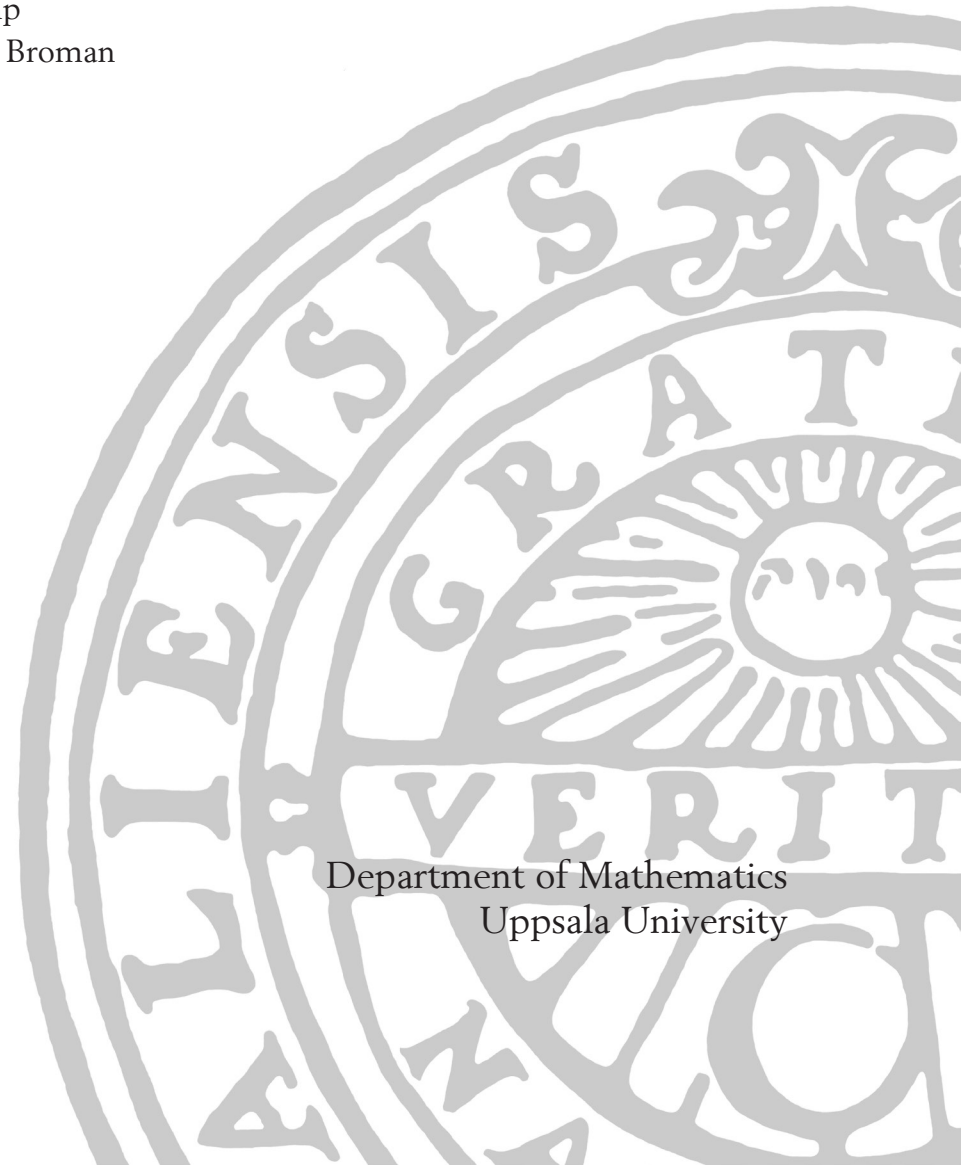
UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2015:36

# Random covering times, a simulation study

Emil Rosén

Examensarbete i matematik, 15 hp
Handledare och examinator: Erik Broman
December 2015



Department of Mathematics
Uppsala University

# Examensarbete C

Emil Rosén

**Abstract**

In this paper the cover times for a set $K$ is studied by computer simulations. It is investigated how the cover time increases when the area of $K$ approaches infinity, both when balls and cylinders are thrown at it. It is also compared and discussed how the cover times behaves for some different shapes of $K$. Further it is investigated how the cover time changes when there are more than one set $K$ to cover, both in the case of ball and cylinder throwing.

## 1 Introduction

The cover time of a large set will be studied, i.e the statistics of how long time it will take to cover a set when throwing other smaller sets at it. Much of the material on the subject have been developed by Svante Jansson and can be found in [2]. Some different aspects of this problem is here addressed by computer simulations. Let $\tau_K$ be the cover time of $K$ (see definition in sec 1.1 and 1.2) when throwing balls at $K$, we will study how $\tau_K$ increases as $K$ grows. We also investigate how $\tau_K$ behaves for some different shapes of $K$. Further simulations are carried out when there are two large sets, $K_1$ and $K_2$, to cover. We then study how their cover times are correlated, and how this correlation decreases when the distance between the sets increases. The reason for selecting a simulation approach is that the problem is intractable to solve analytically, although it is possible to establish asymptotic results as in [2]. The attention will be directed at the two dimensional case, $d = 2$. We make this restriction in order to make it more computational feasible. The different investigations performed are explained more in detail in the following sub sections.

## 1.1 Throwing balls at K

Let $\Phi_d$ denote a homogeneous Poisson process in $\mathbb{R}^d \times \mathbb{R}^+$ and let $(x,t)$ denote a point in $\Phi_d$. We think of $(x,t)$ as a spatial coordinate $x$ together with a time coordinate $t$. Let $B(x,r)$ denote a ball in $\mathbb{R}^d$ with center at $x$ and radius $r$, i.e $B(x,r) = \{y \in \mathbb{R}^d : |x-y| \leq r\}$. Define

$$\tau_K = \tau_K(\Phi_d) = \inf\{T > 0 : K \subseteq \bigcup_{\substack{(x,t)\in\Phi_d \\ t\leq T}} B(x,1)\},$$

i.e. $\tau_K$ is the smallest time such that $K$ is completely covered by balls whose centers are the spatial component of the points in the Poisson process $\Phi_d$. We refer to $\tau_K$ as the cover time of $K$. Thus we can see this as a procedure in which we are throwing balls on $K$ at unite rate until $K$ becomes completely covered.

The first question considered in this paper is how rapidly the cover time of $K$, $\tau_K$, approaches infinity as $|K| \to \infty$ and how this depends on the shape of $K$.

## 1.2 Throwing cylinders at K

The second problem studied is similar to the first one considered and can be described as follows. Let $\mathcal{L}_d$ be the space of all lines in $\mathbb{R}^d$. Let $\Xi_d$ be a Poisson process on $\mathcal{L}_d \times \mathbb{R}^+$ and let $(L,t) \in \mathcal{L}_d \times \mathbb{R}^+$ denote a point in this space. We will again think of $t$ as time. For every $L \in \mathcal{L}_d$ define $\mathfrak{C}(L)$ to be a cylinder with radius one and whose axis is $L$, that is

$$\mathfrak{C}(L) = \{x \in \mathbb{R}^d : \exists y \in L, |x-y| \leq 1\}.$$

As before we define a cover time

$$\tau_K(\Xi_d) = \inf\{T > 0 : K \subseteq \bigcup_{\substack{(L,t)\in\Xi_d \\ t\leq T}} \mathfrak{C}(L)\}.$$

Hence $\tau_K(\Xi_d)$ is the first time when $K$ is completely covered by cylinders. The question studied is again at what rate the cover time, $\tau_K(\Xi_d)$ will increase as $|K|$ increases.

### 1.3 Throwing cylinders at two sets

The third problem considered is to study how the cover times of two disjoint sets $K_1$ and $K_2$ are correlated, and how this correlation decreases when the distance between the two sets increases. This problem is only studied for the cylinder throwing case. In the ball throwing case this problem is trivial since if the sets are separated by more than the diameter of the smaller sets their cover times are independent. This is not the case when throwing cylinders. Regardless of how far apart the sets $K_1$ and $K_2$ are, it can still be the case that one cylinder intersects both $K_1$ and $K_2$, therefore the cover times will not be independent. The joint distribution of the cover times of $K_1$ and $K_2$ is hence different from the cover time of $K_1$ only.

## 2 Theory

### 2.1 Preliminaries

First we introduce some notations and a few theorems which will be used in this paper. Given a set $A$ we will denote its area (in two dimensions) or volume (higher dimensions) by $|A|$, i.e $|A|$ denotes the Lebesgue measure of $A$.

We will let $o(h)$ be the set of all functions $f$ which satisfies

$$\lim_{h\to 0} f(h)/h = 0.$$

We will with a slight abuse of notation say that $f(h) = o(h)$ when $f(h) \in o(h)$.

#### 2.1.1 Some distributions

**Poisson distribution**

A random variable $X$ taking values in the natural numbers will be said to be Poisson distributed with intensity $\lambda$ if

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad k = 0, 1, 2, 3, \dots$$

and we will denote this by $X \sim Po(\lambda)$.

**Binomial distribution**

The random variable $X$ taking values in $\{0, 1, \ldots, N\}$ will be said to have a Binomial distribution if

$$\mathbb{P}(X = n) = \binom{N}{n} p^n (1-p)^{N-n}, \qquad n = 0, 1, \ldots, N,$$

which we will denote by $X \sim Bin(N, p)$.

### 2.1.2 Useful lemmas and propositions

Next follows an inequality which says that $e^x$ lies above its tangent which passes through the point $(0, 1)$.

**Lemma 2.1.**

$$1 + x \leq e^x, \ \forall x \in \mathbb{R}. \tag{1}$$

*Proof.* When $x \geq 0$ we have that

$$1 + x \leq \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x.$$

When $-1 \leq x < 0$ then $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ is an alternating series with monotonically decreasing terms. In this case is well known that the partial sums $S_N(x) = \sum_{n=0}^{N} \frac{x^n}{n!}$ is a lower bound of the series value whenever $N$ is a odd natural number and it is an upper bound whenever $N$ is an even natural number. Hence when $-1 \leq x < 0$ we have that $1 + x = S_1(x) \leq e^x$.

When $x < -1$ it the inequality follows by the fact that $1 + x < 0 < e^x$, hence the inequality follows. □

**Lemma 2.2.**

$$1 - e^{-x} \geq \frac{x}{2}, \quad 0 \leq x < \ln(2). \tag{2}$$

*Proof.* Let

$$f(x) = 1 - e^{-x} - \frac{x}{2}.$$

Then

$$f'(x) = e^{-x} - \frac{1}{2}.$$

Thus when $x < \ln(2)$ we have that $f'(x) > 0$. Since $f(0) = 0$ it follows that when $0 \leq x < \ln(2)$ we have the inequality

$$0 \leq 1 - e^{-x} - \frac{x}{2}.$$

which is what we want to prove. □

The following lemma can be useful if we have an infinite number of events and we wish to decide if an infinite number of these events will occur or only a finite number of them.

**Lemma 2.3** (Borel-Cantelli)**.** *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events in a probability space. Let $A = \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m$ be the event that infinitely many of the events $A_n$ occur. Then:*

1. $\mathbb{P}(A) = 0$ *if* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$.

2. $\mathbb{P}(A) = 1$ *if* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ *and* $A_1, A_2, \ldots$ *are mutually independent events.*

*Proof.* To prove the first part (1) of the lemma, let

$$B_n = \cup_{m=n}^{\infty} A_m.$$

Then $\{B_n\}$ is a decreasing sequence of sets, i.e $B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots$. Therefore by the continuity of the probability measure it follows that $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(B_n)$. Now we have that

$$\mathbb{P}(B_n) \leq \sum_{m=n}^{\infty} \mathbb{P}(A_m).$$

As $n \to \infty$ the sum of the series approaches 0 since it is an convergent series, hence $\mathbb{P}(A) = 0$, which proves the first part of the lemma.

To prove part (b). Note that

$$A^c = \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m^c.$$

Let

$$C_n = \cap_{m=n}^{\infty} A_n^c.$$

Then $\{C_n\}$ is an increasing sequence, i.e $C_1 \subseteq C_2 \subseteq C_3 \subseteq \cdots$. From the continuity of the probability measure it follows that $\mathbb{P}(A^c) = \lim_{n\to\infty} \mathbb{P}(C_n)$. Since the events $A_n^c$ are mutually independent we get that

$$\mathbb{P}(C_n) = \prod_{m=n}^{\infty} \mathbb{P}(A_m^c) = \prod_{m=n}^{\infty} (1 - \mathbb{P}(A_m)) \leq \prod_{m=n}^{\infty} \exp\left(-\mathbb{P}(A_m)\right) = \exp\left(-\sum_{m=n}^{\infty} \mathbb{P}(A_m)\right).$$

Where the inequality is Lemma 2.1. It now follows that

$$\mathbb{P}(A^c) = \lim_{n\to\infty} \mathbb{P}(C_n) \leq \lim_{n\to\infty} \exp\left(-\sum_{m=n}^{\infty} \mathbb{P}(A_m)\right) = 0.$$

Thus we conclude that $\mathbb{P}(A) = 1$ which we wanted to prove. □

The next proposition says that if we have a Poisson distributed random number of balls and we color each of them independent of each other, either red with probability of $p$ or blue with probability $1 - p$. Then the number of red balls will again be Poisson distributed.

**Proposition 2.1.** *Given two random variables $N \sim Po(\lambda)$ and $M \sim Bin(N, p)$ then the variable $M$ is Poisson distributed with parameter $p\lambda$.*

*Proof.* Let $N \sim Po(\lambda)$ and $M \sim Bin(N, p)$ and we define the characteristic function of $M$ as

$$\Psi(t) = \mathbb{E}[e^{itM}].$$

By conditioning on $N$, we get that

$$\mathbb{E}[e^{itM}|N] = \sum_{n=0}^{N} \binom{N}{n} e^{int} p^n (1-p)^{N-n} = (1 + (e^{it} - 1)p)^N.$$

Let $f(t) = 1 + (e^{it} - 1)p$ and we then get that

$$\Psi(t) = \mathbb{E}[\mathbb{E}[e^{itM}|N]] = \mathbb{E}[f(t)^N] = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n f(t)^n}{n!} = e^{-\lambda} e^{\lambda f(t)} = \exp((e^{it} - 1)p\lambda).$$

This is the characteristic function of the Poisson distribution with intensity $p\lambda$ and it therefore follows that $M$ have the distribution $Po(p\lambda)$. □

## 2.2 Poisson Process in $\mathbb{R}^d$

In this subsection some short material about the Poisson process is covered which will be relevant. For a more thorough exposition on the subject see e.g [3]. There are a lot of things that can be

considered to be placed in space in a random fashion, e.g the location of sheeps in a sheep pen at a given time point, the positions of particles in a gas in a gas chamber at some time point or the stars in the night sky. The Poisson process is one way of describing such random positions, however not all of these examples are suitably described with the Poisson process. The points in the Poisson process show no sign of any patterns and no part of space will look more special than any other, as it would if we were looking at the sheeps where we would find a clump of points at one location and non elsewhere. The points of the Poisson process will therefore be distributed over the whole space in a haphazardly way and how one part of space look is independent of any other part of space. Throughout this paper we will talk about the number of points belonging to the Poisson process in some subset $A \subset \mathbb{R}^d$ and we will denote it by $N(A)$. Formally we will denote the Poisson process as follows.

**Definition 2.1.** *A random countable subset $\Xi$ of $\mathbb{R}^d$ is called a Poisson process with intensity $\lambda$ if,*

1. *For all $A \subseteq \mathbb{R}^d$, the random variable $N(A)$ has a Poisson distribution with parameter $\lambda|A|$.*
2. *If $A_1, A_2, \ldots, A_n$ are disjoint subsets of $\mathbb{R}^d$, then $N(A_1), N(A_2), \ldots, N(A_n)$ are independent random variables.*

The independence of different parts of space in the definition of the Poisson process excludes situations as the sheeps in the sheep pen, since if we know that one part of space contains some sheeps we are not likely to find any sheeps far away from that part of space. The Poisson distribution in the definition is inevitable under the independence assumption and the additional assumptions:

1. The points in the process are not too close to each other, thus if we look at a small part of space we are highly unlikely to find more than one point in this part. This assumption can be formulated as
$$\mathbb{P}(N(A) > 1) = o(|A|).$$
2. In a small region of space the probability of finding a point in it is proportional to the area of the region which we formulate as
$$\mathbb{P}(N(A) \geq 1) = \lambda|A| + o(|A|)$$

A derivation of this, following the one given in [3], is here stated. Now consider the ball $B_r = B(0,r)$ and let $q_n(r) = \mathbb{P}(N(B_r) \leq n)$, $p_n(r) = \mathbb{P}(N(B_r) = n)$ and we also define $m(r) = \lambda|B_r|$. We can show that the number of points in $B_r$ is Poisson distributed by deriving a differential equation for $p_n(r)$.

Using assumption 2 we get that the probability of a ring $B_{r+h} \setminus B_r$ containing at least one point is

$$\begin{aligned}\mathbb{P}(N(B_{r+h} \setminus B_r) \geq 1) &= \lambda|B_{r+h} \setminus B_r| + o(|B_{r+h} \setminus B_r|) \\ &= \lambda|B_{r+h}| - \lambda|B_r| + o(h) \\ &= m(r+h) - m(r) + o(h).\end{aligned}$$

The event that $B_r$ contains at most $n$ points and the event that $B_{r+h}$ contains more than $n$ points, using assumption 1, can be calculated as follows

$$\begin{aligned}q_n(r) - q_n(r+h) &= \mathbb{P}(N(B_r) \leq n, N(B_{r+h}) > n) \\ &= \mathbb{P}(N(B_r) = n,\, N(B_{r+h} \setminus B_r) \geq 1) + \sum_{k=0}^{n-1} \mathbb{P}(N(B_r) = k,\, N(B_{r+h} \setminus B_r) > n-k) \\ &= \mathbb{P}(N(B_r) = n)\mathbb{P}(N(B_{r+h} \setminus B_r) \geq 1) + \sum_{k=0}^{n-1} \mathbb{P}(N(B_r) = k)\mathbb{P}(N(B_{r+h} \setminus B_r) > n-k) \\ &= p_n(r)(m(r+h) - m(r) + o(h)) + \sum_{k=0}^{n-1} p_k(r)o(h).\end{aligned}$$

By dividing this relationship with $h$ and letting $h \to 0$ we get

$$-\frac{dq_n}{dr} = p_n \frac{dm}{dr}.$$

Since $p_0 = q_0$ it follows that

$$\frac{dp_0}{dr} = -p_0 \frac{dm}{dr}$$

and the solution to this equation with the initial conditions $p_0(0) = 1, m(0) = 0$ is

$$p_0(r) = e^{-m(r)}.$$

When $n \geq 1$ we notice that $p_n(r) = q_n(r) - q_{n-1}(r)$, hence it follows

$$\frac{dp_n}{dr} = \frac{dq_n}{dr} - \frac{dq_{n-1}}{dr} = (p_{n-1} - p_n)\frac{dm}{dr}. \tag{3}$$

Equation (3) together with the initial condition $p_n(0) = 0$ when $n \geq 1$ gives that

$$p_n(r) = e^{-m(r)} \int_0^r p_{n-1}(t) e^{m(t)} \frac{dm}{dt} dt.$$

Hence by induction we get that

$$p_n(r) = \frac{e^{-m(r)} m^n(r)}{n!}, \quad (n \geq 0).$$

Thus the distribution of $N(B_r)$ is $Po(m(r))$. Therefore with these assumptions the Poisson distribution is inevitable and we can not choose any other distribution.

The next proposition will be important when we generate realizations of the Poisson process, and this is the 'conditional property' of the Poisson process.

**Proposition 2.2.** *Let $\Xi$ be a Poisson process on $\mathbb{R}^d$ and let $A \subseteq \mathbb{R}^d$ be a part of space such that $0 < |A| < \infty$. Given that $N(A) = n$, the $n$ points of $\Xi$ lying in $A$ have the same distribution as $n$ points which are chosen randomly and independently in $A$ according to a uniform distribution in $A$.*

For a proof of this proposition, see [1].

### 2.2.1 Poisson Processes in general spaces

We can also define a Poisson process on more general spaces than $\mathbb{R}^d$. If we have some measurable space $(\Omega, \mathscr{F}, \mu)$, where $\mathscr{F}$ is a $\sigma$-algebra and $\mu$ is a measure, i.e a function from $\mathscr{F}$ to $\mathbb{R}$ which satisfies

1. $\mu(\emptyset) = 0$,
2. $\mu(A) \geq 0$, for every $A \subseteq \mathscr{F}$,
3. If $\{A_i\} \subseteq \mathscr{F}$ is a countable set of disjoint subsets, then $\sum_{i=0}^{\infty} \mu(A_i) = \mu(\cup_{i=0}^{\infty} A_i)$,

we can define the Poisson process as follows

**Definition 2.2.** *The random countable subset $\Xi$ of $\Omega$ is called a Poisson process with mean measure $\mu$ if,*

1. *For any measurable subset $A$ the random variable $N(A)$ has a Poisson distribution with parameter $\mu(A)$.*

2. *If $A_1, A_2, \ldots, A_n$ are disjoint sets in $\mathbb{R}^d$, then $N(A_1), N(A_2), \ldots, N(A_n)$ are independent random variables.*

This definition should be compared with the definition of the Poisson process on $\mathbb{R}^d$. Almost nothing has changed between these definitions, the difference is that in a more general space we don't have the intuitive concept of an area, instead we are replacing this with the concept of a measure.

The cylinder problem can be seen as that we are selecting the time coordinates and the cylinders $x$-coordinate, i.e where the cylinders axis intersects the $x$-axis, according to a Poisson process on $\mathbb{R}^+ \times \mathbb{R}$. Then the angle between the $x$-axis and the cylinder is selected randomly with density $sin(\theta)/2,\ 0 \leq \theta \leq \pi$ and independent of each other. This can be shown to be the only sensible choice of the density of the angle. Any other choice would not create a homogeneous process.

Thus the problem can be seen as a Poisson process on $\mathcal{L}_2 \times \mathbb{R}^+$. We hence want to have the measure of those lines which intersects the larger set $K$ so let

$$\mathscr{L}_K = \{L \in \mathcal{L}_2 : L \cap K \neq \emptyset\}.$$

Let $\Pi_\theta(K)$ be the projection of the set $K$ onto the line which passes through the origin and intersects the $x$-axis with angle $\theta$. Then it can be show that

$$\mu(\mathscr{L}_K) = \int_0^{2\pi} |\Pi_\theta(K)| d\theta, \tag{4}$$

see [4] for a proof of this.

Our next proposition gives an exact condition of when there only will be a finite number of cylinders intersecting a bounded interval of the $y$-axis during a finite time interval. In particular, it shows that the choice $\sin(\theta)/2$ will not create a process where there are an infinite number of cylinders intersecting a bounded interval of the $y$-axis.

**Proposition 2.3.** *Assume that the angle $\theta$ which the cylinders are chosen with are selected from a continuous distribution with density $f(\theta)$, $0 \leq \theta \leq \pi$. Then, under a finite time interval, there are only a finite number of cylinder axis which intersects $I = \{(0, y) : 0 \leq y \leq 1\}$ if and only if*

$$\int_0^\pi \frac{f(\theta)}{sin(\theta)} d\theta < \infty.$$

*Proof.* We measure the angle clockwise from the $x$-axis and we only look at the cylinders which has a time between 0 and 1 and a positive $x$-coordinate. Those with a negative $x$-coordinate can be counted in a similar manner and reach an analogous result. Assume also that the intensity of the Poisson process is $\lambda$. Let $L_{[n,n+1)}$ be the number of cylinders with $x$-coordinate in $[n, n+1)$ that intersects the $y$-axis through $I$. Let $M_{[n,n+1)}$ and $J_{[n,n+1)}$ be the number of cylinders with $x$-coordinate within $[n, n+1)$ that has an angle which is lower than $\arctan(1/n)$ and $\arctan(1/(n+1))$ respectively. Thus we always have that $M_{[n,n+1)} \geq L_{[n,n+1)} \geq J_{[n,n+1)}$ since every cylinder which $L_{[n,n+1)}$ counts have an angle which is less than $\arctan(1/n)$ and every cylinder within the interval which has an angle less than $\arctan(1/(n+1))$ intersects $I$. Define

$$p(x) = \begin{cases} \mathbb{P}(\theta \leq \arctan(1/x)) & \text{if } x > 0, \\ 1/2 & \text{if } x = 0. \end{cases}$$

This is an upper bound of the probability that a cylinder with $x$-coordinate in the interval $[n, n+1)$ intersects the $y$-axis through $I$.

The number of cylinders $N$ with $x$-coordinate within $[n, n+1)$ is Poisson distributed with parameter $\lambda$ and the angles are independent of each other, hence $M_{[n,n+1)}$ is distributed as $Bin(N, p(n))$. By Proposition 2.1 it therefore follows that $M_{[n,n+1)} \sim Po(p(n)\lambda)$. With the same argument we reach the conclusion that $J_{[n,n+1)} \sim Po(p(n+1)\lambda)$. We are now looking at the events $\{L_{[n,n+1)} > 0\}_{n=0}^{\infty}$ and we are interested to know whether or not an infinite number of these event occur. Therefore we bound the following sum

$$\begin{aligned} \sum_{n=0}^{\infty} \mathbb{P}(L_{[n,n+1)} > 0) &\leq \sum_{n=0}^{\infty} \mathbb{P}(M_{[n,n+1)} > 0) = \sum_{n=0}^{\infty} \left(1 - e^{-\lambda p(n)}\right) \\ &\leq \lambda \sum_{n=0}^{\infty} p(n) \leq \lambda p(0) + \lambda \int_0^{\infty} p(x) dx \end{aligned}$$

where the second inequality is Lemma 2.1 and the third inequality follows form the fact that $p(x)$ is decreasing in $x$. We can also obtain a lower bound of the sum, which is

$$\sum_{n=0}^{\infty} \mathbb{P}(L_{[n,n+1)} > 0) \geq \sum_{n=0}^{\infty} \mathbb{P}(J_{[n,n+1)} > 0) = \sum_{n=0}^{\infty} \left(1 - e^{-\lambda p_{n+1}}\right)$$
$$\geq \frac{\lambda}{2} \sum_{n=0}^{\infty} p(n+1) \geq \frac{\lambda}{2} \int_{1}^{\infty} p(x)dx.$$

where the second inequality is Lemma 2.2 and the third inequality follows since $p(x)$ is decreasing.

Thus the sum $\sum_{n=0}^{\infty} \mathbb{P}(L_{[n,n+1)})$ converges if and only if the integral in the bound converges. This integral can be calculated to be

$$\int_0^{\infty} p(x)dx = \int_0^{\infty} \mathbb{P}(\theta \leq \arctan(1/x))dx = \int_0^{\infty} \int_0^{\arctan(1/x)} f(\theta)d\theta dx$$
$$= \int_0^{\pi/2} \int_0^{1/\tan(\theta)} f(\theta)dxd\theta = \int_0^{\pi/2} f(\theta)\frac{\cos\theta}{\sin\theta}d\theta. \tag{5}$$

If this integral is finite, then Borel-Cantelli tells us that only a finite number of the events $\{L_{[n,n+1)} > 0\}_{n=0}^{\infty}$ occur and if it is infinite then an infinite number of these events occur.

Now by symmetry we can do the same thing with the negative part of the $x$-axis and reach the similar conclusion that a finite number of the cylinders with negative $x$-coordinate will intersect $I$ if and only if

$$\int_{\pi/2}^{\pi} f(\theta)\frac{\cos\theta}{\sin\theta}d\theta < \infty. \tag{6}$$

Since the integrals in (5) and (6) both converges if and only if

$$\int_0^{\pi} \frac{f(\theta)}{\sin(\theta)}d\theta$$

is finite, the proposition follows. □

As a remark on the previous proposition, the integral

$$\int_0^{1/2} \frac{x^{1/\sqrt{-\ln(x)}}}{\sin x}dx,$$

is finite which can be shown by making the change of variable $t = \sqrt{-\ln(x)}$ which gives the integral

$$\int_{\sqrt{\ln 2}}^{\infty} 2te^{-t}\frac{e^{-t^2}}{\sin e^{-t^2}}dt \leq \frac{1}{\sin(1/2)} \int_{\sqrt{\ln 2}}^{\infty} te^{-t}dt = \frac{e^{-\sqrt{\ln 2}}\left(1 + \sqrt{\ln 2}\right)}{\sin(1/2)}.$$

Hence, that $f(\theta) < \theta^{\alpha}$ for some $\alpha > 0$ in some neighborhood of $\theta = 0$, i.e a Hölder condition, is not a necessary condition for the proposition, although a sufficient.

# 3 Simulation

## 3.1 Throwing balls

To simulate the first problem, in which we are throwing balls on a set $K$, we need to be able to generate a realization of the Poisson process and decide when the set $K$ is covered. This is done by placing $K$ on a rectangle $R$ such that there becomes some margins in $R$ around $K$ which are larger than the radius of the balls. The realization is created by iteratively taking a time step $\Delta t$. When taking a step from $t$ to $t+\Delta t$ we first decide the number of figures that should be thrown on $K$. We do this by drawing a Poisson distributed random number $N$ with parameter $\Delta t|R|$. Thereafter we decide where in $[t, t+\Delta t] \times R$ these figures should be placed, which is done by randomly drawing the positions from a uniform distribution on $[t, t+\Delta t] \times R$. When we have the positions the figures should be placed at, we are sorting these positions in ascending time order and places the figures one by one on $R$. After we have placed a figure on $R$ we check if $K$ is covered or not (see Sec. 3.4), if it is covered then the time position this figure is at is the cover time of $K$. This continues until $K$ is covered and we have a realization of the cover time of $K$.

## 3.2 Throwing cylinders

The cylinder problem is simulated in the same manner as the first problem. We iteratively take a time step $\Delta t$ until the larger set $K$ is entirely covered. When we generate the number of cylinders which should be created each time step, we use equation (4) for the intensity in the Poisson distribution. The integral is calculated numerically. Following Sec. 2.2.1 the angle of the cylinders which are created is randomly selected with the density

$$f(\theta) = \frac{|\Pi_\theta(R)|}{\mu(\mathscr{L}_R)}, \qquad 0 \leq \theta \leq 2\pi,$$

where $R$ is the enclosing rectangle as when we are throwing balls on the surface, except that $R$ in this case can be a union of two rectangles when we have two surfaces we are throwing the cylinders at, also $\mathscr{L}_R$ is the set of all lines which intersects $R$. Then the position of the cylinder is chosen by randomly taking a point on $\Pi_\theta(R)$ (where $\theta$ is the angle of the cylinders which position we are

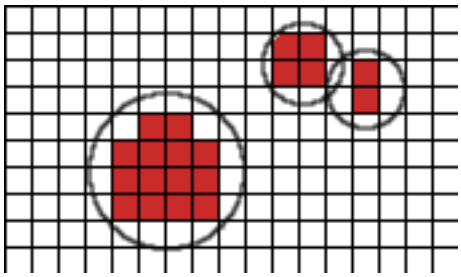


Figure 1: A figure that shows how the surface is decided if it is covered or not, the red squares are decided that they are completely covered.

selecting) according to a uniform distribution. When this point has been chosen the cylinder is then created to go through this point with the randomly chosen angle.

### 3.3 Two sets with increasing distance

Simulations were performed to study how the correlation of the cover time for two sets $K_1$ and $K_2$ varies when the distance between the sets is increasing. In these simulations the sets $K_1$ and $K_2$ were circles with radius $r = 50$. Simulations were only performed when cylinders were thrown at the sets. See [4] for why this creates a correct realization of the process.

### 3.4 Determining if K is covered

To check if the set $K$ is covered or not we split the rectangle $R$ into a grid of small squares, see Fig. 1, which we then place $K^c$ on. Hence the part of this grid which is not covered is $K$. The set $K$ is then determined to be covered if every square in the grid is covered. A square in the grid is determined to be covered if it is completely within one of the smaller figures. As in Fig. 1 the red squares are the squares that are considered to be covered. If we look at the two smaller circles that intersects in the figure we see that there is a square which is covered by the two circles together but not by any of them alone, this square is hence not considered to be covered even though it is. Because of this, we are not necessarily getting the true cover time, since the set $K$ can be covered by all the balls but still some of the grid squares are not covered by a single ball alone. Therefore the cover time we are simulating is only an upper bound of the true cover time.

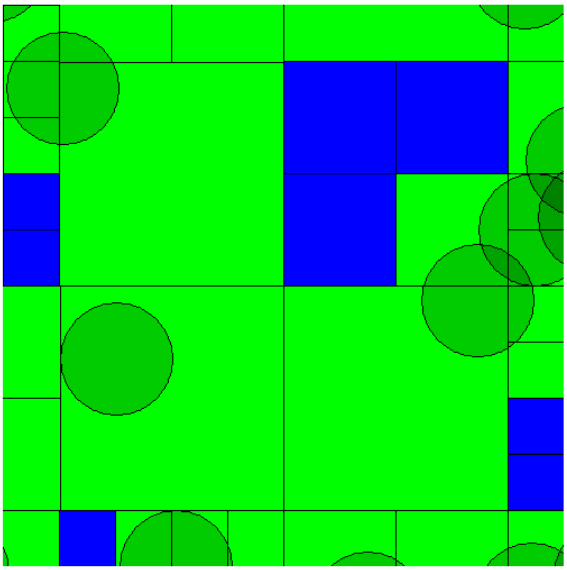


Figure 2: Illustration of how it is determined if $K$ is covered or not. The quad tree have determined that the blue squares are not covered and hence $K$ is not covered.

In order to make this upper bound as tight as possible the grid needs to consist of as small squares as possible. If we have too many squares it would take too much time to check if every square is covered or not. Therefore we are structuring the squares in a quad tree. The root node can be considered to consist of every square in the grid. These grids are then partitioned in four parts a top left, top right, bottom left and bottom right part which are then assigned to be the four child nodes to the root. The splitting of the grid continues in this way. This gives the benefit that we doe not have to check each square separate if they are covered or not, instead we can check this in chunks by only checking entire nodes in the quad tree if they are covered or not.

The tree is not built entirely since this would require too much space, instead we are only splitting the nodes until we can decide whether or not the set $K$ is covered or not. See Fig. 2, Fig. 3, Fig. 4 for an illustration of how the surface is determined if it is covered or not, the rectangles in the figures are the nodes in the tree. The color green on a rectangle indicates that this part is yet to be determined if it is covered or not and a blue color indicates that it is not covered at all. The white rectangles are parts that has been determined that it is covered.

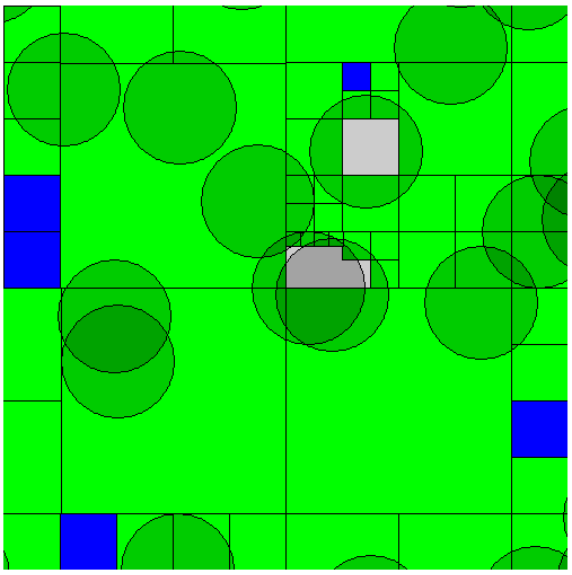


Figure 3: Illustration of how it is determined if $K$ is covered or not. The green squares are partially covered, the blue ones are not covered and the white ones are completely covered.

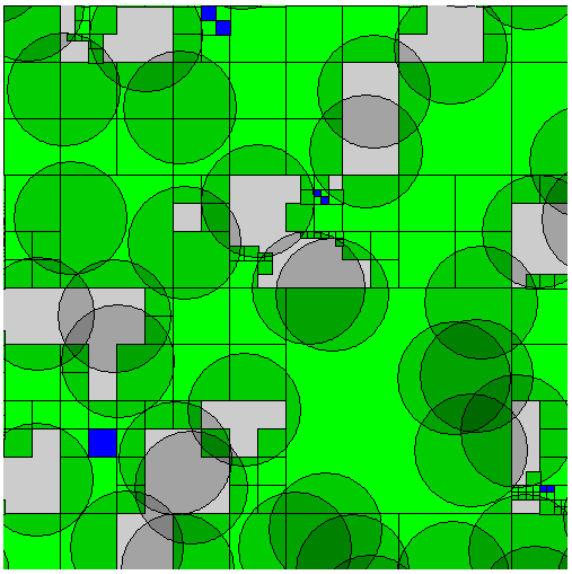


Figure 4: Another illustration of when the quad tree is working on determining if $K$ is covered or not.

## 3.5 Performed simulations and measures

### 3.5.1 Rate of increase for cover times

To study the rate, when throwing balls and cylinders, at which $\tau_K$ increases as $|K|$ increases, $\tau_K$ was measured by simulation experiments. This was done by performing 1000 simulation runs for each value of $|K|$. The mean value for the obtained $\tau_K$, as well as the confidence intervals was computed. This was done for the four different shapes given in the next section.

### 3.5.2 Different shapes of $K$

To make some kind of investigation of how the shape of the set $K$ affects the cover time $\tau_K$ simulations have been performed with the four different shapes

| Shape denotation | Shape description. |
|---|---|
| $K^{(1)}$ | Square with height $h$ equal the width $w$, i.e. $h = w$. |
| $K^{(2)}$ | Rectangle with fixed height $h = 100$ but variable width $w$. |
| $K^{(3)}$ | Rectangle with height $h$ twice the width $w$, i.e. $h = 2w$. |
| $K^{(4)}$ | Circle. |

# 4 Results

## 4.1 Throwing balls

In Fig. 5, Fig. 6, Fig. 7 and Fig. 8 the mean value of the cover time when $|K|$ is increasing and balls are thrown at it, is shown for the shapes $K^{(1)}$, $K^{(2)}$, $K^{(3)}$ and $K^{(4)}$ respectively. Each point in the figures is the mean of 1000 runs and the squares in the grid used to check if the surface is covered or not have the width $10^{-4}$. There is also a 95% confidence interval displayed in the figures. These have been created by assuming that the estimate of the mean cover time has a normal distribution, which are justified by the central limit theorem.

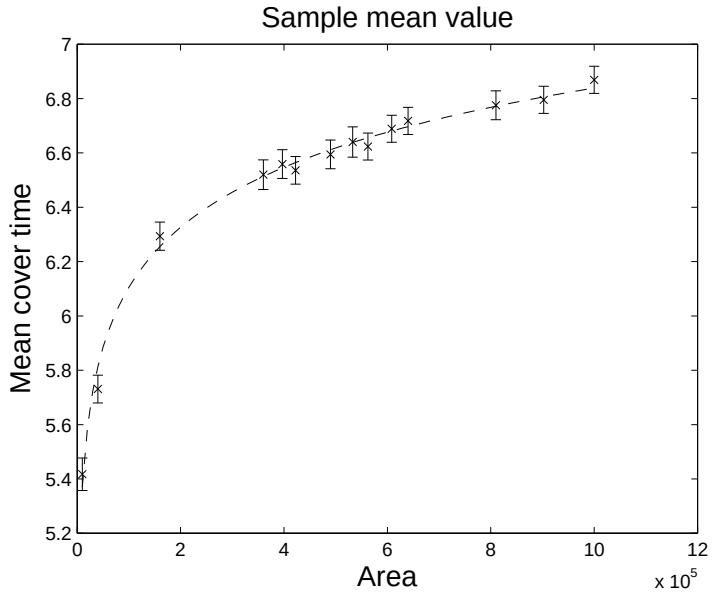



Figure 5: Mean values of the cover time for $K^{(1)}$ when balls are thrown at it. The fitted curve is $y = 0.3188 \log(|K|) + 2.4345$ and the 95% confidence intervals are shown.

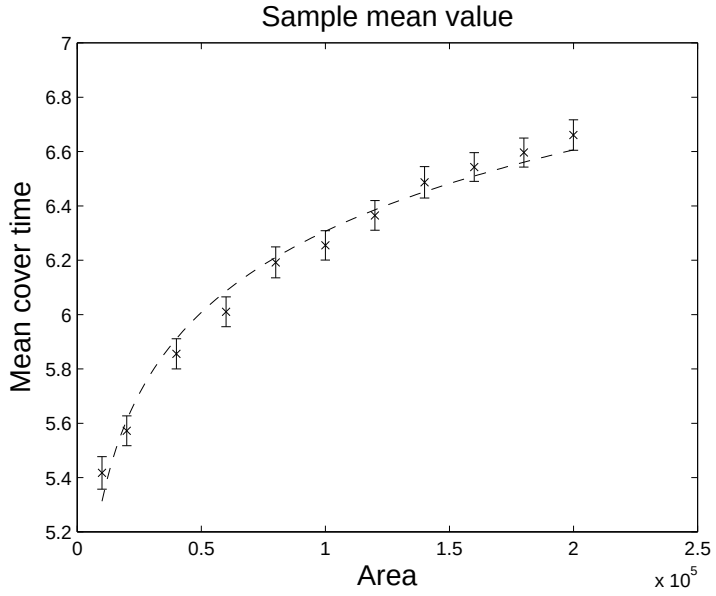



Figure 6: Mean values of the cover time for $K^{(2)}$ when balls are thrown at it. The fitted curve is $y = 0.4317 \log(|K|) + 1.3363$ and the 95% confidence intervals are shown.

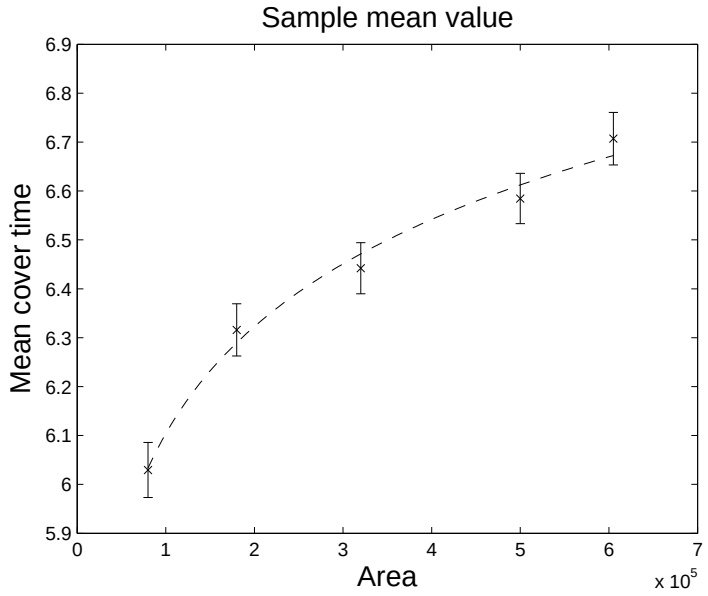



Figure 7: Mean values of the cover time for $K^{(3)}$ when balls are thrown at it. The fitted curve is $y = 0.3159 \log(|K|) + 2.4673$ and the 95% confidence intervals are shown.

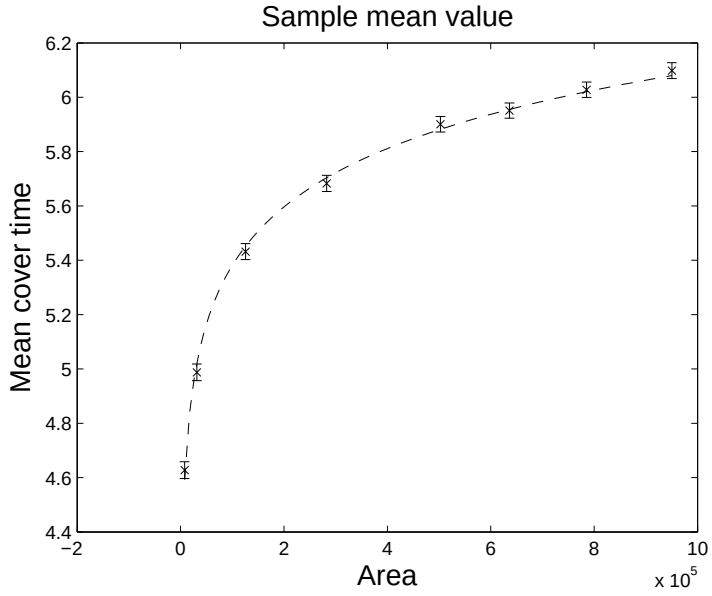



Figure 8: Mean values of the cover time of $K^{(4)}$ when balls are thrown at it. The fitted curve is $y = 0.3103 \log(|K|) + 1.8086$ and the 95% confidence intervals are shown.

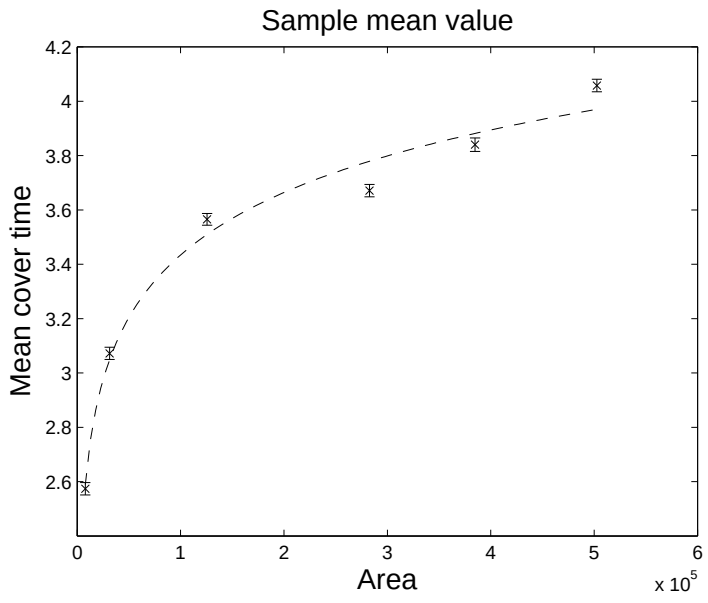



Figure 9: Circle covered by cylinder, $y = 0.3318 \log(|K|) - 0.3865$

## 4.2 Throwing cylinders

In Fig. 9 the mean value of the cover time when $|K|$ is increasing and cylinders are thrown at it, is shown for the shape $K^{(4)}$, i.e. $K$ is circular. Each point in the figures is the mean of 1000 runs and the squares in the grid used to check if the surface is covered or not have the width $10^{-4}$. There is also a 95% confidence interval displayed in the figures.

## 4.3 Two sets with increasing distance

In Fig. 10 the correlation of the cover times for two sets of type $K^{(4)}$ as the distance between them is increasing is shown. The radius of the sets are 50. Each point in the figures is the mean of 40 000 runs and the squares in the grid used to check if the surface is covered or not have the width $10^{-4}$.

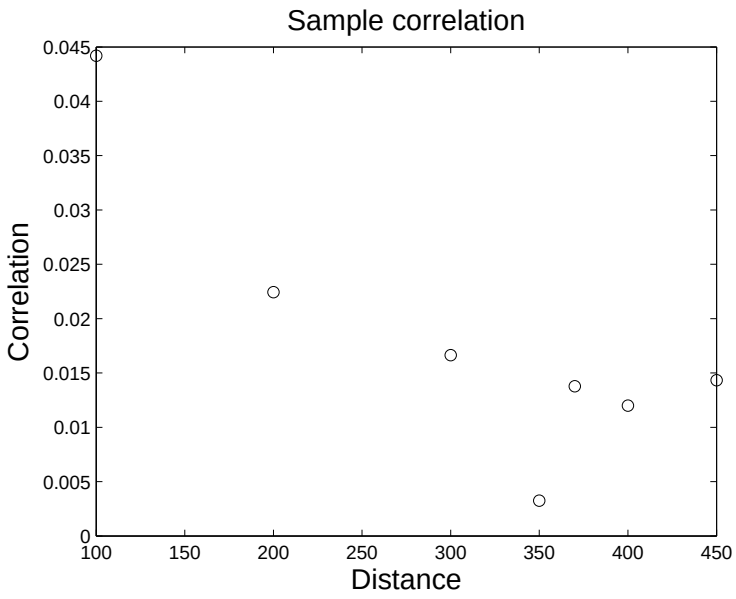



Figure 10: Correlation when covered by cylinders, two circles.

# 5 Discussion

## 5.1 Rate of increase of cover times

As noted before the mean value of $\tau_K$ seems to follow a logarithmic trend when $|K|$ increases. In the case when throwing cylinders the values of $\tau_K$ are generally lower then when throwing balls. This is not surprising as the cylinders typically covers a larger area than a ball when hitting $K$. The impact of the shape of $K$ is discussed in the next subsection.

### 5.1.1 Impact of the shape of $K$

Denote the rate of which $\tau_K$ increases for the four different shapes with $r_1$, $r_2$, $r_3$ and $r_4$. From Fig. 5 - Fig. 8 it can be seen that the $r_1$, $r_3$ and $r_4$ are all similar while $r_2$ is slightly higher. A possible reason for this can be that the area to circumference ratio are different for the different shapes.

A shape that have more circumference relative the area of the shape will have a longer cover time as when the smaller sets hits at the edge of the larger set it will only partially contribute to the covering of the set. Therefore the shapes with a lower area to circumference ratio will have a slower increase in the cover time. As an extreme case, one can consider a rectangle with height

$h = \epsilon$ and width $w = 1/\epsilon$ whose cover time will clearly diverge as $\epsilon \to 0^+$ while the area is always 1.

For shape $K^{(2)}$ the area to circumference ratio increases slower than for the other shapes. In the following table the area to circumference ratio as well as the experimentally obtained increase in $\tau_K$ is shown for the different shapes.

| Shape | Area/Circumference | Experimentally obtained increase of $\tau_K$ |
|---|---|---|
| $K^{(1)}$ | $h/4$ | 0.318 |
| $K^{(2)}$ | $100h/(200+2h)$ | 0.432 |
| $K^{(3)}$ | $h/3$ | 0.316 |
| $K^{(4)}$ | $h/2$ | 0.310 |

## 5.2 Throwing cylinders at two sets

From Fig. 10 it can, as expected, be seen that the correlation decreases as the distance between the two sets is increasing. The correlation is estimated with

$$\rho^* = \frac{\sum_{n=1}^{N}(\tau_n^{(1)} - \bar{\tau}^{(1)})(\tau_n^{(2)} - \bar{\tau}^{(2)})}{\sqrt{\sum_{n=1}^{N}(\tau_n^{(1)} - \bar{\tau}^{(1)})^2 \sum_{n=1}^{N}(\tau_n^{(2)} - \bar{\tau}^{(2)})^2}},$$

where $\tau_n^{(i)}$ is the cover time of surface $i$ in simulation $n$ and $\bar{\tau}^{(i)}$ is the mean of the cover time of surface $i$. It should though be noted that this have been a problematic simulation to perform. The uncertainty in the obtained points is large, and the longer the distance between the two sets become the more important the accuracy of the estimation become. This is because if we want to see the trend in the correlation, we need the error to be small when the correlation is small since otherwise the errors contribution is to significant. The current estimates are based on 40 000 simulation runs which takes about 16 CPU hours. It could be interesting to increase the number of simulations. This is however not conducted in this paper as time limitations will not allow it.

# References


[1] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Probability and Random Processes. OUP Oxford, 2001.

[2] Svante Janson. Random coverings in several dimensions. *Acta Mathematica*, 156(1):83–118, 1986.

[3] J.F.C. Kingman. *Poisson Processes*. Oxford Studies in Probability. Clarendon Press, 1992.

[4] L.A. Santaló. *Integral Geometry and Geometric Probability*. Cambridge Mathematical Library. Beijing World Publishing Corporation (BJWPC), 2004.