

# Asymptotic properties of coalescing random walks

Patrik Dykiel

U.U.D.M. Project Report 2005:15

Examensarbete i matematisk statistik, 10 poäng

Handledare och examinator: Ingemar Kaj

December 2005



Department of Mathematics

Uppsala University



# Asymptotic properties of coalescing random walks

Patrik Dykiel

## Abstract

A Coalescing random walk is an interacting particle system which can be described as the following. Suppose that there is one particle located at each integer position  $[0, n]$  at time point 0 and that a particle is chosen at random at any time point  $t$  among the remaining in  $[1, n]$  and moves one step in some direction, where the direction can be deterministic or stochastic. The particle coalesce with any particle that might be located at the new location. The time, or the number of iterations, until all particles has gathered at position 0, or the system has converged, is then a well defined random variable which is of some interest and which we denote by  $T_n$ .

This paper is of a investigatory nature where we study some asymptotic properties of  $T_n$  in different configurations of the process that was presented above. We use simulation in order to estimate the asymptotic expressions of  $ET_n$  and  $VarT_n$ , we furthermore investigate the distributional behavior of  $T_n$  in different configurations of the process.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Previous and current research in the field . . . . .	1
1.2	Notations and definitions . . . . .	1
1.3	Acknowledgement . . . . .	2
<b>2</b>	<b>Different settings of the process</b>	<b>3</b>
2.1	The basic setting . . . . .	3
2.2	The free setting . . . . .	4
2.3	The random movement setting . . . . .	5
<b>3</b>	<b>The simulation strategy and simulation results</b>	<b>7</b>
3.1	The estimation strategy, introductory simulation results and the functional form of $ET_n$ and $VarT_n$ . . . . .	7
3.2	The estimated parameters of $ET_n$ and $VarT_n$ . . . . .	9
3.3	The distribution of $T_n$ . . . . .	13
<b>4</b>	<b>Concluding remarks</b>	<b>18</b>
<b>A</b>	<b>Programs and routines used in the simulation</b>	<b>20</b>



# 1 Introduction

The aim with this paper is to study and present some different asymptotic properties of a special type of stochastic process which is called a *Coalescing Random Walk*. A coalescing random walk is a stochastic process where the members of the process interact in a particular way. Suppose that there are  $n$  particles, or objects, who are uniformly spread out on the integer positions of the positive real axis. Suppose further that each one of these particles conduct individual random walks on the integer positions. If any of the particles arrive at a position where another particle already is located then the two particles will merge, or coalesce. The new particle will thereafter conduct the same kind of random walk as its ancestors and the process will repeat itself in a similar fashion. There is a kind of natural end state of this process and that is, of course, when only one particle remains. This natural ending state of the process will be the subject under study in this paper.

Although the random walk is easy to understand and very simple to describe, is there a built in complexity which makes it very difficult to handle theoretically. The results that will be presented in this paper will hopefully shed some light on the behavior of different settings of the process. We will in this survey use, and rely on, the heads on approach of simulation in order to reveal some interesting convergence properties of the process in question.

## 1.1 Previous and current research in the field

Coalescing random walks are members of a larger class of stochastic processes which are denoted by *Interacting particle systems*. The roots of interacting particle systems can be traced to different modeling efforts in statistical physics and computer science in the late 1960's. One of the founding fathers to this mathematical theory is Frank Spitzer, whose research in the late sixties and early seventies presented the basic theoretical framework[Gri93]. Many contributions have been made to the theory since then. Some of the more important contributions were made by Richard Arratia, who devoted his Phd studies to the subject, and by Bramson and Griffeath whose work on the asymptotic behavior of the systems has opened new doors in order to describe it.

Recent research has also been conducted in order to describe different asymptotic properties of this kind of random walks. For example van den Berg and Kesten studied the asymptotic density of this kind of stochastic processes, see [vdBK00], and Stephenson has extended their work, see [Ste01]. Larsen and Lyons studied a simple coalescing random walk and presented their results in a paper from 1999, see [LL99]. We will in this paper take off from their work.

## 1.2 Notations and definitions

Suppose that we have a setting that is of the kind that was presented earlier. We can then introduce the following notations.

Let the vector  $\mathbf{X}(t) = \{X_1(t), X_2(t), \dots, X_n(t)\}$  denote *the state of the system at time point  $t$ ,  $t > 0$* , where  $X_i(t)$  informs if there is a particle present at position  $i$ .  $X_i(t)$  is 1 if there is a particle present and 0 otherwise and the system has the initial state  $\mathbf{X}(0) = \{1, 1, \dots, 1\}$ . We can introduce a notation for the *number of particles in the system at time point  $t$*  as  $C_t = \sum_{i=1}^n X_i(t)$ ,

where  $C_0 = n$ . We are now able to make a stringent definition of the stochastic process that we will study.

**Definition 1** *Suppose that the random variables  $X_i(t)$ ,  $1 \leq i \leq n$ , together describe the state of a system  $\mathbf{X}(t)$ . The process then changes its state at any discrete time point  $t$ ,  $t > 0$ , according to the following scheme:*

1. *an existing particle in the system is chosen at random with probability  $1/C_{t-1}$ . Let us denote the index of chosen particle with  $i$  which represents its distance from the origin.*
2. *The chosen particle moves one step in some direction, where the direction might be random, to a position with index  $j$ ,  $j < i$  or  $j > i$ . If another particle is located at  $j$  then the particle at position  $i$  will be absorbed into the particle at position  $j$  and disappears from the system. If  $j$  on the other hand is empty then the particle at  $i$  traverses to position  $j$  and remains in the system.*
3. *The procedure repeats itself until  $C_t = 1$  for some  $t > 0$ .*

The above definition of the process describes a coalescing random walk where the coalescing point of the system is arbitrary among the positions in  $[1, n]$ . If a specific position outside  $[1, n]$  is chosen as coalescing point, that is position 0 or  $n + 1$ , then the process follows the above scheme until  $C_t = 0$ .

The process that is defined above is obviously a discrete stochastic process in discrete time. We can even allow ourselves to call it iterative because of its repeating nature. From definition 1 can we also conclude that there exists an implicitly defined random variable in the process that is of interest to us. That variable is the *convergence time* of the system, which is the time, or the number of iterations, until the system has transformed itself to a state such that  $C_t = 1$ , or  $C_t = 0$  in the case with a specific coalescing point outside the interval. This random variable depends on the number of particles in the system at  $t = 0$  and we will therefore denote it by  $T_n$ , and more formally call it *the convergence time of a system of size  $n$* . We will in this paper explore the nature of  $T_n$  and give extra attention to the expected value and dispersion of the variable, but we will also study the distributional behavior of the variable.

### 1.3 Acknowledgement

The author would like to thank Professor Ingemar Kaj for the introduction to the topic as well as for his invaluable input during the work on this thesis.

## 2 Different settings of the process

The description of the process as seen in definition 1 presents the process in an uncomplicated and straight forward way. The definition also leaves a great amount of degrees of freedom to make some alterations to the configuration of the process that might present some interesting aspects in the dynamics of these kinds of processes.

The simplest setting of our process was studied by Larsen and Lyons, see [LL99]. They studied the setting where the member particles of the process only were allowed to move one step to the left in each iteration. Their configuration of the system also only allowed the particle to move as far as to the position with index zero, which was the coalescing point of the system. From studying the particles in this setting they found an exact expression for  $ET_n$  and an upper bound to  $VarT_n$ . We will get back to this subject later.

We will take off from where Larsen and Lyons stopped and study more complicated settings of the system. We will, for example, ease the movement constraints of the particles and allow them to move over the ends of our original starting interval, that is the integer positions in  $[0, n]$ . We will also add an extra complexity to the process by allowing the direction of the movement to be random. We can summarize the configurations that we will study with the following description:

- i) *The particles will be allowed to move in one direction on the interval and without barrier constraints*
- ii) *The particles will be allowed to move in a random direction with a barrier constraint*

We will denote the two cases as *the free setting* and *the random movement setting*. A short presentation of the differences between these systems along with a more detailed presentation of Larsen and Lyons setting will be made in the following sections.

### 2.1 The basic setting

In Larsen and Lyons setting, as we will regard as the basic setting, is the member particles of the system under a simple, but efficient, constraint that limits the particles possibility to move. In this basic setting can the particle with starting position  $i$  conduct a random walk of maximum length  $i$ . The reason is of course that the particle can not move further to the left than to the origin. This in turn bounds  $T_n$  to take on values in  $[n, \frac{n(n+1)}{2}]$ . The bounds will become more obvious if we study the following picture.

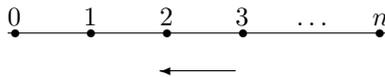


Figure 1: Structure and allowed movement of the particles in the basic setting

The particles are uniformly spread out and the arrow shows the allowed direction for the particles movement. Obviously, the number of moves, or the time, until the system has converged, or gathered at position zero, will be  $n$  if the particle at the far right moves to the left in each iteration. If, on the other hand, the particle at the far left moves to the left in each iteration, then the time until the system converges will be  $1 + 2 + \dots + n = \frac{n(n+1)}{2}$ . Every other series of events in the process will produce a value of  $T_n$  that lies between the values that were presented above. Since  $T_n$  is bounded in this system does it follow that  $ET_n$  and  $VarT_n$  exists and are finite for every fixed  $n$ . A property of this system that makes it easy to work with theoretically is that convergence of the system is guaranteed within  $\frac{n(n+1)}{2}$  iterations. That is, there is no possibility that the system will 'run wild' and iterate into infinity. The other systems that we will study in this paper do not possess this appealing property.

## 2.2 The free setting

In this setting will we allow the particles to move completely free on the interval but in just one direction. We illustrate the process with the picture below, where the arrows once again represents the possible movement patterns of the particles.

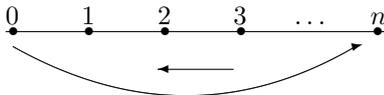


Figure 2: Structure and allowed movement of the particles in the free setting

The key difference of this system from the basic setting is, as can be seen in the illustration, that the length of each particles walk is unconstrained. Since there does not exist any barriers that hold up the particles movement, as position zero in the basic setting, is  $T_n$  unbounded in this setting.

There is another aspect of this system which easily can be realized if we study the movement patterns of the particles. We notice that if a particle is located at the origin and makes a move at a given time point then the particles new location will be position  $n$ . This allows us to make yet another illustration of the process which can be seen in figure 3.

The representation in figure 3, where the arrow show the movement of the particles, is identical with the previous regarding the movement of the particles in the system. We can therefore conclude that the domain of the interacting particles has been transformed into a circle as the barrier, the point in the domain that halted the particles movement, has been removed. How this setting may alter the behavior of the process will be investigated in following sections.

Another important difference between this configuration of the process, to the basic one, is that the location of where on the domain that system converges lacks importance. That is, we are not interested in where the particles gather but that they do.

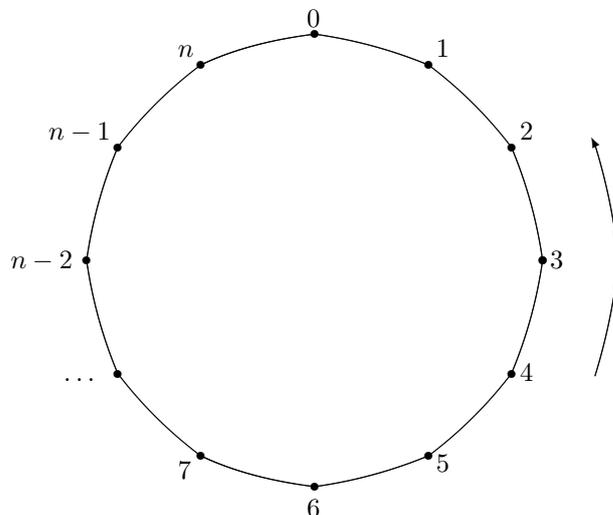


Figure 3: Alternative representation of the free setting

### 2.3 The random movement setting

We will in this setting make another slight alternation to the basic setting of 2.1. In this configuration of the system will the direction of the particles movement be Bernoulli distributed with probability  $p$ . That is, the probability that the selected particle will move to the right is  $p$ , and the corresponding probability that the particle moves to the left is  $q = 1 - p$ . Furthermore, the particles are in this setting allowed to traverse over the far right end of the interval. This makes this particular setting a kind of hybrid of the two system that were presented earlier. Although its resemblance with the basic setting is it once again impossible to bound  $T_n$  because of the semi-unconstrained movement of the particles. The notation semi-unconstrained may be a bit diffuse and may need some closer explanation. Even though we have relaxed the movement constraints of the particles even more by allowing them to move in both directions, there still exists a constraint that forbids the particles to move completely free. The constraint that limits the movement is that the particles are not allowed to make a full lap around their starting position. Since we have not removed the barrier at position zero, does it still exist a fix particle that will absorb any particle that tries to pass it. Although the movement is free, it is not completely free which motivates the above notation. An illustration of the process in this setting can be seen in figure 4 where, once again, the arrows show the allowed movement of the particles.

If we follow the particles movement patterns can we see that if a particle arrives at position zero then it will remain there. That is, the particle at position zero is a barrier (fix point) that will remain static during the systems alterations. Position zero will therefore be the coalescing point of the system. We can, however, by studying the movement pattern of the particles allow ourselves to

make an alternative presentation of the process as can be seen in figure 5.

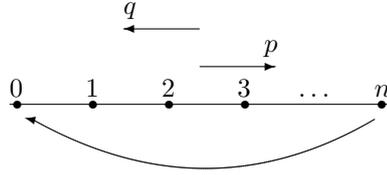


Figure 4: Structure and allowed movement of the particles in the random setting

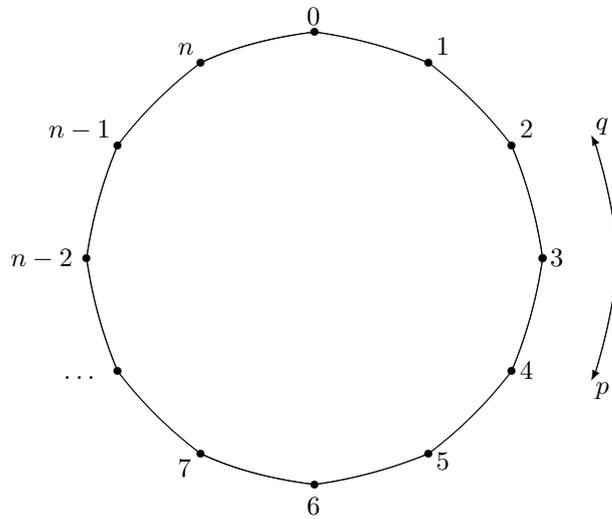


Figure 5: Alternative representation of the random movement setting

The domain on which the particles move can, as seen, once again be translated into a circle. The resemblance of this configuration to the system with the basic setting will become obvious if we set  $p = 0$ . In fact, if  $p = 1$  will we have the completely same setting as the basic one, but that the movement of the particles is reversed. So, in order to investigate how the behavior of the system alters with the 'randomization' of the particles movement must we let the particles to move completely at random. We will therefore set the direction-probabilities to be symmetric, that is  $p = q = \frac{1}{2}$ .

### 3 The simulation strategy and simulation results

The convergence time of the process,  $T_n$ , is as mentioned earlier the subject of especial interest to us. We will by simulation try to find expressions of the expected value and dispersion of this variable, as well as its distribution. We will also investigate if, and in that case how, these expressions change with the alterations of the system configuration, and if these alterations change the distributional behavior of the variable. Since we are searching for these expressions by a simulation approach do we need to know what in fact we're looking for. In the following section will the estimation strategy be introduced and we deduce the form of the sought moments. In the sections that follow is the estimated functions of the moments presented along with estimated densities and a discussion of the possible distribution of  $T_n$ .

#### 3.1 The estimation strategy, introductory simulation results and the functional form of $ET_n$ and $VarT_n$

We will in our search of  $ET_n$  and  $VarT_n$  for our three settings of the process use a very simple and straightforward approach. We will simulate estimates of  $\widehat{ET}_n$  and  $\widehat{VarT}_n$  for different values of  $n$  and by that way obtain a set of points  $(\widehat{ET}_n, n)$  and  $(\widehat{VarT}_n, n)$  on which we can fit functions. The estimation of these functions will be a kind of least square problem and we will use a nonlinear least square algorithm called *The Levenberg-Marquardt algorithm* in the fitting procedure. This algorithm is developed from a wider class of optimization algorithms that are used when solving unconstrained optimization problems, and found to be very efficient in solving least squares problems. For an introduction to the algorithm see [Mar63].

We have earlier mentioned that the value of  $T_n$  is dependent of  $n$  and it is therefore reasonable to believe that we can express  $ET_n$  and  $VarT_n$  as functions of  $n$ . A natural question that arises is of course the one regarding the explicit form of these functions. Larsen and Lyons found in their study that the expected value and variance of  $T_n$  in the basic setting was

$$ET_n = \frac{2n(2n+1)}{3} \binom{2n}{n} \frac{1}{2^{2n}} \quad (1)$$

and

$$VarT_n \leq \frac{(8 + o(1))n^{5/2}}{15\sqrt{\pi}} \quad (2)$$

respectively. They also deduced that these moments could be asymptotically expressed as

$$ET_n \sim 0.752n^{\frac{3}{2}} \quad (3)$$

and

$$VarT_n \sim Cn^{\frac{5}{2}}, \quad (4)$$

where  $C$  is some constant  $C \leq \frac{8}{15\sqrt{\pi}}$ . For proof see [LL99].

Let us now turn our attention towards some introductory estimates of  $ET_n$  and  $VarT_n$  for the three settings. The estimates that can be seen in figures 6 and 7 are based on system sizes up to 50, where the moments of  $T_n$  for each system

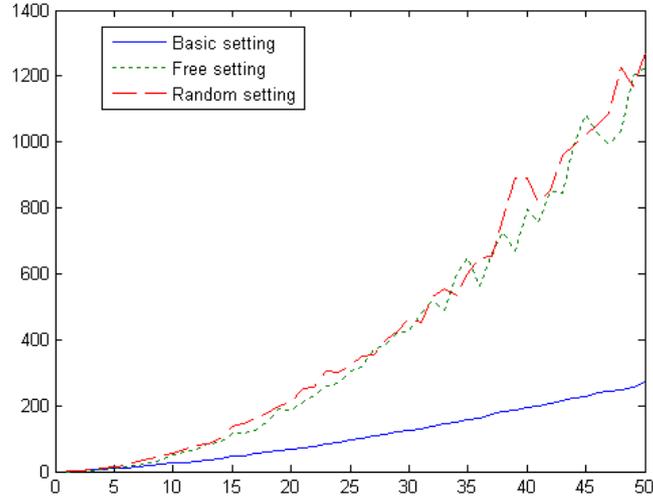


Figure 6: Estimated values of  $ET_n$  for the three settings

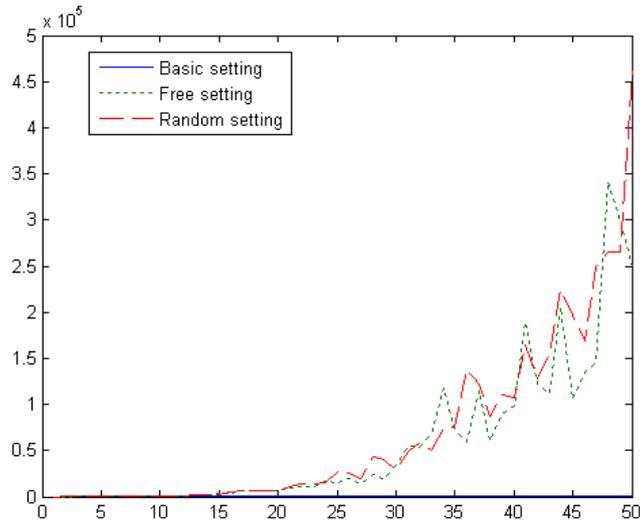


Figure 7: Estimated values of  $VarT_n$  for the three settings

size is based on 50 observations. We find, not surprisingly, that the functions are increasing with  $n$ . It is very obvious that  $ET_n$  grows more rapidly with  $n$  for the free and random setting, than for the basic setting as seen in figure 6. A natural question is now how rapidly these functions actually grow. Since  $\exp 50 \approx 5.18 \cdot 10^{21}$  can we with a peaceful mind conclude that these functions does not grow in an exponential manner, as a consequence must these functions therefore be some powers of  $n$ . We also notice that even though the growth

rate is different is the general appearance of the functions very similar. The functional form of the expected value of  $T_n$  for the free and random movement setting should therefore be the same as the corresponding for the basic setting. And because we know the asymptotic form of  $ET_n$  in the basic setting can we state that the expressions we are searching for is of the form

$$ET_n = an^b, \quad (5)$$

where  $a > 0$ ,  $b > \frac{3}{2}$ .

If we now turn our attention to figure 7 can we see the estimated values of  $VarT_n$  for the three different processes. We once again see that the functions are increasing with  $n$  and that the dispersion of  $T_n$  for the free and the random setting grows more rapidly with  $n$  than the analog for the basic setting. This can of course be explained by the fact that  $T_n$  is bounded in the basic setting and unbounded in the others. We can once again conclude that the functions does not grow in an exponential manner.  $VarT_n$  should therefore be of the same form as  $ET_n$ , but where the exponent of the size factor,  $d$ , must be greater than  $b$ . We can not do any further statements of the ordering of the functions parameters other than simply saying that the scaling parameter of  $VarT_n$ ,  $c$ , must be greater than zero. The explicit expression of  $VarT_n$  that we are searching for is therefore

$$VarT_n = cn^d, \quad (6)$$

where  $d > b$  and  $c > 0$ . The estimated parameters of  $ET_n$  and  $VarT_n$  will be presented in the next section.

### 3.2 The estimated parameters of $ET_n$ and $VarT_n$

The estimated exponents of  $ET_n$ , from 500 observations at each  $n$ , for the different processes are presented in figure 8, and the corresponding estimated scaling parameters can be seen in figure 9. We can from the figures deduce that the convergence rate of the estimates of the exponent in all of the configurations is, relatively, much higher than for the scaling constant. We can in figure 8 see that the estimates of the exponent for the three different models show a stable behavior for values of  $n$  that are 50 and larger. The exponent of  $ET_n$  for the basic setting approaches  $\frac{3}{2}$  when  $n$  grows, as presented by Larsen and Lyons, and the corresponding parameters for the other processes approaches 2.

The scaling constants for the different processes does not show an equally apparent stabilizing behavior. We notice that the estimated values of this parameter show an oscillating behavior but seems to stabilize as  $n$  increases. The estimate of the scaling parameter for the basic setting converges to a small neighborhood of its true value of 0.752, the last estimated value was 0.7658, and the estimates of the other two parameters seems to converge to a value close to 0.5, the last estimated value of the constants was 0.4879 for the free setting and 0.4661 for the random movement setting.

A fact that is given from figure 8 and 9, and that cannot be ignored, is the high convergence rate of the parameter estimates in the basic configuration compared with the others. The estimated values of the parameters has for the basic configuration stabilized for as small values of  $n$  as 50, the other two models seems however to need a bit larger  $n$  in order to their parameter estimates to stabilize properly.

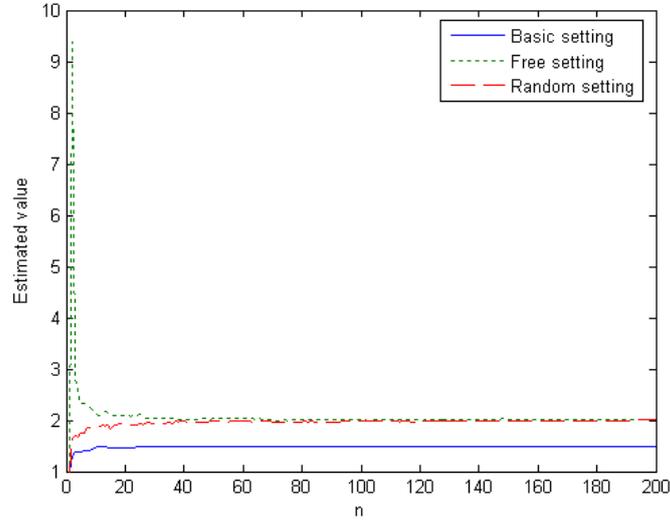


Figure 8: Estimated exponents of  $ET_n$  for the three settings

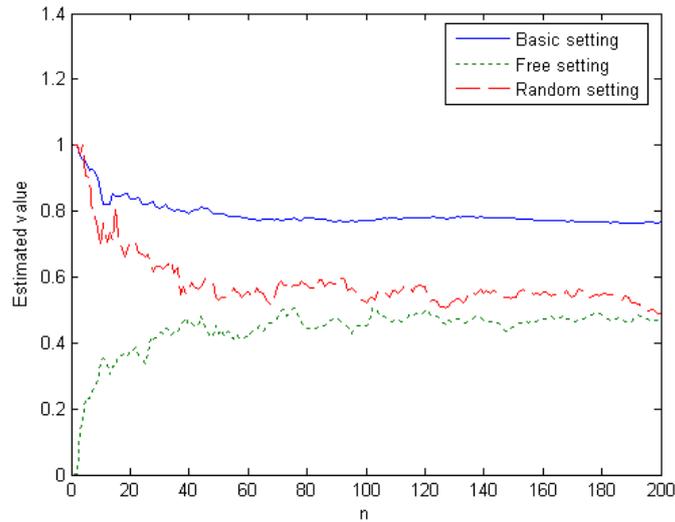


Figure 9: Estimated scaling parameters of  $ET_n$  for the three settings

If we now turn our attention towards the dispersion of  $T_n$  can we in figure 10 and 11 see the corresponding estimated parameters for the three models, where the parameters are estimated from the same number of observations at each  $n$  as for  $ET_n$ . We can in these figures instantly notice that the estimated values of the parameters in  $VarT_n$  does not show the same type of smooth behavior as the corresponding in  $ET_n$ . The convergence rates of the parameter estimates in  $ET_n$  seems however to repeat themselves in the estimations of  $c$

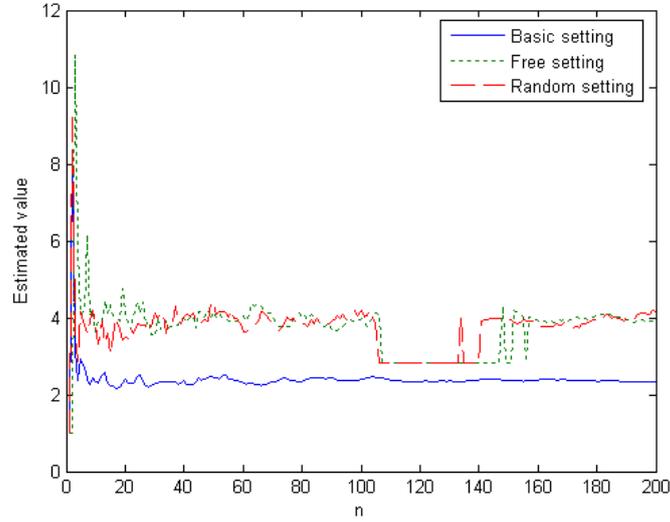


Figure 10: Estimated exponents of  $VarT_n$  for the three settings

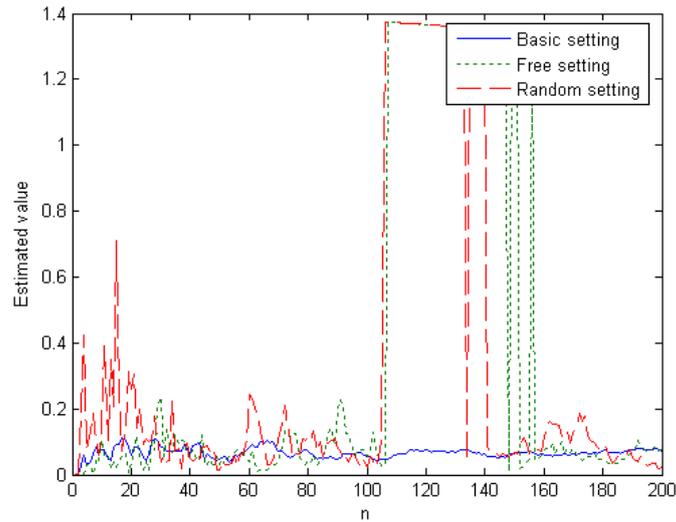


Figure 11: Estimated scaling parameters of  $VarT_n$  for the three settings

and  $d$ , in the sense that the estimates of the exponents stabilize faster than the scaling constants. The parameters in  $VarT_n$  in the basic setting once again stabilize faster than the corresponding in our other two processes. The only parameters that we with some certainty can present a numerical value to is the exponents in the different variances. The exponent in the variance of the basic setting approaches  $\frac{5}{2}$ , as shown by Larsen and Lyons, and the exponents of the variances of  $T_n$  in the free and random movement setting approaches a

value very close to 4. Their scaling constants are more troublesome to pinpoint because of the rather large oscillating behavior of their estimates. The scaling parameter for the basic setting seems to stabilize but a closer investigation shows that the parameter estimates still oscillates. Even when the estimation procedure includes values of  $n$  up to 500 does this oscillating behavior remain. It is therefore, at this point, impossible to present any asymptotic value to the parameter, but we can however state that the true value is less than 0.1.

Let us now denote  $T_n$  for the basic, free and random movement setting with  $T_{B,n}$ ,  $T_{F,n}$  and  $T_{R,n}$  respectively. We can with our simulation results now state that  $ET_n$  and  $VarT_n$  for the different configurations can be asymptotically expressed as

$$ET_{B,n} \sim 0.752n^{\frac{3}{2}} \quad (7)$$

$$VarT_{B,n} \sim Cn^{\frac{5}{2}} \quad (8)$$

$$ET_{F,n} \sim a_1n^2 \quad (9)$$

$$VarT_{F,n} \sim C'n^4 \quad (10)$$

$$ET_{R,n} \sim a_2n^2 \quad (11)$$

$$VarT_{R,n} \sim C''n^4 \quad (12)$$

where  $C < 0.1$ ,  $C' < 0.2$ ,  $C'' < 0.2$  and where  $a_1$  and  $a_2$  are approximately 0.5. Our suggested upper bound of  $C$  in  $VarT_{B,n}$  is consistent with Larsen and Lyons results since  $\frac{8}{15\sqrt{\pi}} \approx 0.3 > 0.1$ .

Even though the Levenberg-Marquardt method was not able to estimate all of the parameters of  $ET_n$  and  $VarT_n$  for our different configurations of the process, can we by a proper re-scaling of our random variables try to obtain more precise estimates of our sought scaling parameters. Our previous simulation results tells us that

$$Var\left(\frac{T_{B,n}}{n^{\frac{5}{4}}}\right) \sim C \text{ as } n \rightarrow \infty,$$

and moreover that

$$E\left(\frac{T_{F,n}}{n^2}\right) \sim a_1, \quad Var\left(\frac{T_{F,n}}{n^2}\right) \sim C' \text{ as } n \rightarrow \infty.$$

Where a similar result is obtained for  $T_{R,n}$ . So, we can by calculate the mean and sample variance of the re-scaled variables, hopefully, obtain more precise and useful estimates of the scaling parameters. Estimated values of the scaling parameters with this approach can be seen in table 1 and 2 below.

$n$	Estimated $C$
100	0.0359
500	0.0279
1000	0.0276
2000	0.0263

Table 1: Estimated values of  $Var\left(\frac{T_{B,n}}{n^{\frac{5}{4}}}\right)$ , calculated from 500 observations

Variable	Mean	Variance
$\frac{T_{F,100}}{100^2}$	0.4928	0.0479
$\frac{T_{R,100}}{100^2}$	0.4966	0.0463
$\frac{T_{F,500}}{500^2}$	0.5051	0.0510
$\frac{T_{R,500}}{500^2}$	0.4967	0.0496
$\frac{T_{F,1000}}{1000^2}$	0.5010	0.0422
$\frac{T_{R,1000}}{1000^2}$	0.5195	0.0564

Table 2: Estimated values of scaling parameters in (9)-(12) calculated from 500 observations

Our initial idea of the asymptotic value of  $C$  in  $VarT_{B,n}$ , as was presented by the search algorithm, was not completely far off. The estimated values of this parameter for different system sizes, as seen in table 1, shows however that the value of this parameter in fact is a bit smaller than 0.1. The estimates of  $C$  decreases as  $n$  grows but the estimates for the three largest system sizes all lies close to 0.026, which makes it reasonable to believe that the asymptotic value of  $C$  is approximately 0.026.

The estimates of the scaling parameters in the expected value and variance of  $T_{F,n}$  and  $T_{R,n}$  does present a bit more accurate image of their asymptotic values. All of the estimated values for the scaling constant of the expected value lies close to 0.5, and the corresponding parameter value for the scaling constant of the variance seems to be close to 0.05. We can therefore formulate the following conjecture.

**Conjecture 1** *The asymptotic variance of  $T_{B,n}$  can be expressed as*

$$VarT_{B,n} \sim Cn^{\frac{5}{2}},$$

where  $C \approx 0.026$  and the asymptotic expected value and variance  $T_{F,n}$  and  $T_{R,n}$  can be expressed as

$$ET_{F,n} \sim a_1n^2, ET_{R,n} \sim a_2n^2,$$

and

$$VarT_{F,n} \sim C'n^4, VarT_{R,n} \sim C''n^4$$

respectively where  $a_1, a_2 \approx 0.5$  and  $C', C'' \approx 0.05$ .

### 3.3 The distribution of $T_n$

Larsen and Lyons did in their study of the basic setting a presumption where they stated that  $T_n$  obeys a central limit theorem. We will in this section

empirically investigate if this presumption holds and investigate if the same presumption holds for  $T_{F,n}$  and  $T_{R,n}$ .

We will investigate if the normality assumption holds by studying the estimated density functions of  $T_n$  for the different processes. The estimation of the density function of  $T_n$  will be conducted by using a similar approach to the one we used in our estimation of  $ET_n$  and  $VarT_n$ . That is, we will simulate a number of observations from some  $T_n$  and use a kernel smoother in order to acquire a nonparametric estimate of  $f_{T_n}$ . The estimates of the density will be based on varying sample sizes of observations from systems of varying sizes. The reason to this approach is to investigate if  $T_n$  asymptotically tends to some distribution as the size of the system grows,  $n \rightarrow \infty$ , or as the sample size grows,  $m \rightarrow \infty$ , or a combination of the both.

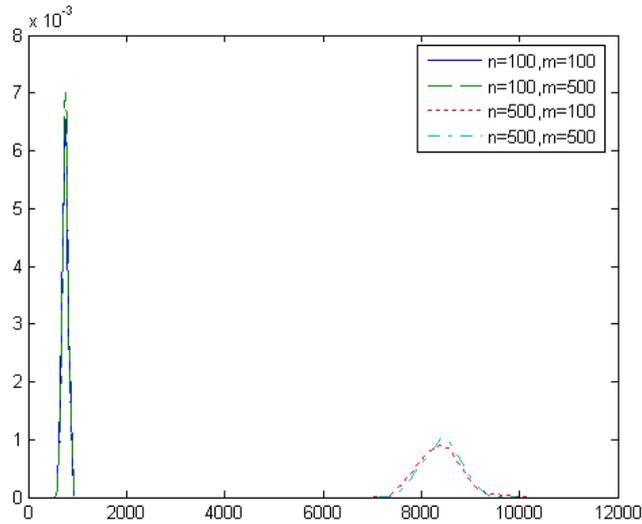


Figure 12: Estimated density functions for  $T_{B,n}$  from different sample sizes and varying system sizes

The first variable that we will study is  $T_{B,n}$  and the estimated densities for different sample and system sizes can be seen in figure 12. The densities have a symmetric appearance and the location and dispersion of the the larger systems are both bigger than for the smaller systems. This behavior of the densities is consistent with our results in the previous section since  $ET_{B,n}$  and  $VarT_{B,n}$  are dependent of  $n$ . The normality assumption seems reasonable at this point because of the symmetry in the estimated densities, and if this assumption is true then

$$\frac{T_{B,n} - ET_{B,n}}{\sqrt{\frac{VarT_{B,n}}{C}}} = \frac{T_{B,n} - 0.752n^{\frac{3}{2}}}{n^{\frac{5}{4}}} \sim N(0, C) \quad (13)$$

holds when  $n$  is sufficiently large, and  $C \approx 0.026$  by our simulation results. The estimated densities of the re-scaled variables can be seen in figure 13 along with the density of  $N(0, 0.026)$ .

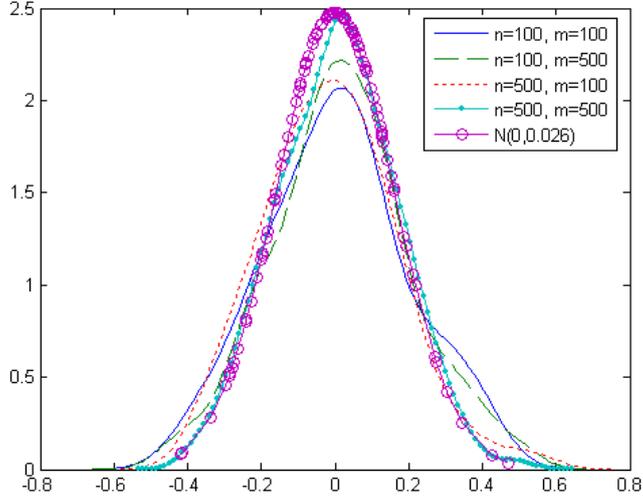


Figure 13: Estimated density functions for (13) from different sample sizes and varying system sizes, along with a  $N(0, 0.026)$

Figure 13 reveals that the sample size,  $m$ , in general does not have a big influence on the estimated densities. The densities for both system sizes become a bit softer as the sample size increases but the symmetry does not change. The system size seems to have a bigger influence on the density since the estimates become more robust for the larger systems and does not show any asymmetric behavior, as the departure from the normal density that the smaller systems show in the right part of the figure. We also see that the densities from the largest system in general agrees more with the appearance of the  $N(0, 0.026)$ . As a consequence can we make the following conjecture.

**Conjecture 2**

$$\frac{T_{B,n} - ET_{B,n}}{\sqrt{\frac{\text{Var}T_{B,n}}{C}}} = \frac{T_{B,n} - 0.752n^{\frac{3}{2}}}{n^{\frac{5}{4}}} \xrightarrow{d} N(0, C) \text{ as } n \rightarrow \infty,$$

where  $C \approx 0.026$ .

Let us now turn our attention towards  $T_{F,n}$  and  $T_{R,n}$ . Since their expected values and variances depend on  $n$  do we need to make a suitable re-scaling of these variables in order to obtain their estimated density functions on the same scale. And because we know that there is an almost quadratic relationship between their expected value and variance might it be suitable to study the densities of the random variables divided by the factor  $n^2$ . The estimated density functions for  $\frac{T_{F,n}}{n^2}$  and  $\frac{T_{R,n}}{n^2}$  can be seen in figure 14 and 15.

Both the figures show a skewed form of the estimated density functions, for all sample and system sizes, which makes an assumption of normality highly inappropriate. The peaks of the density functions are located around 0.5 which strengthens our earlier discoveries.

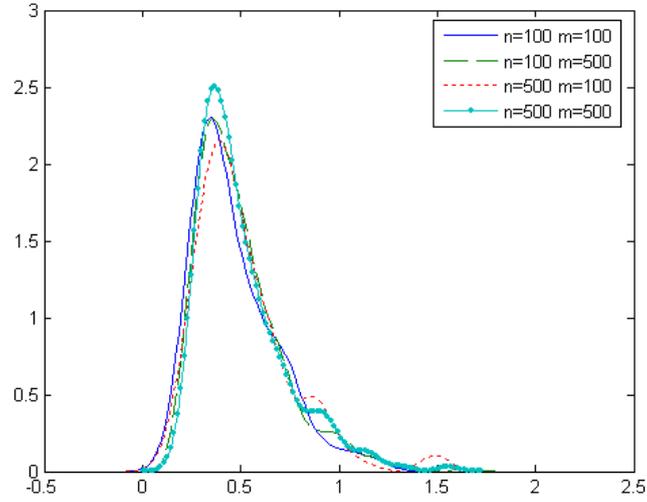


Figure 14: Estimated densities for  $\frac{T_{F,n}}{n^2}$  from different sample and system sizes

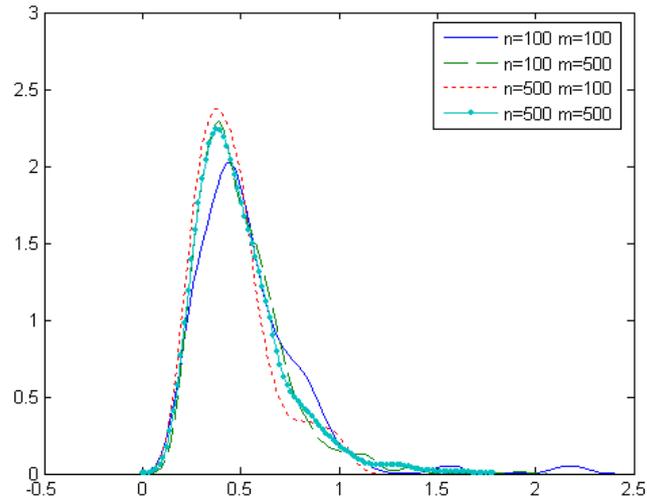


Figure 15: Estimated densities for  $\frac{T_{R,n}}{n^2}$  from different sample and system sizes

Since the estimated densities and moments are strikingly similar for  $\frac{T_{F,n}}{n^2}$  and  $\frac{T_{R,n}}{n^2}$  is it at this point reasonable to believe that these variables follows the same, or at least very similar, distributions. It is of course difficult at this point to deduce which distribution each of these random variables follow, but the estimated densities do however suggest that the distribution is negatively skewed. There is furthermore an almost quadratic relationship between the expected value and variance which narrows down the possible candidates.

Two very reasonable distributions are the *Gamma distribution* and the *Weibull distribution*, and we will denote random variables that follow these distributions as  $X$ , for the gamma distribution, and  $Y$ , for the Weibull distribution. We know that the expected value and variance for these random variables are

$$EX = pa \quad (14)$$

$$VarX = pa^2 \quad (15)$$

$$EY = \alpha^\beta \Gamma(\beta + 1) \quad (16)$$

$$VarY = \alpha^{2\beta} (\Gamma(2\beta + 1) - \Gamma(\beta + 1)^2), \quad (17)$$

where  $p, a, \alpha, \beta > 0$ , and our simulation results suggests that

$$E\left(\frac{T_{F,n}}{n^2}\right) = E\left(\frac{T_{R,n}}{n^2}\right) = \frac{1}{2} \text{ and } Var\left(\frac{T_{F,n}}{n^2}\right) = Var\left(\frac{T_{R,n}}{n^2}\right) = \frac{1}{20}.$$

By solving the obtained equations do we get  $p = 5$ ,  $a = \frac{1}{10}$  for the gamma case, and  $\alpha \approx 1.86$ ,  $\beta \approx -2.49$  for the Weibull case which is an unallowed solution. Hence is it very reasonable to believe that  $\frac{T_{F,n}}{n^2}$  and  $\frac{T_{R,n}}{n^2}$  follows a gamma distribution. The density for the  $\Gamma(5, \frac{1}{10})$  along with the estimated densities can be seen in figure 16 below.

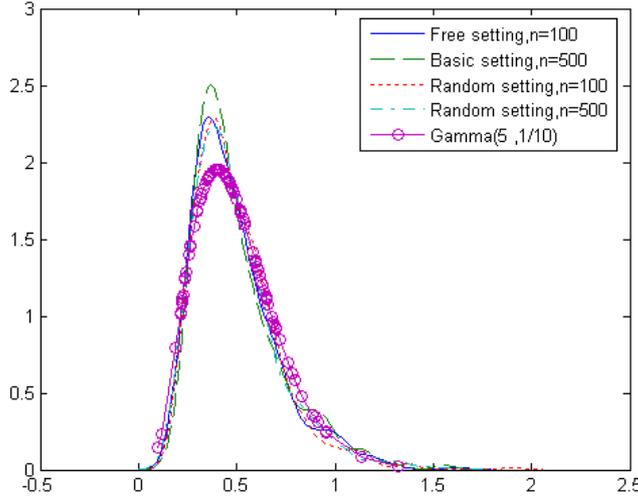


Figure 16: Estimated densities for  $\frac{T_{R,n}}{n^2}$  and  $\frac{T_{F,n}}{n^2}$  based on 500 observations for different  $n$  and the proposed gamma distribution

The gamma distribution agrees very well to our estimated densities and our newly acquired information now makes it possible to us to propose the following conjecture.

### Conjecture 3

$$\frac{T_{F,n}}{n^2} \xrightarrow{d} \Gamma(p_1, a_1) \text{ and } \frac{T_{R,n}}{n^2} \xrightarrow{d} \Gamma(p_2, a_2) \text{ as } n \rightarrow \infty,$$

where  $p_1 a_1 = p_2 a_2 \approx \frac{1}{2}$  and  $p_1 a_1^2 = p_2 a_2^2 \approx \frac{1}{20}$ .

## 4 Concluding remarks

That  $ET_{B,n}$  and  $VarT_{B,n}$  is smaller for any given system size than the corresponding for  $T_{F,n}$  and  $T_{R,n}$  is natural. Let us for simplicity denote  $\mathbf{Y}_{n,k}$  as the set of all states where there are  $k$  particles remaining in a system of size  $n$ . If we study the movement patterns in the configurations under study can we, after some thought, conclude that the basic configuration of the system only can be every state in  $\mathbf{Y}_{n,k}$  once, for a fix  $k$ . In other words, given that the system has transformed itself from a state having  $k+1$  particles to a state with  $k$  particles, let us denote it by  $y_{n,k}$ , can the system never go back to the state  $y_{n,k}$  as the process proceeds. The basic setting is therefore in some sense state consuming. Since the particles only traverse to the left in each iteration does it follow that the process successively transform itself into a state from which a transformation to a state in  $\mathbf{Y}_{n,k-1}$ , say  $y_{n,k-1}$ , is guaranteed. One could say that the probability that the number of member particles will decrease grows as the iterations increase. The same does not hold for the free and random movement setting since they are state recurrent. There exists, in theory, a possibility that all of the remaining  $k$  particles in the free setting will make a full lap from some state and come back to the same state without encountering each other. The possibility that the system will have the same state more than once is even greater in the random movement setting since this event can occur if the selected particle makes two 'jumps' and has empty neighboring positions. Since these processes are state recurrent does not the probability of a transition to some element in  $\mathbf{Y}_{n,k-1}$  approach one with the systems alterations as in the basic setting. As a consequence will these configurations of the system require a larger amount of iterations in order transform into a state with less particles when  $n$  is large and  $k$  is small.

An explanation to the identical, or very similar, results that we have observed between the free and random movement setting is more difficult to present. But the explanation to this phenomenon should be obtained if the transition probabilities that the system transforms itself from some  $y_{n,k}$  to another  $y_{n,k-1}$  was studied. Since the systems are different in the meaning that they alternate in different fashions must these probabilities be identical since, for example, their expected values are the same. Another possibility is that the actual transition probabilities are different but that the partial expected convergence times sum up to the same expression. An outline of a theoretical study that has the aim to explain these results could be as by partitioning the process by regarding the random walks between  $\mathbf{Y}_{n,k}$  and  $\mathbf{Y}_{n,k-1}$  as separate random walks. With this approach could the Markovian theory be used since the probability that the process will take on a certain state only depends on its previous state. Studies of this theoretical kind was not the aim of this paper and the author leaves this work to the interested reader.

## References

- [Gri93] David Griffeath. Frank spitzer's pioneering work on interacting particle systems. *The Annals of Probability*, 21(2):608–621, 1993.
- [LL99] Michael Larsen and Russel Lyons. Coalescing particles on an interval. *Journal of Theoretical Probability*, 12(1):201–205, 1999.
- [Mar63] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.
- [Ste01] David Stephenson. Asymptotic density in a threshold coalescing and annihilating random walk. *The Annals of Probability*, 29(1):137–175, 2001.
- [vdBK00] J. van den Berg and Harry Kesten. Asymptotic density in a coalescing random walk model. *The Annals of Probability*, 28(1):303–352, 2000.

## A Programs and routines used in the simulation

When simulating the value of  $T_n$  and estimating  $ET_n$  and  $VarT_n$  for the different configurations has MATLAB been used. The used program consists of two parts, the first part is general and is used for all configurations. The general part of the program can be seen below.

```
n=input('Specify the largest system size ');
%Asks the user to specify the largest system size
m=input('Specify the number of simulations at each n ');
%Asks the user to specify the sample size at each n
for j=1:n
    %For each system size
    for i=1:m
        %the program calculates the m values of the convergence time for
        %each n here.

    end

    tid(j)=mean(y);
    %Saves the estimated value of ET_n for each n in the vector tid.
    antal(j)=j;
    %Saves the value of n in the vector antal
    vartid(j)=var(y);
    %Saves the estimated value of VarT_n in the vector vartid.
    ansats=[1,1];
    %Makes the initial assumption that the functions ET_n and VarT_n
    %are linear.
    [u,fel]=lsqcurvefit(@funktion,ansats,antal,tid);
    [k,fel]=lsqcurvefit(@funktion,ansats,antal,vartid);
    konstant(j)=u(1);
    exponent(j)=u(2);
    varkonstant(j)=k(1);
    varexponent(j)=k(2);
    %Estimates the parameters of ET_n and VarT_n for values of n so far.
    %The paramters for ET_n are saved in the vectors konstant and exponent,
    %while the parameters of VarT_n are saved in varkonstant and
    %varexponent.
end
ansats=[1,1];
[x,resnorm]=lsqcurvefit(@funktion,ansats,antal,tid);
[k,resnorm]=lsqcurvefit(@funktion,ansats,antal,vartid);
%Estimates ET_n and VarT_n once again for all n and saves the parameters in
%the vectors x and k.
```

How the actual values of  $T_n$  are simulated depends on the configuration of the process. The code for simulating  $T_{B,n}$  can be seen below.

```
iteration=0;
%Sets T_n to be 0.
```

```

x=ones(1,j);
    %Creates the initial state.
while length(find(x))>0
    %While there is a particle remaining in the system the following
    %actions are done.
        z=unidrnd(length(find(x)),1,1);
        %A particle is chosen at random from the remaining.
        u=find(x);
        %Saves the index of the remainig particles in vector u.
        if u(z)==1
            %If the particle at position 1 is chosen
            x(1)=0;
            iteration=iteration+1;
            %Updates the value of T_n
        elseif u(z)>1
            %If a particle at another position is chosen
            x(u(z)-1)=1;
            x(u(z))=0;
            iteration=iteration+1;
            %Updates T_n
        end
    end
end
y(i)=iteration;
    %Saves the m values of T_n in the vector y that will be used when
    %estimating ET_n and VarT_n.

```

The code for simulating  $T_{F,n}$  is:

```

iteration=0;
%Sets T_n to be 0.
x=ones(1,i);
%Creates the initial state.
while length(find(x))>1
    %While there is more than one particle remaining in the system
    %the following actions are done.
        u=find(x);
        %Saves the index of the remainig particles in vector u.
        z=unidrnd(length(find(x)),1,1);
        %A particle is chosen at random from the remaining.
        if u(z)==1
            x(1)=0;
            x(i)=1;
            %If the particle at position 1 is chosen then it is
            %moved to position n.
            iteration=iteration+1;
            %Updates T_n
        elseif u(z)~1
            x(u(z)-1)=1;
            x(u(z))=0;
            %If a particle at another position is chosen then it is
            %moved to the left.
        end
    end
end

```

```

        iteration=iteration+1;
        %Updates T_n
    end
end
y(j)=iteration;
%Saves the m values of T_n in the vector y that will be used when
%estimating ET_n and VarT_n.

```

The code for simulating  $T_{R,n}$  is:

```

iteration=0;
    %Sets T_n to be 0.
    x=ones(1,i);
    %Creates the initial state.
    while length(find(x))>0
        %While there is a particle remaining in the system
        %the following actions are done.
        u=find(x);
        %Saves the index of the remainig particles in vector u.
        z=unidrnd(length(find(x)),1,1);
        %A particle is chosen at random from the remaining.
        dir=unifrnd(0,1);
        %Generates the direction of the particles movement
        if dir<=0.5
            %If the direction is left then the following actions are
            %done.
            if u(z)==1
                %If the particle at position 1 is chosen
                x(1)=0;
                iteration=iteration+1;
                %Updates T_n
            elseif u(z)~=1
                %If a particle at another position is chosen
                x(u(z)-1)=1;
                x(u(z))=0;
                iteration=iteration+1;
                %Updates T_n
            end
        elseif dir>0.5
            %If the direction is right then the following actions are
            %done.
            if u(z)==i
                %If the particle at position n is chosen
                x(i)=0;
                iteration=iteration+1;
                %Updates T_n
            elseif u(z)~=i
                %If a particle at another position is chosen
                x(u(z)+1)=1;
            end
        end
    end

```

```
        x(u(z))=0;
        iteration=iteration+1;
        %Updates T_n
    end
end
end
y(j)=iteration;
%Saves the m values of T_n in the vector y that will be used when
%estimating ET_n and VarT_n.
```

Besides the programs above has existing routines in MATLAB been used to calculate the results in this thesis.