k-1)/2} \sum_{(j_2,\ldots,j_k)\in M(n,k)} 2^{-\sum_{m=2}^{k}(m-1)(j_m-1)} \prod_{m=2}^{k} (2^{m-1} - 1)^{j_m - 1}. \end{aligned}$$

Note that the final product is an odd integer, and that $\sum_{m=2}^{k}(m-1)(j_m - 1)$ achieves its unique maximum on $M(n, k)$ for $j_2 = \cdots = j_{k-1} = 1$, $j_k = n - k + 1$. For the binary expansion of $P(S_k = n)$ this means that there is a final 1 in position

$$k(k-1)/2 + (k-1)(n-k) = (k-1)(n - k/2).$$

In the step from n to $n + 1$ this position moves $k - 1$ steps to the right, which means that $P(S_k = n)$ and $P(S_k = n + 1)$ are different. □

Table 2.1 gives in its third column the maximum likelihood estimator $\hat{n}(k)$ for various k-values, obtained via (2.1). The second column lists the values of an unbiased estimator; see the next section. The fourth column contains the associated ratios. (Here and in the following real numbers are rounded to the given level.) These suggest that asymptotically the maximum likelihood estimator is a constant multiple of the unbiased estimator.

For the investigation of this phenomenon we need the asymptotic behaviour of the random variable X_n. It has been observed by [Lo87] in the context of digital search trees

k	$2^k - 1$	$\hat{n}(k)$	$\hat{n}(k)/(2^k - 1)$
5	31	39	1.258065
8	255	325	1.274510
10	1023	1306	1.276637
14	16383	20925	1.277239
17	131071	167415	1.277285

Table 2.1 Values of the estimators.

and by [Fl85] in the context of approximate counting that there are small fluctuations in the distribution (resp. its moments) of $X_n - \lfloor \log_2 n \rfloor$ as $n \to \infty$. In [DG07] the possible limit points were related to the discretized shifts of one specific random variable. We need some properties of this random variable. For this, let Z_k, $k \in \mathbb{N}$, be independent and exponentially distributed with parameter 1 and let

$$S_\infty := \sum_{k=1}^{\infty} 2^{-k} Z_k. \tag{2.5}$$

It is easy to see that the series converges with probability 1 to a finite value. One of the results in [Lo87] is an explicit formula for the distribution function of S_∞,

$$P(S_\infty \le x) = 1 - \sum_{k=1}^{\infty} a_k \exp(-2^k x), \tag{2.6}$$

with

$$a_k := b \prod_{j=1}^{k-1} (1 - 2^j)^{-1} \text{ for all } k \in \mathbb{N}, \text{ and } b := \prod_{j=1}^{\infty} (1 - 2^{-j})^{-1}.$$

An alternative derivation is given in [DG07]. In view of the rapid decrease of the coefficients a_k this representation as a 'pseudo mixture' (note that the coefficients alternate in sign) can be used to obtain the values of the distribution function numerically; this is used in connection with Figure 3.1 below.

Theorem 2.3 *The random variable S_∞ has a density f_∞ that is strictly increasing on $(0, \zeta)$ and strictly decreasing on (ζ, ∞) for some $\zeta > 0$.*

Proof: The definition (2.5) displays the distribution of S_∞ as the weak limit of convolutions of exponential distributions. It therefore follows from the results in Section 1.4 of [DJ88] that the distribution S_∞ is strongly unimodal, which means that it has a density f_∞ that is increasing on $(0, \zeta)$ and decreasing on (ζ, ∞) for some $\zeta \ge 0$. The smoothing effect of the convolution product yields $\lim_{t \to 0} f_\infty(t) = 0$, hence we must have $\zeta > 0$.

From the representation (2.6) it follows that f_∞ can be extended analytically into the complex half-plane $\{z \in \mathbb{C} : \Re(z) > 0\}$. Suppose now that the mode of f_∞ is not unique. Then unimodality would imply that f'_∞ vanishes on some interval of positive length in this half-plane which in turn would lead to $f'_\infty \equiv 0$; this cannot possibly be the case for a probability density f_∞. The same argument shows that f_∞ is in fact strictly increasing on $(0, \zeta)$ and strictly decreasing on (ζ, ∞). □

The mode ζ of the distribution of S_∞ can be evaluated numerically as

$$\zeta = 0.63864361\cdots.$$

Our next result confirms the conjecture that we derived from the values in Table 2.1.

Theorem 2.4 *As $n \to \infty$, $\hat{n}(X_n)/(2^{X_n} - 1)$ converges in probability to the constant value 2ζ $(= 1.27728722\cdots)$.*

Proof: We know from the proof of Theorem 2.2 that the distribution of S_k has a unique mode ζ_k and that $\hat{n}(k) = \zeta_{k+1}$. It is therefore enough to show that

$$\lim_{k\to\infty} 2^{-k}\zeta_k = \zeta. \tag{2.7}$$

From Lemma 1 in [DG07] and the discussion in Section 3 of [DG07] we obtain that $2^{-k}S_k$ converges in distribution to S_∞, with S_∞ as in (2.5). For the distribution functions F_k of $2^{-k}S_k$ and F_∞ of S_∞ this means that

$$\lim_{k\to\infty} \sup_{x\geq 0} \left|F_k(x) - F_\infty(x)\right| = 0. \tag{2.8}$$

We next introduce a smoothed version of the discrete distributions: For each $k \in \mathbb{N}$ let $\tilde{F}_k$ be the distribution function of $\tilde{S}_k := 2^{-k}(S_k + U)$, where U and S_k are independent and U is uniformly distributed on the unit interval. Clearly, with $\tilde{\zeta}_k$ a mode of the distribution of $\tilde{S}_k$, we have that $\tilde{\zeta}_k$ and $2^{-k}\zeta_k$ differ by at most 2^{-k}, so that (2.7) will follow if we can show that $\tilde{\zeta}_k$ converges to ζ as $k \to \infty$. Also, as $\tilde{F}_k$ arises by linear interpolation from F_k, which is of pure jump type, we have (2.8) with $\tilde{F}_k$ instead of F_k too.

Now let $\epsilon > 0$ be given and suppose that $\tilde{\zeta}_n < \zeta - \epsilon$ for all $n \in A$ for some infinite set $A \subset \mathbb{N}$. Let

$$x_0 := \zeta - \epsilon, \quad x_1 := \zeta - \frac{\epsilon}{2}, \quad x_2 := \zeta,$$

and define κ by

$$\kappa := F_\infty(x_0) + \frac{1}{2}\left(F_\infty(x_2) - F_\infty(x_0)\right) - F_\infty(x_1). \tag{2.9}$$

As f_∞ is strictly increasing on (x_0, x_2) we must have $\kappa > 0$. Note that κ depends on ϵ. Now let $\delta := \kappa/4$ and let k_0 be such that

$$\sup_{x\geq 0} \left|\tilde{F}_k(x) - F_\infty(x)\right| \leq \delta \quad \text{for all } k \geq k_0. \tag{2.10}$$

For any $k \in A$ with $k \geq k_0$ we have that $\tilde{F}_k$ is concave on (x_0, x_2), which together with (2.10) and (2.9) implies

$$\begin{aligned}
\tilde{F}_k(x_1) &\geq \tilde{F}_k(x_0) + \frac{1}{2}\left(\tilde{F}_k(x_2) - \tilde{F}_k(x_0)\right) \\
&\geq \tilde{F}_k(x_0) + \frac{1}{2}\left(F_\infty(x_2) - F_\infty(x_0) - 2\delta\right) \\
&\geq F_\infty(x_0) + \frac{1}{2}\left(F_\infty(x_2) - F_\infty(x_0)\right) - 2\delta \\
&= F_\infty(x_1) + \kappa - 2\delta \\
&> F_\infty(x_1) + \delta,
\end{aligned}$$

which contradicts (2.10).

Put together this shows that $\liminf_{k\to\infty} 2^{-k}\zeta_k \geq \zeta - \epsilon$ for all $\epsilon > 0$. We now let ϵ tend to 0, and a final symmetry argument concludes the proof. □

3 Comparison with the unbiased estimator

It has been noted long ago that $2^{X_n} - 1$ is an unbiased estimator for n, but it is also well known that unbiasedness on its own can lead to suboptimal procedures; see e.g. Section 8.2 in [CH74]. Further, it is part of the statistical folklore that maximum likelihood estimators are asymptotically optimal 'in smooth cases'. Note, however, that in the standard setup n is the sample size and not, as in the present situation, a parameter. In particular, the estimators do not stabilize: Indeed, we use one single integer value k to estimate a parameter that is roughly of magnitude 2^k. As we have seen in the previous section, maximum likelihood leads to a procedure that estimates n by a value that is about 28% larger than the value given by the unbiased estimator. For an understanding of this phenomenon on an informal level we mention that in the tree context the estimator $2^k - 1$ is the number of nodes of the tree that has *all* its external nodes at the level observed on that particular path. While random trees generated by the DST algorithm tend to be fairly balanced the rapid increase of $k \mapsto 2^k$ forces the estimator to become small if it is to be unbiased. Note that $\log_2(2^{X_n} - 1)$ is not an unbiased estimator for $\log_2 n$.

We now consider the probability that the estimator is less than or equal to the parameter to be estimated: Asking for this to be equal to 1/2 would lead to median-unbiasedness, a property that would not be destroyed by monotone transformations. We require the asymptotic distributional behaviour of X_n as $n \to \infty$. This problem was solved in [Lo87], who discovered the fact that we have to pass to subsequences in order to obtain convergence in distribution; see also [Fl85] in the context of approximate counting. In [DG07] a representation of the limit points was obtained: If $(n(m))_{m\in\mathbb{N}}$ is a sequence of integers such that

$$\lim_{m\to\infty} n(m) = \infty \quad \text{and} \quad \lim_{m\to\infty} \{-\log_2 n(m)\} = \eta, \tag{3.1}$$

where $\{x\}$ denotes the fractional part of $x \in \mathbb{R}$, then

$$X_{n(m)} - \log_2 n(m) \;\to_{\text{distr}}\; Z_\eta := \lfloor -\log_2 S_\infty + \eta \rfloor - \eta \tag{3.2}$$

with $m \to \infty$. Here we have written $Y_n \to_{\text{distr}} Y_\infty$ for convergence in distribution of the random variables Y_n to Y_∞ as $n \to \infty$, and S_∞ is as in (2.5). For the two competing estimators this leads to

$$\log_2\bigl(2^{X_{n(m)}} - 1\bigr) - \log_2 n(m) \to_{\text{distr}} Z_\eta, \tag{3.3}$$

$$\log_2\bigl(\hat{n}(X_{n(m)})\bigr) - \log_2 n(m) \to_{\text{distr}} Z_\eta + 1 + \log_2 \zeta. \tag{3.4}$$

From this and the explicit description of the distribution of S_∞ given in (2.6) we can evaluate the limit of the probability that the unbiased estimator or maximum likelihood estimator is less than the parameter to be estimated. In particular, if (3.1) is satisfied, some standard manipulations lead to

$$\lim_{m\to\infty} P\bigl(2^{X_{n(m)}} - 1 \le n(m)\bigr) = P(S_\infty > 2^{\eta-1}), \quad 0 \le \eta < 1,$$

for the unbiased estimator and

$$\lim_{m\to\infty} P\bigl(\hat{n}(X_{n(m)}) \le n(m)\bigr) = \begin{cases} P(S_\infty > 2^{\eta-1}), & \text{if } \eta \ge 1 + \log_2 \zeta, \\ P(S_\infty > 2^{\eta}), & \text{if } \eta < 1 + \log_2 \zeta, \end{cases}$$

for the maximum likelihood estimator. The left part of Figure 3.1 shows the limit probabilities as a function of η. It turns out that for small values of η, i.e. when the parameter is just a bit larger than an integer power of 2, the probability that the unbiased estimator is too small will be rather high (about 0.8). The corresponding probability for the maximum likelihood estimator is closer to $1/2$ for such values. On the interval $[1+\log_2 \zeta, 1)$ the two curves coincide. The maximum deviation from the value $1/2$, which would correspond to median-unbiasedness, is 0.326327 for the unbiased estimator and 0.2504908 for the maximum likelihood estimator.

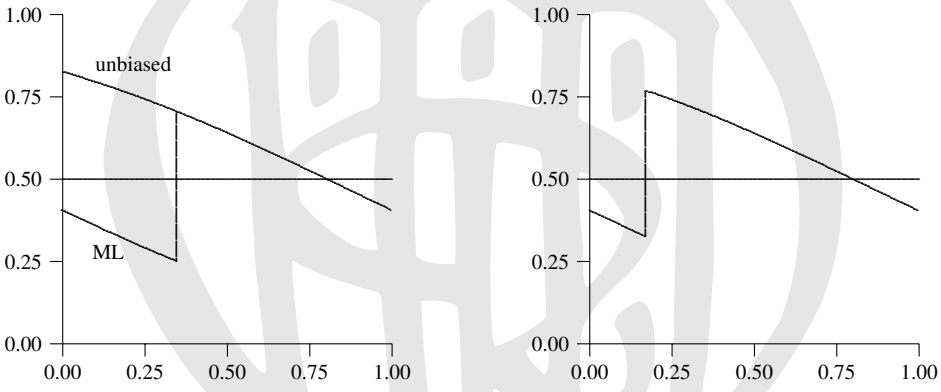

Figure 3.1 Comparison of the unbiased and the maximum likelihood estimator (see text).

Similarly, the right part of Figure 3.1 shows the limiting probability that the maximum likelihood estimator is closer to the parameter than the unbiased estimator, again as

a function of η. Again, this can be evaluated from the above limit relations (3.2) and (3.3): If (3.1) is satisfied, then

$$\begin{aligned}\lim_{m\to\infty} P\Big(\big|\hat{n}(X_{n(m)}) - n(m)\big| < \big|2^{X_{n(m)}} - 1 - n(m)\big|\Big) &= P\big(|Z_\eta + 1 + \log_2 \zeta| < |Z_\eta|\big)\\ &= P\big(Z_\eta < -c\big)\\ &= \begin{cases} P(S_\infty > 2^{\eta-1}), & \text{if } \eta \geq c,\\ P(S_\infty > 2^{\eta}), & \text{if } \eta < c,\end{cases}\end{aligned}$$

with $c := (1+\log_2 \zeta)/2 = 0.1765415\cdots$. In the first equality we have used the continuous mapping theorem, see [Bi68, Theorem 5.1], together with the continuity of the distribution function of S_∞. It turns out that, asymptotically, the maximum likelihood estimator outperforms the unbiased estimator for values of η that are between 0.1765415 and 0.8036598. Hence, from a Bayesian perspective and if we put a uniform prior on $\{\log_2 n\}$, then the subjective probability that the maximum likelihood estimator is closer to the parameter of interest will be about 0.6271183.

4 Confidence intervals

Regarding the birth process $(X_n)_{n\in\mathbb{N}_0}$ as a non-homogeneous renewal process is also useful in connection with confidence bounds. Let $\Psi_\alpha^+(k)$ and $\Psi_\alpha^-(k)$ be the 'strict' and the 'ordinary' α-quantile of the distribution of S_k,

$$\begin{aligned}\Psi_\alpha^+(k) &:= \min\big\{n \in \mathbb{N} :\ P(S_k \leq n) > \alpha\big\},\\ \Psi_\alpha^-(k) &:= \min\big\{n \in \mathbb{N} :\ P(S_k \leq n) \geq \alpha\big\}\end{aligned}$$

for $0 < \alpha < 1$, $k \in \mathbb{N}$. We know from the proof of Theorem 2.2 that $P(S_k \leq n)$ is a binary rational, so for the usual confidence levels $\alpha = 0.1, 0.05, 0.01$ the two quantiles coincide.

The following theorem shows that the Ψ-functions lead to one-sided confidence bounds that are optimal in a specific sense (we do not consider randomized bounds).

Theorem 4.1 (a) *A $100(1-\alpha)\%$ lower confidence bound for n is given by $\Psi_\alpha^+(X_n)$. Further, if $\Phi : \mathbb{N} \to \mathbb{N}$ is an increasing function with the property that*

$$P\big(n \geq \Phi(X_n)\big) \ \geq\ 1-\alpha \quad \textit{for all } n \in \mathbb{N},$$

then $\Phi(k) \leq \Psi_\alpha^+(k)$ for all $k \in \mathbb{N}$.

(b) *A $100(1-\alpha)\%$ upper confidence bound for n is given by $\Psi_{1-\alpha}^-(X_n+1)$. Further, if $\Phi : \mathbb{N} \to \mathbb{N}$ is an increasing function with the property that*

$$P\big(n \leq \Phi(X_n)\big) \ \geq\ 1-\alpha \quad \textit{for all } n \in \mathbb{N},$$

then $\Phi(k) \geq \Psi_{1-\alpha}^-(k+1)$ for all $k \in \mathbb{N}$.

Proof: For notational convenience we abbreviate Ψ_α^+ to Ψ. It follows from the properties of the sequence $(S_k)_{k\in\mathbb{N}}$ that $k \mapsto \Psi(k)$ is increasing and that $\lim_{k\to\infty} \Psi(k) = \infty$. Hence

$$\Psi^{-1} : \mathbb{N} \to \mathbb{N}, \quad \Psi^{-1}(n) := \min\{k \in \mathbb{N} : \Psi(k) \geq n\}$$

is well-defined, and it is easy to check that

$$\Psi^{-1}(n) \leq k \Leftrightarrow n \leq \Psi(k), \qquad \Psi(\Psi^{-1}(n)) \geq n$$

for all $n, k \in \mathbb{N}$. Together with the renewal inversion argument (2.3) this leads to

$$\begin{aligned} P(n \geq \Psi(X_n)) &= P(X_n < \Psi^{-1}(n+1)) \\ &= P(S_{\Psi^{-1}(n+1)} \geq n+1) \\ &\geq P\big(S_{\Psi^{-1}(n+1)} \geq \Psi(\Psi^{-1}(n+1))\big). \end{aligned}$$

Because of

$$P(S_k < \Psi^+(k)) \leq \alpha \quad \text{for all } k \in \mathbb{N}$$

this implies

$$P(n \geq \Psi^+(X_n)) \geq 1-\alpha,$$

so that $\Psi^+(X_n)$ is indeed a $100(1-\alpha)\%$ lower confidence bound for n.

Now suppose that Φ satisfies the condition in the theorem. Let $\{n_i : i \in A\}$ with $n_1 < n_2 < \cdots$ be the range of Φ (which may be finite) and let

$$k_i := \min\{k \in \mathbb{N} : \Phi(k) = n_i\}.$$

The above transformations with Φ instead of Ψ and $n = n_i - 1$ lead to $P(S_{k_i} \geq \Phi(k_i)) \geq 1-\alpha$ and hence $P(S_{k_i} \leq \Phi(k_i) - 1) \leq \alpha$. In view of

$$P(S_{k_i} \leq \Psi(k_i) - 1) \leq \alpha < P(S_{k_i} \leq \Psi(k_i))$$

this implies $\Phi(k_i) \leq \Psi(k_i)$ for all $i \in A$. As Φ and Ψ are both increasing this in turn implies $\Phi \leq \Psi$.

The proof of the second part is quite similar: With $\Psi = \Psi_{1-\alpha}^-$ we obtain

$$\begin{aligned} P(n \leq \Psi(X_n + 1)) &= P(X_n + 1 \geq \Psi^{-1}(n)) \\ &= P(S_{\Psi^{-1}(n)-1} \leq n) \\ &\geq P\big(S_{\Psi^{-1}(n)-1} \leq \Psi(\Psi^{-1}(n) - 1)\big) \\ &\geq 1-\alpha. \end{aligned}$$

The same arguments as in the first part also provide the optimality. □

We know from Section 2 that $2^{-k}S_k$ converges to S_∞ in distribution. By Theorem 2.3, the limit has a continuous and strictly increasing distribution function F_∞. Taken together this implies the convergence of the quantiles, so that

$$\lim_{k\to\infty} 2^{-k}\Psi_\alpha^\pm(k) = Q_\infty(\alpha), \tag{4.1}$$

where

$$Q_\infty(y) := \inf\{x \in \mathbb{R} : F_\infty(x) \geq y\}, \quad 0 < y < 1,$$

denotes the quantile function associated with the distribution of S_∞. Table 4.1 contains the values of the 90% confidence bounds for various k, together with approximations obtained from (4.1) and $Q_\infty(0.1) = 0.4051573$, $Q_\infty(0.9) = 1.75722$. We see that the approximate bounds are quite close to the exact bounds even for relatively small values of k and that the approximate bounds are conservative, at least for these particular k-values.

k	$\Psi^+_{0.1}(k)$	$2^k Q_\infty(0.1)$	$\Psi^-_{0.9}(k+1)$	$2^{k+1} Q_\infty(0.9)$
5	13	12.96	110	112.46
8	104	103.72	898	899.70
10	415	414.88	3597	3598.79

Table 4.1 Confidence bounds and approximations.

As usual, the one-sided bounds can be combined to obtain confidence intervals $I_\alpha(X_n)$ of the form

$$I_\alpha(k) = \big[\Psi^+_\beta(k), \Psi^-_{1-\alpha+\beta}(k+1)\big],$$

with some $\beta \in (0, \alpha)$. We now assume that $\alpha < 1/2$. A standard way to split the error probability is to choose $\beta = \alpha/2$, which results in equal-tailed confidence intervals. For example, with $\alpha = 0.05$ and $k = 7$ we obtain the interval [34, 627]. Note, however, that the upper confidence bound is 538 for these values of α and k, which because of $n \geq k$ would lead to the considerably shorter confidence interval [7, 538].

Depending on circumstances we might wish to choose β such that the length of the confidence interval is as small as possible, or we might wish to minimize the ratio of the upper and lower bound. Neither of these choices would be equivariant under strictly increasing transformations. The likelihood approach leads naturally to the idea of likelihood-based confidence regions, see e.g. [CH74, p. 218], where it is required that

$$L(n|k) \geq L(n'|k) \quad \text{for all } n \in I_\alpha(k),\ n' \notin I_\alpha(k). \tag{4.2}$$

We know from Lemma 2.1 that $n \mapsto L(n|k)$ is proportional to the probability mass function associated with S_{k+1} and we also know from the proof of Theorem 2.2 that the distribution of S_{k+1} is unimodal. This implies that (4.2) leads to regions that are in fact intervals, as in the three other cases mentioned previously. For the maximum likelihood confidence intervals we do have equivariance under strictly monotone transformations so that, for example, it does not matter whether we consider n or $\log_2 n$ as the quantity of primary interest.

The distributional asymptotics (4.1) provide an approximation for $\beta = \beta(\alpha)$ and hence for $I_\alpha(k)$ if k is large. Indeed, in the limit the requirement of minimal length or

minimal ratio leads to the problem of minimizing

$$\beta \mapsto 2 \cdot Q_\infty(1-\alpha+\beta) - Q_\infty(\beta),$$
$$\beta \mapsto \frac{Q_\infty(1-\alpha+\beta)}{Q_\infty(\beta)}$$

respectively, on the interval $(0, \alpha)$. For the likelihood intervals the corresponding asymptotic problem is that of finding $0 < x_0 < x_1 < \infty$ such that $f_\infty(x_0) = f_\infty(x_1)$ and $F_\infty(x_1) - F_\infty(x_0) = 1 - \alpha$. These problems can all be tackled by standard numerical methods; again, the series representation (2.6) turns out to be useful. In Table 4.2 the resulting β-values are given for some standard confidence levels. It turns out, for example, that the shortest length intervals are close to the intervals that consist of the deterministic lower bound and the upper $100(1-\alpha)\%$ confidence bound, in agreement with the above specific numerical example for $\alpha = 0.05$ and $k = 7$. For the minimal ratio intervals are closer to the equal-tailed case, maximum likelihood splits are in between.

α	0.10000	0.05000	0.01000
minimal length	0.00331	0.00128	0.00015
minimal ratio	0.05770	0.02941	0.00610
maximum likelihood	0.00856	0.00323	0.00037

Table 4.2 Asymptotic split probabilities for various confidence levels and interval types.

References

[Bi68] Billingsley, P. (1968) *Convergence of Probability Measures.* Wiley, New York.

[CH74] Cox, D. R. and Hinkley, D. V. (1974) *Theoretical Statistics.* Chapman and Hall, London.

[DJ88] Dharmadhikari, S. and Kumar, J. (1988) *Unimodality, Convexity, and Applications.* Academic Press, San Diego.

[DG07] Dennert, F. and Grübel, R. (2007) Renewals for exponentially increasing lifetimes, with an application to digital search trees. *Ann. Appl. Prob.* **17**, 676–687.

[Fl85] Flajolet, Ph. (1985) Approximate counting: a detailed analysis. *BIT* **25**, no. 1, 113–134.

[KP91] Kirschenhofer, P. and Prodinger, H. (1991) Approximate counting: an alternative approach. *RAIRO Inform. Théor. Appl.* **25**, 43–48.

[Lo87] Louchard, G. (1987) Exact and asymptotic distributions in digital binary search trees. *Theor. Inform. Appl.* **21**, 479–496.

[Ma92] Mahmoud, H.M. (1992) *Evolution of Random Search Trees.* Wiley, New York.

[Mo78] Morris, R. (1978) Counting a large number of events in small registers. *Comm. ACM* **21**, 840–842.

[Pr94] Prodinger, H. (1994) Approximate counting via Euler transform. Number theory (Rakova dolina, 1993). *Math. Slovaca* **44**, 569–574.

[SF96] Sedgewick, R. and Flajolet, Ph. (1996) *An Introduction to the Analysis of Algorithms.* Addison-Wesley, Reading.

Florian Dennert
Institut für Mathematische Stochastik
Leibniz Universität Hannover
Postfach 6009
30060 Hannover
Germany
dennert@stochastik.uni-hannover.de

Rudolf Grübel
Institut für Mathematische Stochastik
Leibniz Universität Hannover
Postfach 6009
30060 Hannover
Germany
rgrubel@stochastik.uni-hannover.de