Fast calculation of boundary crossing probabilities for Poisson processes

Amit Moscovich-Eiger\textsuperscript{a,}\textsuperscript{*}, Boaz Nadler\textsuperscript{a}

\textsuperscript{a}Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel.

Abstract

We present a fast $O(n^2 \log n)$ algorithm for calculating the probability that a one-dimensional Poisson process will stay within arbitrary boundaries that are bounded by $n$. This algorithm is faster than previous $O(n^3)$ methods, and can be used to compute $p$-values for continuous goodness-of-fit statistics.

Keywords: Boundary crossing, Poisson process, Empirical process, Goodness of fit, Brownian motion, First passage

1. Introduction

Let $X_1, \ldots, X_n \sim U[0, 1]$ be $n$ i.i.d. random variables. Consider the associated binomial process $\eta_n(t) = \sum_i 1(X_i \leq t)$, which is equal to $n$ times the empirical cdf of the $X_i$’s. For two arbitrary functions $g, h : [0, 1] \to \mathbb{R}$, the corresponding two-sided non-crossing probability is defined as

$$\Pr \{ \forall t \in [0, 1] : g(t) \leq \eta_n(t) \leq h(t) \}. \quad (1)$$

This quantity plays a fundamental role in a wide range of applications. Examples include the $p$-values of sup-type continuous goodness-of-fit statistics (Noé\textsuperscript{[1]}, Khmaladze and Shinjikashvili\textsuperscript{[2]}, Steck\textsuperscript{[3]}, Durbin\textsuperscript{[4]}); confidence bands for distribution functions (Owen\textsuperscript{[5]}, Frey\textsuperscript{[6]}); change-point detection (Worsley\textsuperscript{[7]}); and sequential testing (Dongchu\textsuperscript{[8]}). These applications may involve large sample sizes, hence fast procedures to compute this probability are of interest.

One popular approach is to estimate Eq. (1) using Monte-Carlo methods. In the simplest of these methods one repeatedly generates $X_1, \ldots, X_n$ and counts the number of times that the inequalities $g(t) \leq \eta_n(t) \leq h(t)$ are satisfied for all $t$. In order to obtain a probability estimate with standard error $< \epsilon$, one typically needs to repeat this procedure at least $O(1/\epsilon^2)$ times, leading to a total running time of $O(n/\epsilon^2)$. When the non-crossing probability is very close to 1 (or to 0), a particularly large number of simulations is necessary, or else this method will yield the estimate 1 (or 0).

While this approach is applicable to some settings, the focus of this paper is on the fast computation of the exact probability in Eq. (1) and an analogous probability involving the Poisson process. Currently, for general boundary functions, the fastest algorithms for computing the probability in Eq. (1) require $O(n^3)$ time (Noé\textsuperscript{[1]}, Friedrich and Schellhaas\textsuperscript{[9]}, Khmaladze and Shinjikashvili\textsuperscript{[2]}). In the one-sided case, when either $g(t) \leq 0$ or $h(t) \geq n$ for all $0 \leq t \leq 1$, the probability in Eq. (1) can be computed in $O(n^2)$ operations (Noé and Vandewiele\textsuperscript{[10]}, Kotel’Nikova and Khmaladze\textsuperscript{[11]}, Moscovich-Eiger et al.\textsuperscript{[12]}).

The main contribution of this paper is the introduction of a fast $O(n^2 \log n)$ algorithm to compute the two-sided probability in Eq. (1). This is done by investigating a closely related problem involving a homogeneous Poisson process. Specifically, let $\xi_n(t) : [0, 1] \to \{0, 1, 2, \ldots\}$ be a homogeneous Poisson process.
The following conditional probability is also of interest, the probability of this process is given by

\[ \Pr \{ \forall t \in [0, 1] : g(t) \leq \xi_n(t) \leq h(t) \} . \]  

(2)

The key observation in this paper, described in Section 2, is that the recursive solution to Eq. (3) given by Khmaladze and Shinjikashvili [2] can be described as a series of at most \( 2n \) truncated linear convolutions involving vectors of length at most \( n \). With the aid of the Fast Fourier Transform (FFT), each convolution can thus be computed in \( O(n \log n) \) operations, yielding a total running time of \( O(n^2 \log n) \).

Section 3 describes the reduction from the conditional Poisson non-crossing probability of Eq. (2) to the binomial non-crossing probability of Eq. (1). In section 4 we present an application of the proposed method to the computation of \( p \)-values for a continuous goodness-of-fit statistic. Comparing the running times of our algorithm to that of Khmaladze and Shinjikashvili [2] shows that our method yields significant speedups for large sample sizes.

Finally, we note that since Brownian motion can be described as a limit of a Poisson process, one may apply our method to approximate the boundary crossing probability and first passage time of a Brownian motion as described in Khmaladze and Shinjikashvili [2]. The latter quantity has multiple applications in finance and statistics (Siegmund [13], Chicheportiche and Bouchaud [14]). In this case a potentially practical method for estimating various quantities related to Brownian motion in 2- or 3 dimensions.

2. Boundary crossing probability for a Poisson process

In this section we describe our proposed algorithm for the fast computation of the two-sided non-crossing probability of a Poisson process, given in Eq. (2). We assume that \( g(t) \leq h(t) \) for all \( t \in [0, 1] \) and that \( g(0) \leq 0 \leq h(0) \), as otherwise the non-crossing probability is simply zero. Also, since the Poisson process is monotone, we may assume w.l.o.g. that \( g(t) \) and \( h(t) \) are monotone non-decreasing. We start by describing the recursion formula of Khmaladze and Shinjikashvili [2] whose direct application yields an \( O(n^3) \) algorithm. We then show how to reduce the computational cost to \( O(n^2 \log n) \) operations.

For every integer \( i \in [0, g(1)] \), let \( t^g_i = \inf \{ t \in [0, 1] : g(t) > i \} \) be the first time the function \( g(t) \) passes the integer \( i \). Similarly for every integer \( i \in [h(0), h(1)] \), let \( t^h_i = \sup \{ t \in [0, 1] : h(t) < i \} \) be the last time the function \( h(t) \) is smaller than \( i \). Let \( T(g) = \{ t^g_i \}_{0 \leq i \leq g(1)} \) and \( T(h) = \{ t^h_i \}_{h(0) \leq i \leq h(1)} \) be the set of all integer passage times for the two functions. It is easy to verify that a non-decreasing function \( f : [0, 1] \to \{0, 1, 2, \ldots\} \) satisfies \( g(t) \leq f(t) \leq h(t) \) for all \( t \in [0, 1] \) if and only if it satisfies these conditions at all discrete times \( t \in T(g) \cup T(h) \cup \{1\} \). Hence, to compute the probabilities in equations (1), (2) and (3), it suffices to analyze these inequalities only at a finite set of times.

Definition 1. For any \( s \in [0, 1] \) and \( m \in \{0, 1, 2, \ldots\} \), let \( Q(s, m) \) be the probability that \( \xi_n(s) = m \) and that the Poisson process \( \xi_n \) with intensity \( n \) does not cross the boundaries \( g(t), h(t) \) up to time \( s \).

\[ Q(s, m) := \Pr \{ \forall t \in [0, s] : g(t) \leq \xi_n(t) \leq h(t) \text{ and } \xi_n(s) = m \} . \]
Clearly $Q(0, 0) = 1$. Let $0 = t_0 \leq t_1 \leq \ldots \leq t_N = 1$ denote the sorted set of times from $T(g) \cup T(h) \cup \{1\}$. For any $i \in \{0, \ldots, N - 1\}$ and any $m \in \{0, 1, 2, \ldots\}$ the Chapman-Kolmogorov equations give

$$Q(t_{i+1}, m) = \begin{cases} \sum_\ell Q(t_i, \ell) \cdot \Pr[Z_i = m - \ell] & \text{if } g(t_{i+1}) \leq m \leq h(t_{i+1}) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $Z_i$ is a Poisson random variable with intensity $n(t_{i+1} - t_i)$ and the sum is taken over all $g(t_i) \leq \ell \leq m$. This formula was proposed by Khmaladze and Shinjikashvili [2] to compute $Q(1, n)$.

The recursive relation of Eq. (4) suggests the following procedure: Start by calculating all probabilities at time $t_1$ for $m \leq n$, namely $Q(t_1, 0), \ldots, Q(t_1, n)$. Then, calculate all probabilities at time $t_2$, etc. Since each $Q(t_i, m)$ is the sum of up to $m + 1$ terms, the total running time is at most $O(n^3)$, but may be smaller if the boundary functions $g(t), h(t)$ are close to each other.

Next, we describe a faster procedure. Let $Q_t = (Q(t_1, 0), Q(t_1, 1), \ldots, Q(t_1, n))$ and let $\pi_\lambda = (\Pr[Z = 0], \Pr[Z = 1], \ldots, \Pr[Z = n])$, where $Z \sim \text{Poisson}(\lambda)$. The key observation is that the vector $Q_{t_{i+1}}$ in Eq. (4) is nothing but a truncated linear convolution of the vectors $Q_{t_i}$ and $\pi_{n(t_{i+1} - t_i)}$. Hence we may apply the circular convolution theorem to compute it in the following fashion:

1. Append $n$ zeros to the end of the two vectors $Q_{t_i}$ and $\pi_{n(t_{i+1} - t_i)}$, denoting the resulting vectors $Q^{2n}$ and $\pi^{2n}$ respectively.
2. Compute the Fourier transform of the zero-extended vectors $F\{Q^{2n}\}$ and $F\{\pi^{2n}\}$, using the FFT algorithm.
3. Use the convolution theorem to obtain the Fourier transform of the convolution of the two zero-extended vectors,$$C^{2n} = F\{Q^{2n} \ast \pi^{2n}\} = F\{Q^{2n}\} \cdot F\{\pi^{2n}\},$$where $\ast$ denotes cyclic convolution and $\cdot$ denotes pointwise multiplication.
4. Compute the inverse Fourier transform of $C^{2n}$ to yield the vector $Q_{t_{i+1}}$ where $$Q_{t_{i+1}}(m) = \begin{cases} F^{-1}\{C^{2n}\}(m) & \text{if } g(t_{i+1}) \leq m \leq h(t_{i+1}) \\ 0 & \text{otherwise} \end{cases}.$$ For more details on the FFT and the computation of discrete convolutions, see Press et al. [15, Chapters 12, 13]. Using the FFT algorithm, each convolution takes $O(n \log n)$ time. Repeating this for all $t \in T(g) \cup T(h) \cup \{1\}$ yields a worst-case total running time of $O(n^3 \log n)$, but in fact it may be much lower if the functions $g(t)$ and $h(t)$ are close to each other. Specifically, note that at each step we need only to compute the cyclic convolution of two vectors of length $2w_i$ where $w_i = h(t_i) - g(t_i) + 1$, thus the total running time is $\sum_{i=0}^{N-1} O(w_i \log w_i)$ which can be as low as $O(n)$.

3. Boundary crossing probability for a Binomial process

We now return to the problem of calculating the probability given in Eq. (1), that a Binomial process will cross the boundaries prescribed by the two functions $g(t)$ and $h(t)$. Similarly to the previous section, we consider the probabilities$$R(s, m) = \Pr[\forall t \in [0, s] : g(t) \leq \eta_n(t) \leq h(t) \text{ and } \eta_n(s) = m].$$Let $0 = t_0 \leq t_1 \leq \ldots \leq t_N = 1$ be as before, and let $Z_{t,i} \sim \text{Binomial}(n - \ell, \frac{t_{i+1} - t_i}{t_{i+1} - t_i})$. The Chapman-Kolmogorov equations give the recursive relations of Friedrich and Schellhaas [9]

$$R(t_{i+1}, m) = \begin{cases} \sum_\ell R(t_i, \ell) \cdot \Pr[Z_{t,i} = m - \ell] & \text{if } g(t_{i+1}) \leq m \leq h(t_{i+1}) \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$
In contrast to Eq. (4), the expression for $R_{t_{i+1}}$, the vector of probabilities at time $t_{i+1}$, is not in the form of a straightforward convolution, and hence cannot be directly computed using the FFT. While not the focus of our work, we note that by some algebraic manipulations, it is possible to compute Eq. (5) using a convolution and an additional $O(n)$ operations. Instead, we present a simpler construction that builds upon the results of the previous section. To this end we recall a well-known reduction from the Binomial to the Poisson case (Shorack and Wellner [16, Chapter 8, Proposition 2.2]):

**Lemma 1.** The distribution of a Binomial process $\eta_n(t)$ is identical to that of a Poisson process $\xi_n(t)$ with intensity $n$, conditioned on $\xi_n(1) = n$.

According to this lemma, the non-crossing probability of a Binomial process $\eta_n$ can be efficiently computed by a simple reduction to the Poisson case

$$\Pr[\forall t : g(t) \leq \eta_n(t) \leq h(t)] = \Pr[\forall t : g(t) \leq \xi_n(t) \leq h(t) | \xi_n(1) = n] = \frac{\Pr[Poisson(n) = n]}{\Pr[Poisson(n) = n]}.$$

(6)

### 4. Computing p-values for goodness-of-fit statistics

The results of the previous sections can be used to compute the $p$-value of various supremum-based continuous goodness-of-fit statistics, and their power against specific alternatives. Such statistics include the Anderson-Darling supremum statistic [17], weighted versions of the Kolmogorov-Smirnov statistic, the Berk-Jones statistic [18], Phi-Divergence statistics [19], the Calibrated Kolmogorov-Smirnov statistic [17], and others. Our algorithm may also be applied to one-sided statistics such as the Higher-Criticism statistic of Donoho and Jin [20] or to the one-sided versions of any of the above-mentioned statistics.

First, we give a short description of the classical continuous goodness-of-fit testing problem. Let $x_1, x_2, \ldots, x_n$ be $n$ i.i.d. samples of a real-valued one dimensional random variable $X$. We wish to assess the validity of a null hypothesis that $X$ follows a known (and fully specified) continuous distribution function $F$ against an unknown and arbitrary alternative $G$,

$$\mathcal{H}_0 : X \sim F \quad vs. \quad \mathcal{H}_1 : X \sim G \quad with \quad G \neq F.$$

Let $u_i = F(x_i)$ be the probability integral transform of the $i$-th sample, and let $u_{(1)} \leq u_{(2)} \leq \ldots \leq u_{(n)}$ be the sorted sequence of transformed samples. Under the null, each $u_i$ is uniformly distributed in $[0, 1]$ and therefore $u_{(i)}$ is distributed as the $i$-th order statistic of a uniform distribution. Many goodness-of-fit statistics measure the deviation of each $u_i$ from its expectation according to some function, and then take the maximum of these deviations. Specifically, let $T_{\downarrow(1)}, \ldots, T_{\downarrow(n)}$ be a sequence of non-decreasing functions and let $T_{\uparrow(1)}, \ldots, T_{\uparrow(n)}$ be decreasing functions. One can define one-sided goodness-of-fit statistics by $T_{\downarrow} := \max_i T_{\downarrow(i)}(u_{(i)})$ and $T_{\uparrow} := \max_i T_{\uparrow(i)}(u_{(i)})$. and a two-sided statistic by $T := \max\{T_{\downarrow}, T_{\uparrow}\}$. It is easy to verify that $T \leq t$ if and only if $T_{\downarrow(i)}(t) \leq u_{(i)} \leq T_{\uparrow(i)}(t)$ holds for all $i$. Therefore the $p$-value of $T = t$ is equal to

$$\Pr[T \geq t | \mathcal{H}_0] = 1 - \Pr[\forall i : T_{\downarrow(i)}^{-1}(t) \leq U_{(i)} \leq T_{\uparrow(i)}^{-1}(t)|U_1, \ldots, U_n \overset{i, i.d.}{\sim} U[0, 1]].$$

(7)

If we take $g(t)$ and $h(t)$ to be the empirical distribution functions of the sets $\{T_{\downarrow(i)}^{-1}(t), \ldots, T_{\downarrow(n)}^{-1}(t)\}$ and $\{T_{\uparrow(i)}^{-1}(t), \ldots, T_{\uparrow(n)}^{-1}(t)\}$ respectively, then it easily follows that the probability of Eq. (7) is equal to that of Eq. (6) which we can compute in time $O(n^2 \log n)$.

Next, we compare the empirical run-time of our procedure in the context of $p$-value calculations as in Eq. (7) for the one-sided and two-sided CKS goodness-of-fit statistics [12]. To this end we wrote an efficient implementation of the proposed procedure using the FFTW3 library by Frigo and Johnson [21] and compared it to a direct implementation of the Khmaladze and Shinjikashvili [2] recursion relations.
Table 1: Large sample running times for computing \( p \)-values of the two-sided and one-sided CKS goodness-of-fit statistics. In all cases, the value of the CKS statistic was chosen such that its \( p \)-value (equal to the boundary crossing probability) is roughly 5%.

In addition to that, we implemented the \( O(n^2) \) one-sided algorithm of [12]. Both two-sided procedures are numerically stable using standard double-precision (64-bit) floating point numbers, even for very large sample sizes. In contrast, the one-sided procedure necessitates the use of extended-precision (80-bit) floating point numbers in addition to a careful numerical implementation. It can only be used with sample sizes up to about \( n = 50,000 \) before one must resort to slower high-precision floating point numbers implemented in software. C++ source code for all procedures is freely available at http://www.wisdom.weizmann.ac.il/~amitmo.

Table 1 lists running times on a 2010 Intel® Xeon® X5660 CPU using a single thread. In all cases, our procedure provides a significant improvement relative to the other two-sided procedure. However, it really shines when the distance between the lower and upper boundary functions is large, as exemplified by the one-sided case. Interestingly, while the two-sided \( O(n^2 \log n) \) procedure is asymptotically slower than the one-sided \( O(n^2) \) procedure, this does not make a big difference in practice, even for the large sample sizes we have tested.

References

[1] M. Noé, The calculation of distributions of two-sided Kolmogorov-Smirnov type statistics, The annals of mathematical statistics 43 (1) (1972) 58–64, URL http://doi.org/10.1214/aoms/1177692700.

[2] E. Khmaladze, E. Shinjikashvili, Calculation of noncrossing probabilities for poisson processes and its corollaries, Advances in applied probability 33 (2001) 702–716, URL http://doi.org/10.1239/aap/1005091361.

[3] G. P. Steck, Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions, The annals of mathematical statistics 42 (1) (1971) 1–11, URL http://doi.org/10.1214/aoms/1177693490.

[4] J. Durbin, Distribution theory for tests based on sample distribution function, SIAM, URL http://doi.org/10.1137/1.9781611970586, 1973.

[5] A. B. Owen, Nonparametric likelihood confidence bands for a distribution function, Journal of the american statistical association. 90 (430) (1995) 516–521, URL http://doi.org/10.2307/2291062.

[6] J. Frey, Optimal distribution-free confidence bands for a distribution function, Journal of statistical planning and inference 138 (2008) 3086–3098, URL http://doi.org/10.1016/j.jspi.2007.12.001.

[7] K. J. Worsley, Confidence regions and tests for a change point in a sequence of exponential family random variables, Biometrika 73 (1) (1986) 91–104, URL http://doi.org/10.1093/biomet/73.1.91.
[8] S. Dongchu, Exact computation for some sequential tests, Sequential analysis 17 (2) (1998) 127–150, URL http://doi.org/10.1080/07474949808836403.

[9] T. Friedrich, H. Schellhaas, Computation of the percentage points and the power for the two-sided Kolmogorov-Smirnov one sample test, Statistical papers 39 (1998) 361–375, URL http://doi.org/10.1007/BF02927099.

[10] M. Noé, G. Vandewiele, The calculation of distributions of Kolmogorov-Smirnov type statistics including a table of significance points for a particular case, The annals of mathematical statistics 39 (1) (1968) 233–241, URL http://doi.org/10.1214/aoms/1177698523.

[11] V. F. Kotel’Nikova, E. V. Khmaladze, On computing the probability of an empirical process not crossing a curvilinear boundary, Theory of probability & its applications 27 (1983) 640–648, URL http://doi.org/10.1137/1127075.

[12] A. Moscovich-Eiger, B. Nadler, C. Spiegelman, The calibrated Kolmogorov-Smirnov test, Tech. Rep., URL http://arxiv.org/abs/1311.3190, 2014.

[13] D. Siegmund, Boundary crossing probabilities and statistical applications, The annals of statistics 14 (2) (1986) 361–404, URL http://doi.org/10.1214/aos/1176349928.

[14] R. Chicheportiche, J.-P. Bouchaud, Some applications of first-passage ideas to finance, in: First-passage phenomena and their applications, URL http://doi.org/10.1142/9789814590297_0018, 2014.

[15] W. H. Press, B. P. Flannery, S. A. Teukolsky, V. W. T., Numerical recipes in C: the art of scientific computing, Cambridge University Press, 2nd edn., URL http://apps.nrbook.com/c/index.html, 1992.

[16] G. R. Shorack, J. A. Wellner, Empirical processes with applications to statistics, SIAM, URL http://doi.org/10.1137/1.9780898719017, 2009.

[17] T. Anderson, D. Darling, Asymptotic theory of certain ”goodness of fit” criteria based on stochastic processes, The annals of mathematical statistics 23 (2) (1952) 193–212, URL http://projecteuclid.org/euclid.aoms/1177729437.

[18] R. H. Berk, D. H. Jones, Goodness-of-fit test statistics that dominate the Kolmogorov statistics, Probability theory and related fields 47 (1979) 47–59, URL http://doi.org/10.1007/BF00533250.

[19] L. Jager, J. A. Wellner, Goodness-of-fits via phi-divergences, Annals of statistics 35 (5) (2007) 2018–2053, URL http://doi.org/10.1214/009053607000000244.

[20] D. Donoho, J. Jin, Higher criticism for detecting sparse heterogeneous mixtures, Annals of statistics 32 (3) (2004) 962–994, URL http://doi.org/10.1214/009053604000000265.

[21] M. Frigo, S. G. Johnson, The design and implementation of FFTW3, Proceedings of the IEEE 93 (2) (2005) 216–231, URL http://doi.org/10.1109/JPROC.2004.840301.