Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Differential Equations for Random Processes and Applications for the Balls and Bins Model

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

BACHELOR OF SCIENCE
in
APPLIED MATHEMATICS

by

Diederik Gerrit Pieter van Engelenburg

Delft, the Netherlands
August 2017

BSc thesis APPLIED MATHEMATICS

"Differential Equations for Random Processes and Applications for the Balls and Bins Model"

DIEDERIK GERRIT PIETER VAN ENGELENBURG

Delft University of Technology

Supervisor

Dr. W.M. Ruszel

Thesis committee

Dr. B. van den Dries   Dr. J. L. A. Dubbeldam

August, 2017   Delft

Differential Equations for Random Processes and Applications for the Balls and Bins Model

Diederik G.P. van Engelenburg

August 31, 2017

Contents

Chapter 1

Introduction

"International trade has long been concerned with aggregated patterns - what and how much countries trade with each others." (Armenter & Koren, 2014, p. 2127) Trade from one country to any other can be categorized by the country to which it is traded and of what type it is. This "categorical nature" of the data is important, because "the number of observations is low relative to the number of possible classifications" (Armenter & Koren, 2014, p. 2128). I model trade from on country to others by "the assignment of observations to categories as balls falling into bins" (Armenter & Koren, 2014, p. 2128). The observation is a "discrete unit" of trade - the ball - and is thrown into one of the bins, which represent "mutually exclusive categories" (Armenter & Koren, 2014, p. 2128). The probability of a ball falling into a bin is determined only by the size of the bin and the probability of a ball hitting a particular trade type is the same over all countries. The model is usefull, because it can be used to indentify theories when data is sparse and tells "which statistics are driven by the sparsity [of the data ...] and those that are not - and require a model to posit the correct joint distribution across categories in order to reproduce the fact" (Armenter & Koren, 2014, p. 2150).

However, the balls and bins model is not used just for modelling trade. The model can be seen as "the classical methafor for the multinomal distribution" (Corrado, 2011) and it can be used for a goodness-of-fit test as described by Ogay (2016, p. 25). This test is based on the range of the multinomial distribution, which can be approximated by the range in the balls and bins model[1] (Ogay, 2016, p. 39).

In chapter 3, I present a theorem by Wormald (1997) which I will apply on the margins in the balls and bins model. The theorem itself is quite general for random processes linked to random graph processes, and establishes "a connection between [random variables defined on the process]

[1]Here, the range is the difference between the maximum amount of balls and the minimum amount of balls in any bin.

and an associated differential equation or system of differential equations" (Wormald, 1997, p. 3), of which the solutions approximate the actual value of the random variables. The theorem is for some "sequence of random processes indexed by n"(Wormald, 1997, p.3) (in this thesis n is the amount of bins), for which one wants to know what happens when $n \to \infty$, and it turns out that, given some particular paramters, the random variables are often "sharply concentrated at almost any given time" (Wormald, 1995, p.1217) near the solutions of the differential equation(s).

First, I prove the general theorem as presented by Wormald (1997, p.35, theorem 5.1), which was written down compact and, to my mind, incomplete. I offer a version of the proof which shows percise bounds (which I use for the applications) and proves all statements made by Wormald. Moreover, I update and replace some assumptions in the general version of the theorem, and show that one assumption in original version of the theorem (Wormald, 1995, p. 1219, theorem 1, 2) can be left out completely. The complete proof is the most important part of the thesis.

Second, in chapter 4, I present two models that are close to the balls and bins model. The connection to this model is new, as the theorem comes from theory on random graph processes. The first model assumes that the bin-sizes are all equal and the random variables on this process are the amount of bins with l balls after t balls are thrown. The second model is a slight modification of the balls and bins model, by changing the probability that a particular bin is hit with some amount of balls at each discrete step in time from t to $t+1$ to something relative to the amount of balls in the bin at time t. The random variables in this model are the amount of balls in particular bins. The advantage of this model is that it can be used for any (starting) distribution of bin sizes.

For both models, I show that the random variables meet the assumptions of Wormald's general theorem (the one as in (Wormald, 1997)) and then describe the maximum, minimum and range distribution in terms of these variables. The results obtained are usefull when the amount of balls thrown is relatively small compared to the number of bins and when the number of bins is quite large, because in those cases the variables are concentrated close enough to the solutions in the differential equations with high probability.

Last, I offer future research possibilities that can be used to offer better bounds for the concentration of the random variables, because it seems that these could be found by either slightly modifying the models or the proof of Wormald's theorem.

Chapter 2

Mathematical Framework

This chapter lays down the most important definitions and theorems needed to understand the work presented later-on. What is displayed below are assumptions and definitions that I use, as to remove any ambiguity. For specific proofs and theorems, using the same definitions, I refer to the original works (Wormald, 1995, and 1997).

2.1 Real Analysis

Real Analysis and probability theory are closely linked. To truly understand probability, measure theory is needed . Therefore, I show the broader definitions (of measure and measurable functions) before showing the probabilistic ones (probability measure and random variable respectively).

Definition 2.1. *A function $f : D \to \mathbb{R}$, $D \subset \mathbb{R}^n$ is called lipschitz (contiuous) on D in l^∞ with contstant $L > 0$ if*

$$|f(x_1, \dots, x_n) - f(y_1, \dots, y_n)| \leq L \max_{1 \leq i \leq n} |x_i - y_i|$$

for all $(x_1, \dots, x_n), (y_1, \dots, y_n) \in D$.

There is another definition of a function being lipschitz continuous, but in the l^1 space. The results in theorem 3.3 and 3.1 are the same for both. For the l^1 norm, change $L \max_{1 \leq i \leq n} |x_i - y_i|$ to $L \sum_{i=1}^{n} |x_i - y_i|$ in the definition above.

For the following definitions, assume that Ω is some sample space.

Definition 2.2. *A collection Σ of subsets of Ω is called σ-algebra if*

- $\Omega \in \Sigma$
- *for every $A \in \Sigma$, $\Omega \setminus A \in \Sigma$.*
- *for every $A_1, \dots, A_n, \dots \in \Sigma$, $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.*

There also exist σ-algebra's generated by a set $B \subset 2^{\Omega}$ (written as $\sigma(B)$), which is the smallest σ-algebra containing B, where 2^{Ω} is the powerset of Ω (which is, itself, a σ-algebra when Ω is countable). By abuse of notation, sometimes $\sigma(C)$ is written as the σ-algebra generated by a set $C \subset \Omega$. What is meant is the σ-algebra generated by $\{C\}$.

Definition 2.3. *Take Σ a σ-algebra*[1]*. A function $\mu : \Sigma \to [0, \infty]$ is called a measure if*

- *for the emptyset $\emptyset \in \Sigma$, $\mu(\emptyset) = 0$,*
- *for all $E_1, \dots, E_n, \dots \in \Sigma$ pairwise disjoint,*

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$$

Definition 2.4. *Let Ω, E be some sets and Σ, Λ σ-algebra's on Ω, E respectively. A function $f : \Omega \to E$ is called measurable if for every $B \in \Lambda$, $f^{-1}(B) = \{x \in \Omega : f(x) \in B\} \in \Sigma$.*

Next to the abstract definitions above, this thesis often uses asymptotics, for which the conventional big- and small O-notation is used.

Definition 2.5. *Let $f : \mathbb{N} \to \mathbb{R}$ be some function. Define $O(f(n))$ as the set that contains all functions $g : \mathbb{N} \to \mathbb{R}$ such that there exists some $C > 0$ with*

$$|g(n)| \leq C|f(n)|,$$

and, if there exists some n_0 such that $f(n) \neq 0$ for all $n \geq n_0$, define $o(f(n))$ as the set that contains all functions $g : \mathbb{N} \to \mathbb{R}$ such that

$$\lim_{n\to\infty} \frac{g(n)}{f(n)} = 0$$

Clearly, the notion of g being $o(f(n))$ is stronger than that of $O(f(n))$, hence $o(f(n)) \subset O(f(n))$. Often I write $g(n) = o(f(n))$ for $g(n) \in o(f(n))$, as does Wormald (1995, p. 1219). Moreover, the notation of O and o is used for $n \to \infty$ and uniform over all other variables for which a function may be defined in this thesis.

2.2 Probability Theory

In probability theory, often the 'universe' worked in is denoted by the set Ω, wich does not need to have any particular structure (Jacob & Protter, 2004, p. 3, 7). Intuitively, this is "the state space, [...] the set of all possible outcomes" (Jacob & Protter, 2004, p. 3)

[1]Over Ω, that is.

Definition 2.6. *Let Σ be some σ-algebra on Ω. The measure $\mathbb{P} : \Sigma \to [0,1]$ is a probability measure if $\mathbb{P}(\Omega) = 1$.*

Often, $(\Omega, \Sigma, \mathbb{P})$ is called a probability triplet or probability space. Here, Σ is a σ-algebra on Ω and $\mathbb{P}$ a probability measure on Σ.

Definition 2.7. *A set $A \subset \Omega$ is said to happen almost surely (a.s.) if $\mathbb{P}(\Omega \setminus A) = 0$*

Lemma 2.1. *Let $(\Omega, \Sigma, \mathbb{P})$ be some probability triplet and $B \in \Sigma$ such that $\mathbb{P}(B) > 0$. Define the conditional probability $\mathbb{P}_B : \Sigma \to [0,1]$ by*

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

then $\mathbb{P}_B$ is a well-defined probability measure.

Instead of $\mathbb{P}_B(A)$, most of the time the notation $\mathbb{P}(A|B)$ is used.

Proof. Because Σ is a σ-algebra, it is closed under (countable) intersections. Since $1 \geq \mathbb{P}(B) > 0$, and because $A \cap B \subset B$ thus (by monotonicity of the probability measure) $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$, it must hold that

$$0 \leq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1,$$

hence $\mathbb{P}_B$ maps into $[0,1]$. It is clear that $\mathbb{P}_B(\Omega) = 1$ and $\mathbb{P}_B(\emptyset) = 0$ because $\mathbb{P}$ is a probability measure.

Last, take $\{A_i\}_{i=1}^{\infty}$ a pairwise disjoint family of sets in Σ. Clearly, $\{A_i \cap B\}_{i=1}^{\infty}$ is also a pairwise disjoing family of sets in Σ, hence indeed

$$\mathbb{P}_B\left(\bigcup_{i=1}^{\infty} A_i\right) = \frac{1}{\mathbb{P}(B)}\mathbb{P}\left(\bigcup_{i=1}^{\infty}(A_i \cap B)\right) = \frac{1}{\mathbb{P}(B)}\sum_{i=1}^{\infty}\mathbb{P}(A_i \cap B) = \sum_{i=1}^{\infty}\mathbb{P}_B(A_i).$$

Thus $\mathbb{P}_B$ is a well-defined probability measure on Ω with respect to Σ. □

Definition 2.8. *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability triplet and I an index set. The set of evenst $\{A_i\}_{i=1}^{n}$ with $A_i \in \Sigma$ for each i is said to be (mutually) independent if for every $I \subset \{1, 2, \ldots, n\}$*

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \prod_{i \in I}\mathbb{P}(A_i)$$

Definition 2.9. *Let $(\Omega, \Sigma, \mathbb{P})$ be some probability triplet and (E, Λ) some state space. A random variable (r.v.) X is a measurable function $X : \Omega \to E$ (with repsect to Σ, Λ).*

This definition of a random variable is sufficient to calculate $\mathbb{P}(\{\omega : X(\omega) \in B\})$ for all $B \in \Lambda$, because these sets are elements of Σ, and $\mathbb{P}$ is only defined on Σ.

For each countable state-space, there exists some probability measure defined on the whole σ-algebra 2^Ω which is "characterized by its values on the axioms: $p_\omega = \mathbb{P}(\{\omega\}), \omega \in \Omega$" (Jacob & Protter, 2004, p. 22).

In the following, all random variables map into $\mathbb{R}$ and the σ-algebra on $\mathbb{R}$ is taken to be the Borel set $\mathcal{B}$, which is the σ-algebra generated by the open subsets of $\mathbb{R}$.

With the notion of probability triplets and random variables at hand, I can start defining the expected value. For this, one needs to understand the notion of a simple random variable. $X : \Omega \to \mathbb{R}$ is said to be simple[2] if $X = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$ with $a_i \in \mathbb{R}$ and $A_i \in \Sigma$ such that $\{A_i\}_{i=1}^{n}$ is a partition of Ω.

Definition 2.10. *The expected value of a simple random variable $X : \Omega \to \mathbb{R}$ is defined by*

$$\mathbb{E}[X] = \sum_{i=1}^{n} a_i \mathbb{P}(A_i)$$

Definition 2.11. *Let $X : \Omega \to \mathbb{R}$ be a positive random variable (that is, $X(\omega) \geq 0$ for all $\omega \in \Omega$). Then the expected value of X is defined by*

$$\mathbb{E}[X] = \sup \{\mathbb{E}[Y] : Y \leq X \text{ and } Y \text{ a simple r.v.}\}$$

Notice that it is possible to have $\mathbb{E}[X] = \infty$. For general definition of expected values, one can write $X = X^+ - X^-$, where $X^+(\omega) = \max(0, X(\omega))$ and $X^-(\omega) = -\min(0, X(\omega))$. These are both positive random variables (Jacob & Protter, 2004, p. 52).

Definition 2.12. *A (general) random variable has finite expected value if $\mathbb{E}[X^+] < \infty$ and $\mathbb{E}[X^-] < \infty$ and one writes*

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

If either $\mathbb{E}[X^+] = \infty$ or $\mathbb{E}[X^-] = \infty$ (but not both), then $\mathbb{E}[X]$ is still defined (as ∞ or $-\infty$ respectively). If $\mathbb{E}[X^+] = \infty$ and $\mathbb{E}[X^-] = \infty$ then X admits no expected value.

Much general properties of this mapping from the space that contains all integrable random variables (on $(\Omega, \Sigma, \mathbb{P})$) is known, such as that it is linear, monotone, monotone convergence, Fatou's lemma, Dominated Convergence, etc. I leave these theorems out of this thesis, but they can be found in for instance (Jacob & Protter, 2004, pp. 52-53) and (Rosenthal, 2006, pp.46-49, 103, 104). At any occurence of the use of a specific theorem, I refer to it in this thesis.

[2] $\mathbb{1}_A : \Omega \to [0, 1]$ is the identy function on subsets $A \subset \Omega$, meaning that $\mathbb{1}_A(\omega) = 1$ if $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ if $\omega \in \Omega \setminus A$.

2.2.1 Conditional Expectation

For the notion of conditional expectation, one must understand the concept of the $\mathcal{L}^1$- and L^1-space. Given a probability triplet $(\Omega, \Sigma, \mathbb{P})$, $\mathcal{L}^1(\Omega, \Sigma, \mathbb{P})$ is the space with all (real) random variables with finite expected value of the absolute value of the variables. The space $L^1(\Omega, \Sigma, \mathbb{P})$ is the set containing all equivalence classes of real random variables where each class contains almost surely equal random variables[3] that are in $\mathcal{L}^1$. Because almost surely equal random variables have the same expected value, the space L^1 is used. By abuse of notation, one often writes "take $Y \in L^1(\Omega, \Sigma, \mathbb{P})$". What is meant is: take a random variable Y in the equivalence class defined by Y in L^1. (Jacob & Protter, 2004, p. 53)

Theorem 2.1. *Let $Y \in L^1(\Omega, \Sigma, \mathbb{P})$ and let $\mathcal{G}$ be a sub-σ-algebra of Σ. There exists a unique element $\mathbb{E}[Y|\mathcal{G}]$ of $L^1(\Omega, \mathcal{G}, \mathbb{P})$ such that*

$$\mathbb{E}[YX] = \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]X],$$

for all bounded $\mathcal{G}$-measurable X, which satisfies

- *If $Y \geq 0$, then $\mathbb{E}[Y|\mathcal{G}] \geq 0$*
- *the map $Y \mapsto \mathbb{E}[Y|\mathcal{G}]$ is linear.*

The theorem above defines conditional expectation (on sub-σ-algebra's) and comes right from (Jacob & Protter, 2004, see theorem 23.4, p.202), where a proof can also be found. Furthermore, this conditional expectation has most of the same properties of the expected value, under almost sure equality, such as dominated convergence, Fatou's lemma, monotone convergence, etc.

Definition 2.13. *Let $X : \Omega \to \mathbb{R}$ be a general random variable and $Y \in L^1(\Omega, \Sigma, \mathbb{P})$. The conditional expectation $\mathbb{E}[Y|X]$ is defined $\mathbb{E}[Y|\sigma(X)]$, where $\sigma(X)$ is the σ-algebra generated by X, which is defined[4] by*

$$\sigma(X) = \{A \subset \Omega : X^{-1}(B) = A \text{ for some } B \in \mathcal{B}\}$$

2.2.2 Martingales

For the proof of the theorem of Wormald (1995), a martingale is created. Fix some probability triplet $(\Omega, \Sigma, \mathbb{P})$. A martingale is defined on an increasing[5] sequence of sub-σ-algebras $(\mathcal{F}_n)_{n=0}^{\infty}$ of Σ. This is called a filtration.

[3] Two random variables X, Y are said to be almost equal if $\mathbb{P}(\{w : X(\omega) \neq Y(\omega)\}) = 0$

[4] That this is indeed a σ-algebra see theorem 8.1 by Jacob and Protter (2004). Moreover, it can be extended to random variables X that map in $\mathbb{R}^n$ with respect to the borel-set $\mathcal{B}^n$.

[5] Meaning that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, for each $n \geq 0$.

Definition 2.14. *Let T be some countable set, $(\Omega, \Sigma, \mathbb{P})$ a probability triplet and (S, Λ) some measurable state-space. The function*

$$X : T \times \Omega \to S$$

such that $X(t, \omega) : \Omega \to S$ is, for each $t \in T$, a random variable on the probability triplet is called a discrete-time random process. Often, it is written as $\{X_t\}_{t \in T}$.

If T is countable, it is possible to look at the stochastic process X as the function

$$X : \Omega \to S^\infty,$$

where $X_t : \Omega \to S$ is a random varialbe for each $t \in T$. If $(S^\infty, \Lambda^\infty)$ is some measurable state-space, then this itself defines a random variable.

Definition 2.15. *A sequence of random variables $\{X_n\}_{n=0}^\infty$ (or a discrte-time random process) is called a martingale if*

(i) $\mathbb{E}[|X_n|] < \infty$, *for all* $n \geq 0$.

(ii) X_n *is* $\mathcal{F}_n$*-measurable for each* $n \geq 0$ *and*

(iii) $\mathbb{E}[X_n | \mathcal{F}_m] = X_m$, *for all* $0 \leq m \leq n$

This definition follows the one by Jacob and Protter (2004, p. 211). Moreover, if the equality in (iii) in the definition above is changed to $\leq$, $\{X_n\}_{n=0}^\infty$ is called a super-martingale and if it is changed to $\geq$, $\{X_n\}_{n=0}^\infty$ is called a sub-martingale.

2.2.3 Specifications and Conditional Probability

Most of the following lemma's and claims are very simple corollaries of standard theorems or definitions and therefore left out of the main text in an attempt to minimize boredom of the reader.

Definition 2.16. *Let $\{Y_t\}_{t \geq 0}$ be a discrete time random process with $t \in \mathbb{N}_{\geq 0}$ defined on some probability triplet $(\Omega, \Sigma, \mathbb{P})$. The history of $\{Y_t\}_{t \geq 0}$ is defined as*

$$H_t = \sigma(X_s, s \leq t) = \sigma\left(\bigcup_{s=0}^{t} \{A \subset \Omega : X_s(B) \in \Sigma \text{ for some } B \in \mathcal{B}\}\right)$$

Notice that H_t is a filtration, because $H_t \subset H_{t+1}$.

Claim 2.1. *Take $n \in \mathbb{N}$ and let $\{Y_t\}_{t \geq 0}$ be some discrete time random process with $t \in \mathbb{N}_{\geq 0}$ and H_t the history of the process. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be* *a lipschitz continuous function and define h as the function* $h : \mathbb{R} \to \mathbb{R}$ *that maps* $x \mapsto f\left(\frac{t}{n}, \frac{x}{n}\right)$*, then*

$$\mathbb{E}[h(Y_t)|H_{t+j}] = h(Y_t),$$

for each $j \in \mathbb{N}_{\geq 0}$*. .*

Proof. Take $i \in \mathbb{N}_{\geq 0}$. Notice that since f is a continous map by assumption and $x \mapsto \frac{x}{n}$ is also continous in x and for each n, it must hold that $h(x)$ is continous.

Furthermore, Y_t is H_t-measurable and since $H_t \subset H_{t+1} \ldots \subset H_{t+i}$ it also holds that Y_t is H_{t+i}-measurable. Take $U \in \mathbb{R}$ open. Then $h^{-1}(U)$ is open by continuity of h and by measurability of Y_t, $Y_t^{-1}(h^{-1}(U))$ is H_{t+i}-measurable. In other words, $h(Y_t)$ is H_{t+i}-measurable and therefore,

$$\mathbb{E}[h(Y_t)|H_{t+i}] = h(Y_t)$$

by for instance theorem 25.3 in Probability Essentials (Jacob & Protter, 2004, p. 204). □

Notice that claim 2.1 is a 'specification' of a general composition of two measurable functions. This is shown in, for instance and although through an exercise, Principles of Real Analysis (Aliprantis & Burkinshaw, 1998). As a last step, the proof uses a characterization of conditional expectation which holds intuitively and luckely can be proven.

The following lemma I used is not written down this particular way most of the time, but it is - to some extend - just the generalised Markov theorem.

Lemma 2.2. *Let* Y *be a random variable,* $f : \mathbb{R} \to \mathbb{R}$ *be a non-negative, non-decreasing borel-measurable function and* $a \geq 0$*, then:*

$$\mathbb{P}(Y > a) \leq \frac{\mathbb{E}[f(Y)]}{f(a)}$$

Proof. Since f is non-decreasing, it is clear that $Y > a$ implies that $f(Y) \geq f(a)$. Thus, by monotonicity of $\mathbb{P}$, it must hold that[6] $\mathbb{P}(Y > a) \leq \mathbb{P}(f(Y) \geq f(a))$. Now, notice that $f \circ Y : \Omega \to \mathbb{R}$ is a non-negative random variable (as f is borel-measurable, see for one the proof of claim 2.1) and $f(a) \geq 0$.

The remaining part follows from the (simplest) Markov-inequality, which I will proof for the fun of it. Since $f(Y)$ is non-negative, one can see that $f(a)\mathbb{1}_{\{f(Y) \geq f(a)\}} \leq f(Y)$. Now, by monotonicity of the expected value, it holds that:

$$f(a)\mathbb{P}(f(Y) \geq f(a)) = \mathbb{E}[f(a)\mathbb{1}_{\{f(Y) \geq f(a)\}}] \leq \mathbb{E}[f(Y)]$$

which proves the lemma. □

[6] Although equality is not necesarry for this proof, it can be easily shown (for, of course, $\geq$ instead of $>$)

The following claim is fundamental for the proof of Azuma's Lemma, where the proof is left out. I add it here for convenience, since I, for one, do not believe it to be trivial enough to be left out completely.

Claim 2.2. *For each $a \geq 0$ and $|x| \leq 1$ (in $\mathbb{R}$), it holds that*

$$e^{ax} \leq \cosh(a) + x \sinh(a)$$

Proof. The map $y \mapsto e^{ay}$ is convex and because $|x| \leq 1$, $\frac{1+x}{2} \in [0,1]$ and thus $1 - \frac{1+x}{2} = \frac{1-x}{2}$, which gives us that

$$e^{ax} = \exp\left(a\frac{1+x}{2} - a\frac{1-x}{2}\right) \leq \frac{1+x}{2}\exp(a) + \frac{1-x}{2}\exp(-a)$$

where the last term equals $\cosh(a) + x\sinh(a)$ □

For the proof, mainly for lemma 3.2 and 3.4 it is needed to look at conditional probability on a sub-σ-algebra. I offer the following definition. To see that such a probability measure is well defined, I offer lemma 2.3.

Definition 2.17. *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability triplet and let $\mathcal{G} \subset \Sigma$ some sub-σ-algebra. Take $A \in \Sigma$. The conditional probability $\mathbb{P}(A|\mathcal{G})$ or $\mathbb{P}_\mathcal{G}(A)$ is defined by*

$$\mathbb{P}(A|\mathcal{G}) = \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$$

Lemma 2.3. *Let $(\Omega, \Sigma, \mathbb{P})$ be probability triplet and $\mathcal{G} \subset \Sigma$ a sub-σ-algebra. Then $(\Omega, \Sigma, \mathbb{P}_\mathcal{G})$ is, almost surely, a probability triplet with respect to $\mathbb{P}$, where $\mathbb{P}_\mathcal{G}(A) = \mathbb{P}(A|\mathcal{G})$ for each $A \in \Sigma$.*

The lemma above implies the lemma below and tells the intuition behind conditional probability on σ-algebra's - why one can define them that way.

Proof. What I will prove is the following: (i) $\mathbb{P}_\mathcal{G}(\Omega) = 1$ a.s., (ii) $0 \leq \mathbb{P}_\mathcal{G}(A) \leq 1$ a.s. for each $A \in \Sigma$ and (iii) for each collection $\{A_i, i \in \mathbb{N}\}$ of disjoint sets in Σ and $A = \bigcup_{i=1}^{\infty} A_i$ $\mathbb{P}_\mathcal{G}(A) = \sum_{i=1}^{\infty} \mathbb{P}_\mathcal{G}(A_i)$.

(i). Since $\Omega \in \mathcal{G}$ (by definition of σ-algebra's), the random variable $\mathbb{1}_\Omega$ is $\mathcal{G}$-measurable. This implies that $\mathbb{E}[\mathbb{1}_\Omega|\mathcal{G}] = \mathbb{1}_\Omega$ a.s. (Jacob & Protter, 2004, see for instance theorem 23.5, p. 204), so indeed $\mathbb{P}_\mathcal{G}(\Omega) = 1$ a.s.

(ii). Take $A \in \Sigma$. Since $0 \leq \mathbb{1}_A \leq 1$ surely, it is also known by Jacob and Protter (2004, theorem 23.5.a. p. 202) that $0 \leq \mathbb{E}[\mathbb{1}_A|\mathcal{G}] \leq 1$.

(iii). Let $\{A_i : i \geq 1\}$ be a collection of disjoint sets in Σ. It is known that

$$\mathbb{E}\left[\bigcup_{i=1}^{\infty} A_i \middle| \mathcal{G}\right] = \sum_{i=1}^{\infty} \mathbb{E}[A_i|\mathcal{G}],$$

a.s. by Lebesque's dominated convergence theorem for conditional expectation (Jacob & Protter, 2004, theorem 23.8.c, p. 205). □

Lemma 2.4. *Let* $(\Omega, \Sigma, \mathbb{P})$ *be a probability triplet and* $\mathcal{G} \subset \Sigma$ *a sub-*σ*-algebra, then for each* $A, B \in \Sigma$ *with* $\mathbb{P}_{\mathcal{G}}(B) > 0$,

$$\mathbb{P}_{\mathcal{G}}(A) = \mathbb{P}_{\mathcal{G}}(A|B)\mathbb{P}_{\mathcal{G}}(B) + \mathbb{P}_{\mathcal{G}}(A|B^c)\mathbb{P}_{\mathcal{G}}(B^c)$$

almost surely, where $\mathbb{P}_{\mathcal{G}}(A) = \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$.

Proof. This follows immidiately from the same theorem that proves this for a regular probability measure, as

$$\mathbb{P}_{\mathcal{G}}(A|B) = \frac{\mathbb{P}_{\mathcal{G}}(A \cap B)}{\mathbb{P}_{\mathcal{G}}(B)},$$

almost surely. □

Lemma 2.5. *Let* $(\Omega, \Sigma, \mathbb{P})$ *be a probability triplet and* $\mathcal{G} \subset \Sigma$ *a sub-*σ*-algebra. Then for each* $A \in \Sigma$,

$$\mathbb{P}(A) = \mathbb{E}[\mathbb{P}(A|\mathcal{G})].$$

Proof. This is theorem 23.3.c. in Probability Essentials (Jacob & Protter, 2004) when noticing that $\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A]$ and $\mathbb{P}(A|\mathcal{G}) = \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$. □

Chapter 3

Wormald's Theorem

The theorem of Wormald is stated below, but I proof it in a vast amount of different lemma's for readability purposes. First, I follow the proof by Wormald (1997, theorem 5.1, p. 35). Later, I also state the original theorem of Wormald (1995), as was published. The latter I proved first, but the version below I used most, therefore it is presented here.

The structure in the proofs of both theorems is more or less the same, but the details are so different that I found it impossible to present one proof implicating both. Therefore, a proof of the orignal theorem of Wormald (1995, theorem 1, 2 p. 1219) is added in the appendix (because I like it very much). I explain the difference after I state the original version (see theorem 3.3).

3.1 Preliminaries

Before the theorem can be stated, I must offer some general setting following Wormald (1997). In the following, random processes are discrete time processes. Take $(\Omega, \Sigma, \mathbb{P})$ a probability triplet, (S, Λ) some measurable space and let Q be a process $Q : \mathbb{N}_{\geq 0} \times \Omega \to S$, hence for each $i \in \mathbb{N}_{\geq 0}$, $Q(i, \cdot)$ takes values in S and $Q(i, \cdot) : \Omega \to S$ is a random variable with respect to (S, Λ) and is written as Q_i.

Let $\{S^{(n)}\}_{n\in\mathbb{N}}$ be a sequence of sets. Consider a sequence of random processes $\{Q^{(n)}\}_{n\in\mathbb{N}}$ on a sequence of probability triplets $(\Omega_n, \Sigma_n, \mathbb{P}_n)$. Take $n \in \mathbb{N}$ and $i \in \mathbb{N}_{\geq 0}$, then $Q_i^{(n)}$ is a random variable that maps Ω_n into $S^{(n)}$.

Take $a \in \mathbb{N}$ and let $1 \leq l \leq n$. Define $S^{(n)+}$ as the set of all $h_t = (q_0, q_1, \ldots, q_t)$ such that $q_i \in S^{(n)}$ for each i and for $t = 0, 1, 2, \ldots$ and function $y_t^{(l)} : S^{(n)+} \to \mathbb{R}$ and the random counterpart of $y_t^{(l)}(h_t)$ by $Y_t^{(l)}$ (hence the function that maps $(S^{(n)})^\infty \to \mathbb{R}$ via the histories of random processes).

Last, let $D \subset \mathbb{R}^{a+1}$ and define T_D the stopping time for the variables $Y^{(1)}, \ldots, Y^{(a)}$ as

$$T_D = \min\left\{ t \in \mathbb{N}_{\geq 0} : \left(\frac{t}{n}, \frac{Y_t^{(1)}}{n}, \ldots, \frac{Y_t^{(a)}}{n} \right) \notin D \right\},$$

and so that T_D is defined as ∞ when there is no such t.

3.2 Theorem statement

Theorem 3.1. *(As in (Wormald, 1997)) For $1 \leq l \leq a$, where a is fixed, let $y_t : S^{(n)+} \to \mathbb{R}$ and $f_l : \mathbb{R}^{a+1} \to \mathbb{R}$ such that for some constant C_0 and all l, $|y_l(h_t)| \leq C_0 n$ for all $h_t \in S^{(n)+}$ for all n. Let $Y_t^{(l)}$ denote the random counterpart of $y_l(h_t)$. Assume the following three conditions hold, where in (ii) and (iii) D is some bounded, connected open set containing the closure of*

$$\{(0, z_1, \ldots, z_a) : \mathbb{P}(Y_0^{(l)} = z_l n, 1 \leq l \leq a) \neq 0 \text{ for some } n\}$$

(i) (Boundedness hypothesis) For some functions $\beta = \beta(n) \geq 1$ and $\gamma = \gamma(n)$, the probability that

$$\max_{1 \leq l \leq a} |Y_{t+1}^{(l)} - Y_t^{(l)}| \leq \beta$$

conditional upon H_t is at least $1 - \gamma(n)$ for $t < T_D$.

(ii) (Trend hypothesis) For some function $\lambda_1 = \lambda_1(n) = o(1)$, for all $l \leq a$,

$$\left| \mathbb{E}[Y_{t+1}^{(l)} - Y_t^{(l)} | H_t] - f_l(\tfrac{t}{n}, \tfrac{Y_t^{(1)}}{n}, \ldots, \tfrac{Y_t^{(a)}}{n}) \right| \leq \lambda_1$$

for $t < T_D$

(iii) (Lipschitz hypothesis) Each function f_l is continuous and satisfies a Lipschitz condition, on

$$D \cap \{(t, z_1, \ldots, z_a) : t \geq 0\}$$

with the same Lipschitz constant for each l.

Then the following are true.

(a) For $(0, \hat{z}_1, \ldots, \hat{z}_a) \in D$ the system of differential equations

$$\frac{\mathrm{d}z_l}{\mathrm{d}x} = f_l(x, z_1, \ldots, z_a), l = 1, \ldots, a$$

has a unique solution in D for $z_l : \mathbb{R} \to \mathbb{R}$ passing through

$$z_l(0) = \hat{z}_l,$$

for each $1 \leq l \leq a$ and which extends to points arbitrarily close to the boundary of D.

(b) *Let $\lambda > \lambda_1 + C_0 n\gamma$ with $\lambda = o(1)$. For a sufficiently large constant C, with probability $1 - O\left(n\gamma + \frac{\beta}{\lambda}\exp\left(-\frac{n\lambda^3}{\beta^3}\right)\right)$,*

$$Y_t^{(l)} = nz_l(\tfrac{t}{n}) + O(\lambda n)$$

uniformly over all $0 \leq t \leq \sigma n$ and for each l, where z_l is the solution in (a) with $\hat{z}_l = \frac{1}{n}Y_0^{(l)}$, and $\sigma = \sigma(n)$ is the supremum of those x to which the solution can be extended before reaching within l^∞-dinstance $C\lambda$ of the boundary of D.

3.3 Proof of Wormald's theorem

I will prove the theorem by Wormald (1997) using different lemma's, finally adding up to the full theorem. I must add that in the following, all three main assumptions (i), (ii) and (iii) of central theorem hold, but I will explicitely mention when they are used. I will not prove (a) as it is a well-known result in Differential Equations (Wormald, 1995), see for instance Hurewiz (1958, p. 32, theorem 11 and 12) for a proof.

As a last remark before I start the proof, I need to say that at first, I assume $a = 1$ (and $l = 1$ for that matter). For simplicity write Y_t for $Y_t^{(l)}$, f for f_l and z for z_l.

3.3.1 Transformation to Martingale

Before I can start, the following claim - albeit it is almost trivial - is needed. Define for all that follows, for $\lambda > \lambda_1$ as in (b):

$$w(n) = \left\lceil \frac{n\lambda(n)}{\beta(n)} \right\rceil \tag{3.1}$$

Claim 3.1. *If $\frac{\beta}{\lambda} > n^{1/3}$, the probability in (b) is bounded above by one, and thus trivial.*

Proof. Suppose $\frac{\lambda}{\beta} \leq \frac{1}{n^{1/3}}$. Then

$$\exp\left(-\frac{n\lambda^3}{\beta^3}\right) \geq \exp\left(-n\frac{1}{n}\right) = \frac{1}{e}.$$

Thus, $\frac{\beta}{\lambda}\exp(-n\frac{\lambda^3}{n^3}) \leq n^{1/3}e^{-1} \to \infty$ as $n \to \infty$. Hence, indeed, the probability in (b) is unrestricted. □

From this point forward, I can thus safely assume that $\frac{\beta}{n} \leq n^{1/3}$. Moreover, assume that $\lambda(n) < 1$ for all n, which is of no harm because $\lambda(n) = o(1)$.

Claim 3.2. *Suppose $\gamma = 0$. There exists some $C > 0$ such that if $\left(\frac{t}{n}, \frac{Y_t}{n}\right) \in D$ is at least at $C\lambda(n)$ distance from the boundary of D, then so is $\left(\frac{t+k}{n}, \frac{Y_{t+k}}{n}\right)$ for each $0 \leq k < w(n)$*

Proof. Take $\left(\frac{t}{n}, \frac{Y_t}{n}\right) \in D$ such that it is of at least l^∞ distance $C\lambda$ from the boundary (C will be determind later). Notice that, for some $C' > 0$

$$\frac{t+w}{n} - \frac{t}{n} \leq \frac{t + \frac{n\lambda}{\beta} + 1}{n} - \frac{t}{n} \leq C'\lambda(n),$$

because $\beta \geq 1$ and $\frac{\beta}{\lambda} \leq n^{1/3}$, thus $\frac{1}{n} \leq \frac{1}{n^{2/3}} \leq \frac{\lambda}{\beta} \leq \lambda$.

For the change in Y_t notice that, by assumption that $\gamma = 0$, it is at most

$$\left|\frac{Y_{t+w}}{n} - \frac{Y_t}{n}\right| \leq \frac{\beta w}{n} \leq C''\lambda,$$

because $\beta w \leq \beta + n\lambda$ and $\frac{\beta}{\lambda} \leq n^{1/3}$, thus (by positiviy of λ_1 and thus λ), $\beta \leq \lambda n^{1/3}$. I can conclude the above. Let $\epsilon > 0$ and take $C = \max\{C' + \epsilon, C'' + \epsilon\}$.

By the above, and because D is connected and open, $\left(\frac{t+w}{n}, \frac{Y_{t+w}}{n}\right)$ cannot be outside of D. □

Lemma 3.1. *Assume that, almost surely,*

$$|Y_{t+k+1} - Y_{t+k}| \leq \beta(n)$$

for all k or equivalently that $\gamma = 0$ (I call this assumption (a1)). Take $t \geq 0$ and assume that $\left(\frac{t}{n}, \frac{Y_t}{n}\right) \in D$ is l^∞-distance at least $C\lambda$ from the boundary of D (I call this assumption (a2)). Then there exists a function $g(n) = O(\lambda)$ such that

$$M_k = Y_{t+k} - Y_t - kf\left(\frac{t}{n}, \frac{Y_t}{n}\right) - kg(n)$$

is a supermartingale with respect to the σ-algebra's created by $H_t, H_{t+1}, \ldots, H_{t+w}$.

Notice that the map $\omega \mapsto f\left(\frac{t}{n}, \frac{Y_t(\omega)}{n}\right)$ is H_t-measurable (see also claim 2.1), which will be used in the proof of this lemma.

Proof. First, I show that, for $0 \leq k < w$

$$\begin{aligned}\mathbb{E}[Y_{t+k+1} - Y_{t+k} | H_{t+k}] &= f\left(\frac{t+k}{n}, \frac{Y_{t+k}}{n}\right) + O(\lambda_1) \\ &= f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + O(\lambda). \qquad (3.2)\end{aligned}$$

The first equality follows directly from assumption (ii). For the second equality, notice that $\frac{k}{n} \leq \frac{\beta k}{n}$ (which trivially holds) as $\beta \geq 1$. Therefore,

$$\left|\frac{t+k}{n} - \frac{t}{n}\right| \leq \frac{k\beta}{n}.$$

Moreover, by the assumption (a1), one gets

$$|Y_{t+k} - Y_t| \leq k\beta$$

by induction on assumption (i) in k. Hence, also

$$\left|\frac{Y_{t+k}}{n} - \frac{Y_t}{n}\right| \leq \frac{k\beta}{n}$$

and thus, by assumption (iii) - that f is Lipschitz-continous on D[1],

$$\left|f\left(\frac{t+k}{n}, \frac{Y_{t+k}}{n}\right) - f\left(\frac{t}{n}, \frac{Y_t}{n}\right)\right| \leq L\frac{k\beta}{n}$$

As a last remark for part one, notice that $k < w$, hence

$$\frac{k\beta}{n} \leq \frac{w\beta}{n} \leq \lambda + \frac{\beta}{n} \leq \left(1 + \frac{1}{n^{1/3}}\right)\lambda = O(\lambda)$$

Thus indeed, equation (3.2) holds.

This finishes the first part. In the second part, I prove that M_k is indeed a supermartingale. Notice, at first, that the existence of a function $g(n) = O(\lambda)$ follows directly from part one of this proof, as $g(n)$ can be taken such that

$$g(n) = \max_{0 \leq k < w(n)} \left\{\left|\mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}] - f\left(\frac{t}{n}, \frac{Y_t}{n}\right)\right|\right\}$$

Secondly, it follows that

$$\begin{aligned}
\mathbb{E}[M_{k+1}|H_{t+k}] &= \mathbb{E}\left[Y_{t+k+1} - Y_t - (k+1)f\left(\frac{t}{n}, \frac{Y_t}{n}\right) - (k+1)g(n)\middle|H_{t+k}\right] \\
&= \mathbb{E}\left[Y_{t+k+1} - Y_{t+k}\middle|H_{t+k}\right] + Y_{t+k} - Y_t \\
&\quad - (k+1)f\left(\frac{t}{n}, \frac{Y_t}{n}\right) - (k+1)g(n) \\
&\leq Y_{t+k} - Y_t - kf\left(\frac{t}{n}, \frac{Y_t}{n}\right) - kg(n) \\
&= M_k,
\end{aligned}$$

hence M_k is a super-martingale. The second equality here holds because Y_t, Y_{t+k} and $f\left(\frac{t}{n}, \frac{Y_t}{n}\right)$ are H_{t+k}-measurable[2] and the inequality follows from part one of the proof and the choice of $g(n)$. This shows the lemma. □

[1] That these elements are indeed in D follows from claim 3.2

[2] This is made more specific by claim 2.1

The choice of $g(n)$ here is very helpfull in a later part of the proof. Of course, chosing

$$g'(n) = \max_{0 \le k < w(n)} \left\{ \mathbb{E}[Y_{t+k+1} - Y_{t+k} | H_{t+k}] - f\left(\frac{t}{n}, \frac{Y_t}{n}\right) \right\}$$

would suffice in the proof of this lemma. The particular $g(n)$ - as in the proof of the lemma - has two nice properties, which is the reason for chosing it here already: it is positive and it also lowerbounds $\mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}]$. This last property guarantees that one can also generate a submartingale that looks a lot like the supermartingale M_k, which happens in lemma 3.5.

3.3.2 Azuma's inequality

Next, I also need Azuma's lemma or something closely related to it, which offers upper-bounds to 'states' in martingales that do not change too rapidly over time. I will state it - and prove it - as below, following the work of Wormald (1995)

Lemma 3.2. *Let $M_0, M_1, \ldots$ be a supermartingale with respect to a sequence of sub-σ-algebra's $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots$, $M_0 = 0$ and $\mathcal{F}_0$ empty*[3] *and $|M_{i+1} - M_i| \le c$ always. Then for all $\alpha > 0$:*

$$\mathbb{P}(M_i \ge \alpha c) \le \exp\left(-\frac{\alpha^2}{2i}\right)$$

A proof of a similar lemma is given by Shamir and Spencer (1987, p. 121, Theorem 3) and, of course, Azuma (1967, p. 357). However, before I can start proving the lemma of Azuma, I must state another lemma, which will be usefull when proving that of Azuma. As a last remark, I closely follow the proof of Azuma, although I rephrase some parts to fit the main theorem better.

Lemma 3.3. *Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra and Y a random variable with $|Y| \le 1$ and $\mathbb{E}(Y|\mathcal{G}) \le 0$ (a.s.) Then, for each $a \ge 0$ it holds that*

$$\mathbb{E}[\exp(aY)|\mathcal{G}] \le \exp(\tfrac{a^2}{2}) \textit{ a.s.}$$

Proof. Notice that for each $a > 0$, by claim 2.2, we find that (on whole Ω), using $|Y| \le 1$,

$$e^{aY} \le \cosh(a) + Y \sinh(a).$$

Therefore,

$$\mathbb{E}[e^{aY}|\mathcal{G}] \le \cosh(a) + \mathbb{E}[Y|\mathcal{G}] \sinh(a) \text{ a.s.}$$

[3] That is, of course, $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Here, it is used that $\mathcal{F}_0$ is empty. It holds also if $\mathcal{F}_0$ is not empty, but a conditional space (on $\mathcal{F}_0$) is used instead.

by linearity of conditional expectation and by[4] Jacob and Protter (2004, see lemma 23.1, p. 201). Now since $\mathbb{E}[Y|\mathcal{G}] \leq 0$ a.s. and $\sinh(a) \geq 0$ for each $a \geq 0$, it follows (read a.s. at each (in)equality)

$$\mathbb{E}[e^{aY}|\mathcal{G}] \leq \cosh(a) = \sum_{n=0}^{\infty} \frac{a^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty} \frac{(a^2)^n}{2^n n!} = e^{\frac{a^2}{2}}$$

using that $(2n)! \geq 2^n n!$, which finishes the proof of this lemma. $\square$

With the lemma above known, I am able to prove Azuma's lemma. To do so, I show the conditions needed for lemma 3.3, then I start proving Azuma's lemma itself.

Proof of lemma 3.2. Let $\alpha > 0$. Write $M'_i = \frac{M_i}{c}$ for each $i \in \mathbb{N}_{\geq 0}$ and denote $Y_i = M'_i - M'_{i-1}$ for each $1 \leq i$. Clearly, $|Y_i| \leq 1$ and also

$$\begin{aligned}\mathbb{E}[Y_i|M'_{i-1}] &= \mathbb{E}\left[\frac{M_i}{c} - \frac{M_{i-1}}{c}\middle|M'_{i-1}\right] \\ &= \frac{\mathbb{E}[M_i|M'_{i-1}] - M_{i-1}}{c} \leq \frac{M_{i-1} - M_{i-1}}{c} = 0,\end{aligned}$$

by linearity of conditional expectation and the fact that $(M_i)_{i\geq 0}$ (and therefore also $(M'_i)_{i\geq 0}$) is a supermartingale and notice that

$$\begin{aligned}\mathbb{E}[\exp(\tfrac{\alpha}{n}M'_n)] &= \mathbb{E}\left[\exp(\tfrac{\alpha}{n}(Y_n + M'_{n-1})\right] \\ &= \mathbb{E}[\exp(\tfrac{\alpha}{n}Y_n)\exp(\tfrac{\alpha}{n}M'_{n-1})] \\ &= \mathbb{E}[\mathbb{E}[\exp(\tfrac{\alpha}{n}Y_n)|M'_{n-1}]\exp(\tfrac{\alpha}{n}M'_{n-1})]\end{aligned}$$

where the last equality holds by the definition of conditional expectation and by 'taking out what is known'[5] (Jacob & Protter, 2004, see theorem 23.7 p. 204). By lemma 3.3, taking $a = \frac{\alpha}{n}$, we find that $\mathbb{E}[\exp(\frac{\alpha}{n}Y_n)|M'_{n-1}] \leq \exp(\frac{\alpha^2}{2n^2})$, so it holds that

$$\mathbb{E}[\exp(\tfrac{\alpha}{n}M'_n)] \leq \exp\left(\tfrac{\alpha^2}{2n^2}\right)\mathbb{E}[\exp(\tfrac{\alpha}{n}M'_{n-1})].$$

Using the same arguments for $n-1, n-2, \ldots, 1$ and the notion that $M'_0 = \frac{M_0}{c} = 0$, we thus get that

$$\mathbb{E}[\exp(\tfrac{\alpha}{n}M'_n)] \leq \prod_{i=1}^{n} \exp\left(\frac{\alpha^2}{2n^2}\right)\mathbb{E}[\exp(\tfrac{\alpha}{n}M'_0)] = \exp\left(\frac{n\alpha^2}{2n^2}\right).$$

[4]The lemma by Jacob and Protter (2004) holds when splitting any random variable into $Y = Y^+ - Y^-$, because Y is clearly integrable by the fact that $|Y| \leq 1$.

[5]This action is legitimate as $|M'_{n-1}|$ is clearly bounded above by $n-1$. Therefore, it and $\exp(\frac{\alpha}{n}M'_{n-1})$ are integrable. A same, yet easier, reasoning holds for $\exp(\frac{\alpha}{n}Y_n)$.

Now notice that because $x \mapsto \exp(\frac{\alpha}{n}x)$ is positive and non-decreasing (since $n, \alpha \geq 0$), the Markov-bound[6] tells

$$\mathbb{P}(M'_n > \alpha) \leq \frac{\mathbb{E}[\exp(\frac{\alpha}{n}M'_n)]}{\exp(\frac{\alpha^2}{n})} \leq \frac{\exp(\frac{n\alpha^2}{2n^2})}{\exp(\frac{\alpha^2}{n})} = \exp\left(\frac{-\alpha^2}{2n}\right).$$

The only thing left to see to finish this proof is that the events $\{M'_n > \alpha\}$ and $\{M_n > \alpha c\}$ coincide. As $n \in \mathbb{N}$ and $\alpha > 0$ were arbitrary, the lemma is proven. □

3.3.3 Concentration of Y_{t+w} and Y_t

Now that Azuma's lemma is derived and I have shown that a variation on assumption (i'), (ii) and (iii) implicitly transform differences in Y_t into a martingale through a 'clever' trick, I can start showing concentration of Y_t. To do so, I state the following lemma - something that follows almost immediately from Azuma's lemma.

Within, conditional probability is used - which is not the case for Azuma's lemma. However, H_t needs to be 'known' for the creation of the martingale $M_0, M_1, \ldots, M_w$ as it is a martingale with respect to $H_t, H_{t+1}, \ldots, H_{t+w}$. From this point on, I look at the space conditioned on the history up to time t; the space conditioned on H_t.

Lemma 3.4. *The assumptions of lemma 3.1 hold. $M_0, M_1, \ldots, M_w$ as in the proof of lemma 3.1, then*

$$\begin{aligned}\mathbb{P}\left(Y_{t+w} - Y_t - wf\left(\frac{t}{n}, \frac{Y_t}{n}\right) \geq wg(n) + \kappa\beta\sqrt{2w\delta}\,\middle|\, H_t\right) \\ \leq \exp(-\delta)\end{aligned} \tag{3.3}$$

for each $\alpha > 0$ and some $\kappa > 0$.

Proof. This proof first shows that the supermartingale $M_0, M_1, \ldots, M_w$ suffies the assumptions in Azuma's lemma and then shows the actual result in the lemma. To do so, notice that

$$\begin{aligned}|M_{k+1} - M_k| &= \left|Y_{t+k+1} - Y_{t+k} - f\left(\frac{t}{n}, \frac{Y_t}{n}\right) - g(n)\right| \\ &\leq |Y_{t+k+1} - Y_{t+k}| + \left|f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + g(n)\right| \\ &\leq \beta + \left|f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + g(n)\right|\end{aligned}$$

for each $k \in \{0, 1, \ldots, w-1\}$, which follows from the definition of M_k and because of assumption (a1) in lemma 3.1. Moreover, because $g(n) = O(\lambda)$ it

[6] See for instance lemma 2.2

certainly is also $g(n) = O(1)$ and f is Lipschitz on D, continuous everywhere and D is bounded, hence $|f|$ is upperbounded by its maximum on[7] $D \cup \partial D$, thus also at most $f = O(1)$, and thus $|g(n) + f\left(\frac{t}{n}, \frac{Y_t}{n}\right)| = O(1)$.

Although the above holds, a more elegant[8] approach to the bound can be found using another property of f. Notice that, by choice of $g(n)$, it holds that there exists some $C' > 0$ such that

$$-C'\lambda \leq g(n) \leq C'\lambda,$$

and similarly, by assumption (ii),

$$-\lambda \leq \mathbb{E}[Y_{t+1} - Y_t | H_t] - f\left(\frac{t}{n}, \frac{Y_t}{n}\right) \leq \lambda$$

it thus follows that, by assumption (i) and (a1),

$$-\lambda - \beta \leq f\left(\frac{t}{n}, \frac{Y_t}{n}\right) \leq \lambda + \beta.$$

This gives, again, that

$$-(C'+1)\lambda - \beta \leq f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + g(n) \leq (C'+1)\lambda + \beta,$$

thus, indeed, there exists some $\kappa > 0$ such that $\beta + |f(\frac{t}{n}, \frac{Y_t}{n}) + g(n)| \leq \kappa\beta$ (because $\lambda = o(1)$ and $\beta \geq 1$). Hence

$$|M_{k+1} - M_k| \leq \kappa\beta,$$

or in words that the super-martingale differences are bounded above uniformly over $k = 0, 1, \ldots, w-1$. Last, one must see that

$$\mathbb{E}[M_0] = \mathbb{E}[Y_t - Y_t - 0 \cdot f(\tfrac{t}{n}, \tfrac{Y_t}{n}) - 0 \cdot g(n)] = 0.$$

This concludes part one of this proof, as the assumptions in lemma 3.2 are met.

Now take $c = \kappa\beta$ and $\alpha = \sqrt{2w\delta}$. From lemma 3.2 we get that

$$\mathbb{P}(M_w \geq \alpha c | H_t) \leq \exp\left(-\frac{\alpha^2}{2w}\right)$$

which is equivalent to

$$\mathbb{P}\left(M_w \geq \kappa\beta\sqrt{2w\delta}\middle| H_t\right) \leq \exp\left(\frac{-\delta 2w}{2w}\right) = \exp(-\delta).$$

To finish the proof, notice that $M_w = Y_{t+w} - Y_t - wf(\frac{t}{n}, \frac{Y_t}{n}) - wg(n)$, hence

$$\mathbb{P}\left(M_w \geq \alpha c\right) = \mathbb{P}\left(Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n}) \geq wg(n) + \alpha c | H_t\right).$$

□

[7] Here, ∂D is the boundary of D, conform convential notation.

[8] Either of these approaches work for what is needed. Dependent on the particular use of this theorem, both can be chosen to get the best possible value of κ.

There are a few things left to do to derive at a main result on concentration. First, I state another lemma which tells something about the concentration of $Y_{t+w} - Y_t - wf(\frac{t}{n}, \frac{Y_t}{n})$.

Lemma 3.5. *Suppose again all assumptions in lemma 3.1 hold. Then*

$$\mathbb{P}\left(|Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq wg(n) + \kappa\beta\sqrt{2wn\frac{\lambda^3}{\beta^3}}\Big|H_t\right) \leq 2\exp\left(-n\frac{\lambda^3}{\beta^3}\right)$$

Proof. The first part of this proof defines a submartingale $(K_k)_{k=0}^{w(n)}$. Here, I use the particular choice of $g(n)$. Define

$$K_k = Y_{t+k} - Y_t - kf(\tfrac{t}{n}, \tfrac{Y_t}{n}) + kg(n).$$

This is a submartingale. To see so, notice that

$$\begin{aligned}\mathbb{E}[K_{k+1}|H_{t+k}] &= \mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}] - (k+1)f(\tfrac{t}{n}, \tfrac{Y_t}{n}) + (k+1)g(n) \\ &\geq Y_{t+k} - Y_t - kf(\tfrac{t}{n}, \tfrac{Y_t}{n}) + kg(n),\end{aligned}$$

because $-g(n) \leq \mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}] - f(\frac{t}{n}, \frac{Y_t}{n}) \leq g(n)$ by definition of $g(n)$. Therefore $\mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}] \geq f(\frac{t}{n}, \frac{Y_t}{n}) - g(n)$. This finishes part one of the proof.

Now, it follows that $-K_k$ is a supermartingale. To see that the differences are bounded by the same bound as for the supermartingale $(M_n)_{n=0}^{w}$, notice that

$$\begin{aligned}|-K_{k+1} + K_k| = |K_{k+1} - K_k| &\leq |Y_{t+k+1} - Y_{t+k} - f(\tfrac{t}{n}, \tfrac{Y_t}{n}) + g(n)| \\ &\leq |Y_{t+k+1} - Y_{t+k}| + |f(\tfrac{t}{n}, \tfrac{Y_t}{n}) - g(n)| \\ &\leq \beta + |f(\tfrac{t}{n}, \tfrac{Y_t}{n}) - g(n)|.\end{aligned}$$

By the proof of lemma 3.4, this is bounded above by[9]

$$|f(\tfrac{t}{n}, \tfrac{Y_t}{n}) - g(n)| \leq (C' + 1)\lambda + \beta$$

hence

$$|-K_{k+1} + K_k| \leq \kappa\beta$$

By lemma 3.2 it follows, taking $\alpha = \sqrt{2\omega\delta}$ and $c = \kappa\beta$ that[10]

$$\mathbb{P}(K_w \leq -\alpha c|H_t) = \mathbb{P}(-K_w > \alpha c|H_t) \leq \mathbb{P}(-K_w \geq \alpha c|H_t) \leq \exp(-\delta).$$

This equivalent to

$$\mathbb{P}\left(Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n}) + wg(n) \leq -\alpha c\;|H_t\right) \leq \exp(-\delta).$$

[9]Here, it is possible to redefine the C_0 in lemma 3.4.
[10]See, for instance, lemma 3.4, or its proof.

And clearly, it holds that

$$\mathbb{P}(Y_{t+w} - Y_t + wf(\tfrac{t}{n}, \tfrac{Y_t}{n}) \leq -wg(n) - \alpha c|H_t) \leq \exp(-\delta)$$

Implying also

$$\mathbb{P}(|Y_{t+w} - Y_t + wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq wg(n) + \alpha c|H_t) \leq 2\exp(-\delta).$$

Now, the proof is almost finished. Take $\delta = n\frac{\lambda^3}{\beta^3}$ and it is done. $\square$

3.3.4 Concentration of $Y_t - nz(\frac{t}{n})$

The concentration of $Y_t - nz(\frac{t}{n})$ - that what the theorem is all about - I show in this section. I use a proof by induction, and to do so, I break this part down into three lemma's, of which I first state the 'biggest', but I prove this last. The reason lies in the definitions and assumptions within the biggest lemma. In all that follows, take

$$\delta = n\frac{\lambda^3}{\beta^3} \tag{3.4}$$

Lemma 3.6. *Define $k_i = iw$ for $i = 0, 1, \ldots, i_0$, with $i_0 = \lfloor \frac{\sigma n}{w} \rfloor$, where w as in equation 3.1 and σ as in result (b) of theorem 3.1. Then it holds that*

$$\mathbb{P}\left(|Y_{k_j} - z(\tfrac{k_j}{n})n| \geq B_j \text{ for some } j \leq i\right) = O(ie^{-\delta}). \tag{3.5}$$

with

$$B_j = Bw\left(\lambda + \frac{w}{n}\right)\left(\left(1 + \frac{Bw}{n}\right)^j - 1\right)\frac{n}{Bw},$$

for some $B > 0$.

I will prove this lemma by induction. To do so, let me introduce some definitions for readability.

$$A_1 = Y_{k_i} - z(\tfrac{k_i}{n})n$$
$$A_2 = Y_{k_{i+1}} - Y_{k_i}$$
$$A_3 = z(\tfrac{k_i}{n})n - z(\tfrac{k_{i+1}}{n})n$$

Now that I have introduced the framework for inductive proof, it is time to state two other lemma's that will help me prove the main lemma (3.6) of this section.

Lemma 3.7. *There exists some $B' > 0$ such that*

$$\left|A_2 - wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)\right| < B'w\lambda$$

with probability $1 - O(e^{-\delta})$, where w as in equation 3.1, λ as in result (b) of theorem 3.1 and δ as in 3.4.

Proof. Assume the induction hypothesis, equation 3.5. Note that

$$\mathbb{P}\left(\left|A_2 - wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)\right| \geq B'w\lambda \middle| H_{k_i}\right)$$
$$= \mathbb{P}\left(|Y_{k_{i+1}} - Y_{k_i} - wf(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n})| \geq B'w\lambda \middle| H_{k_i}\right).$$

Taking $t = k_i$, and noticing that $k_{i+1} - k_i = w$, one gets that this is equivalent to

$$\mathbb{P}\left(|Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq B'w\lambda \middle| H_t\right). \tag{3.6}$$

Notice that, in lemma 3.5, $wg(n) = O(w\lambda)$ by choice of $g(n)$ and see furthermore that

$$\kappa\beta\sqrt{2wn\frac{\lambda^3}{\beta^3}} \leq \kappa\beta\sqrt{2wn\left\lceil\frac{n\lambda}{\beta}\right\rceil\frac{\lambda^2}{\beta^2}} = \kappa\sqrt{2}w\lambda = O(w\lambda),$$

hence there exists some B' such that, by lemma 3.5 (because $t = k_i \leq i_0 w \leq \sigma n$) the probability in equation 3.6 is upperbounded by $2e^{-\delta} = O(e^{-\delta})$. Lemma 2.5 tells that, because $Y = O(e^{-\delta})$ implies that $\mathbb{E}[Y] = O(e^{-\delta})$, the above also holds in the whole probability space, not just conditioned on H_t. This shows that indeed,

$$\left|A_2 - wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)\right| < B'w\lambda$$

has probability $1 - O(e^{-\delta})$. □

Lemma 3.8. *Let z be a solution in (a), then there exists some $B'' > 0$ such that*

$$\left|A_3 + wz'\left(\tfrac{k_i}{n}\right)\right| \leq \frac{B''w^2}{n}$$

with w as in equation 3.1.

Proof. A general approach[11] would be to notice that, because z is continuously differentiable (on the desired domain), there exists a $x \in \left[\frac{k_i}{n}, \frac{k_{i+1}}{n}\right]$ such that

$$z'(x) = \frac{z\left(\frac{k_{i+1}}{n}\right) - z\left(\frac{k_i}{n}\right)}{\frac{w}{n}},$$

[11] Another approach, for different assumptions on f can be found in the notes on this proof. It could be that, when assuming f to be analytic, for instance, a better bound can be found.

because again $\frac{k_{i+1}}{n} - \frac{k_i}{n} = \frac{w}{n}$ by the Mean Value Theorem (Vuik, van Beek, Vermolen, & van Kan, 2007). Hence,

$$\begin{aligned}\left| z'\left(\frac{k_i}{n}\right) - \frac{z\left(\frac{k_{i+1}}{n}\right) - z\left(\frac{k_i}{n}\right)}{\frac{w}{n}} \right| &= \left| z'\left(\frac{k_i}{n}\right) - z'(x) \right| \\ &= \left| f\left(\frac{k_i}{n}, z\left(\frac{k_i}{n}\right)\right) - f(x, z(x)) \right| \\ &\leq \max\left\{ \left| \frac{k_i}{n} - x \right|, \left| z\left(\frac{k_i}{n}\right) - z(x) \right| \right\},\end{aligned}$$

because f is lipschitz with constant L. This is, however, in case of the l^∞ norm lipschitz assumption, for the l^1, just take the sum of the two. Next, notice that

$$\left| z(x) - z\left(\frac{k_i}{n}\right) \right| = \left| \int_{\frac{k_i}{n}}^{x} f(t, z(t)) \mathrm{d}t \right| \leq \left(x - \frac{k_i}{n} \right) \max_{t \in [0, \sigma(n)]} |f(t, z(t))|,$$

which is a value that f takes on the closed interval because it is continuous by assumption[12] and notice too that $x - \frac{k_i}{n} \leq \frac{w}{n}$ by choice of x. Hence, indeed,

$$\left| z'\left(\frac{k_i}{n}\right) - \frac{z\left(\frac{k_{i+1}}{n}\right) - z\left(\frac{k_i}{n}\right)}{\frac{w}{n}} \right| \leq L \max\left\{ w, w \max_{t \in [0, \sup_n \sigma(n)]} |f(t, z(t))| \right\} \leq B'' w,$$

for some $B'' > 0$. Last, notice that

$$\left| A_3 + wz'\left(\frac{k_i}{n}\right) \right| = \left| z'\left(\frac{k_i}{n}\right) - \frac{z\left(\frac{k_{i+1}}{n}\right) - z\left(\frac{k_i}{n}\right)}{\frac{w}{n}} \right|,$$

which finishes the proof of this lemma. □

Now, it is time to prove lemma 3.6, as all the building blocks are present to do so.

Proof of lemma 3.6. First, notice that $z(0) = \frac{Y_0}{n}$, hence the induction hypothesis (equation 3.5) holds for $n = 0$. Second, it is helpful to see that

$$\left| A_3 + wf\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right) \right| = \left| A_3 + wz'\left(\tfrac{k_i}{n}\right) - wz'\left(\tfrac{k_i}{n}\right) + wf\left(\tfrac{k_i}{n}\right) \right|.$$

From lemma 3.8 and the fact that z is a solution to (a) (i.e. that $z'(x) = f(x, z(x))$ on D), it is known that[13]

$$\left| A_3 + wf\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right) \right| \leq \frac{B'' w^2}{n} + \left| wf\left(\frac{k_i}{n}, z\left(\tfrac{k_i}{n}\right)\right) - wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right) \right|,$$

[12] Thus so is $z(x)$. However, this bound seems to be quite abusive. Moreover, the actual bound is given by some $A \geq \sigma(n)$ that is constant in n.

[13] Here, I use that $\frac{k_i}{n} \leq \sigma$ by choice of i_0.

for some (constant) $C \geq 0$. Now it is time to use - once more - the fact that f is Lipschitz continuous. This implies that

$$\left|f\left(\tfrac{k_i}{n}, z(\tfrac{k_i}{n})\right) - f\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right)\right| \leq L\left|z(\tfrac{k_i}{n}) - \frac{Y_{k_i}}{n}\right|,$$

for the lipschitz-constant L. Furthermore, the induction hypothesis (equation 3.5) offers that $|Y_{k_i} - z(\frac{k_i}{n})n| < B_i$ with probability $1 - O(ie^{-\delta})$. This shows that the following inequality holds,

$$\left|f\left(\tfrac{k_i}{n}, z(\tfrac{k_i}{n})\right) - f\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right)\right| \leq L\frac{B_i}{n},$$

with probability $1 - O(ie^{-\delta})$. Hence I can deduce that

$$\left|A_3 + wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)\right| \leq \frac{B''w^2 + B''wB_i}{n},$$

with probability $1 - O(ie^{\delta})$ because $L \leq B''$ for sure by the proof in lemma 3.8

Now, let me move to the final part. To do so, notice that

$$\begin{aligned}|A_1 + A_2 + A_3| &= \left|Y_{k_i} - z(\tfrac{k_i}{n})n + Y_{k_{i+1}} - Y_{k_i} + z(\tfrac{k_i}{n})n - z(\tfrac{k_{i+1}}{n})n\right| \\ &= \left|Y_{k_{i+1}} - z(\tfrac{k_{i+1}}{n})\right|.\end{aligned}$$

Without further ado, lemma 3.7 and 3.8 offer the following upper bound[14] when taking $B = \max\{B', B''\}$:

$$\begin{aligned}|A_1 + A_2 + A_3| &\leq |A_1| + \left|A_2 - wf\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right)\right| + \left|A_3 + wf\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right)\right| \\ &\leq B_i + Bw\lambda + \frac{Bw^2 + BwB_i}{n}, \\ &= Bw\lambda + \frac{Bw^2}{n} + \frac{(Bw+n)B_i}{n} \\ &= B_{i+1}\end{aligned}$$

with probability at most $(1 - O(ie^{-\delta})(1 - O(e^{-\delta})) = 1 - O((i+1)e^{-\delta})$. This finishes the proof, as it shows the induction hypothesis, equation 3.5. □

As this lemma is proven, I am ready to show that - given assumption (a2) - the theorem of Wormald holds. This is exactly what the lemma below tells.

Lemma 3.9. *Given assumption (a1), (a2) and $a = 1$, (b) in theorem 3.3 holds.*

[14] For a more heavy 'calculation', showing the last equality, see claim 3.3

Proof. Take $0 \leq t \leq \sigma n$ and let $i = \lfloor \frac{t}{w} \rfloor$ (then $k_i \leq t \leq k_{i+1}$) and it is known that

$$\mathbb{P}\left(\left|Y_{k_i} - z\left(\frac{k_i}{n}\right)\right| \leq B_i\right) = 1 - O(ie^{-\delta}),$$

by lemma 3.6. Furthermore, it is usefull to see that

$$B_i = \left(\lambda + \frac{w}{n}\right)\left(\left(1 + \frac{Bw}{n}\right)^i - 1\right)n = O(n\lambda + w)$$

Since $i \leq \lfloor \frac{\sigma(n)n}{w} \rfloor$, clearly $i = O(\frac{n}{w})$ since D is bounded (and thus also bounded in its first element).

Now, notice that $t - k_i \leq w$, hence the change in Y is upperbouned by βw. The change in z is upperbounded by $O(w)$ because

$$n\left|z\left(\frac{t}{n}\right) - z\left(\frac{k_i}{n}\right)\right| \leq n\int_{\frac{k_i}{n}}^{\frac{t}{n}} |f(x, z(x))|\mathrm{d}x \leq n\frac{w}{n} \max_{x \in [0, \sup_n \sigma(n)]} |f(x_1, z(x_1))|,$$

where $x = (x_1, \ldots, x_n)$ because z, f are continuous by assumption, by result (a) and because D is bounded.

See, too, that $w\beta = O(n\lambda)$ and $w = O(n\lambda)$ because β is bounded below, by definition of w and because $n\lambda \to \infty$ as $n \to \infty$ since $\frac{\lambda}{\beta} \geq \frac{1}{n^{1/3}}$ and thus $\lambda \geq \frac{\beta}{n^{1/3}} \geq \frac{1}{n^{1/3}}$.

I can conclude from lemma 3.6 that, indeed, with probability $1 - O(\frac{n}{w}e^{-\delta})$,

$$\left|Y_t - z\left(\frac{t}{n}\right)n\right| = O(\lambda n).$$

The last thing to notice is that $\frac{n}{w} \leq \frac{\beta}{\lambda}$ by definition of w, which finishes the proof of this lemma. □

3.3.5 Final generalisations

The final generalisation - that the scaled points $\frac{k_i}{n}$ and $\frac{Y_{k_i}}{n}$ indeed are l^∞ distance at least $C\lambda$ away from the boundary - makes the proof complete for $\gamma = 0$. The last step is to go from $\gamma = 0$ to arbitrary γ. This shows that the full theorem of Wormald (1997) holds, for $a = 1$, which is generalised thereafter. To see this, I present the following lemma.

Lemma 3.10. *All the points* $\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)$ *are at least* l^∞ *distance* $C'\lambda$ *away from the boundary of* D*, where* $k_i = iw$*,* $i = 0, 1, \ldots, i_0$*,* $i_0 = \lfloor \frac{\sigma n}{w} \rfloor$*, for some* $C' > 0$ *large enough with probability* $1 - O(\frac{n}{w}e^{-\delta})$*.*

Proof. The induction hypothesis in lemma 3.5 is the main part of this generalizations.

By choice of k_i, that is $k_i \leq \sigma n$, $\frac{k_i}{n} \leq \sigma$, hence $z\left(\frac{k_i}{n}\right)$ is at least l^∞ distance $C\lambda$ away from the boundary by definition of σ (and for each i, the distance between $\frac{k_i}{n}$ and the boundary of D is already upperbounded by σ hence trivial). Let $x = (x_1, x_2) \in \partial D$, then

$$\left| z\left(\frac{k_i}{n}\right) - x_2 \right| \geq C\lambda$$

And the probability that

$$\left| \frac{Y_{k_i}}{n} - z\left(\frac{k_i}{n}\right) \right| \geq \frac{B_i}{n} \tag{3.7}$$

for any $i \leq i_0$ is $O(i_0 e^{-\delta})$. Hence

$$\left| \frac{Y_{k_i}}{n} - x_2 \right| \geq \left| C\lambda - \frac{B_i}{n} \right| \geq (C - C_0)\lambda,$$

for some C_0 for which $B_i \leq C_0 \lambda n$ for each n and each $0 \leq i \leq i_0$, with probability $1 - O(\frac{n}{w} e^{-\delta})$. Hence, chosing C so that it is at least bigger than C_0, offers a suitable lowerbound. Define $C' = C - C_0$. Here C is defined, or at least lower-bounded.

Last, it must be noted that there exists some n such that $\left(0, \frac{Y_0}{n}\right) \in D$ and because D is open and containing the closure of this point[15], there must exist some n such that this point is far enoug away from the boundary, since $\lambda = o(1)$. □

Notice that this does not change the result, because the events for which equation 3.7 holds is equivalent to the events in lemma 3.6 and thus does not change the probability there (because it could be included in the induction hypothesis).

Lemma 3.11. *Let γ an arbitrary function in n, then (b) still follows.*

I will offer a proof of this statement in line with the theorem as proven by (Wormald, 1997, p. 38) here, which uses the assumption that $|Y(t)| \leq C_0 n$ for some $C_0 > 0$ and all t. There is another approach, which can be found in the notes on this proof and which I use to proof Wormald's original theorem (Wormald, 1995). The other approach does not assume $|Y(t)| \leq C_0 n$, but restricts γ to some extend.

Proof. There are two cases, the elements of Ω_n for which $|Y_{t+1} - Y_t| \leq \beta$ and those that are not. Condition all steps for the bounds of the martingale (lemma 3.5 and 3.4) on the event that the inequality $|Y_{t+1} - Y_t| \leq \beta$ holds.

This changes the difference between the expected change and $f\left(\frac{t}{n}, \frac{Y_t}{n}\right)$ as in lemma 3.1 by at most $C_0 n \gamma$, hence

$$\mathbb{E}[Y_{t+k+1} - Y_{t+k} | H_{t+k}] = f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + O\left(\lambda_1 + C_0 n\gamma + \frac{k\beta}{n}\right).$$

[15] This holds for all $1 \leq l \leq a$

Take for λ_1 in the rest of the proof $\lambda_1 + C_0 n\gamma$ and the result follows, with probability $1 - O(n\gamma)$, because the probability that any of the martingale-differences is not bounded above by $\kappa\beta$ is at most $\sigma n\gamma$ (because $0 \le t \le \sigma n$). Substracting this, indeed

$$Y_t = nz(\tfrac{t}{n}) + O(\lambda n)$$

with probability $1 - O\left(n\gamma + \frac{\beta}{\lambda}\exp\left(-n\frac{\lambda^3}{\beta^3}\right)\right)$, where $\lambda > \lambda_1 + C_0 n\gamma$ which is exactly result (b) for $a = 1$. $\square$

The last thing that is left to be shown is that the proof works exactly the same for $a \neq 1$, hence for arbitrary $a \in \mathbb{N}$.

Lemma 3.12. *For $a > 1$, the result (b) still follows.*

Proof. To see this, notice that the probability of

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i),$$

by definition of a probability measure and thus the probability that any of the events $\{|Y_l(k_j) - z(k_j/n)n| > B_j\}$ occur (in l) is upperbounded by the sum of the probability of either event occuring, hence by $O(aie^{-\alpha})$ in lemma 3.6. Hence, change the induction hypothesis to

$$\mathbb{P}\left(\left|Y_{k_j}^{(t)} - z\left(\frac{k_j}{n}\right)n\right| > B_j \text{ for some } j \le i\right) = O(aie^{-\delta}),$$

for all $1 \le l \le a$. Checking for $i+1$, one just has to check for each variable, hence adding the probabilities that it does not fail. This shows the induction and thus the proof, because a is constant by choice.[16] Last, it changes the probability that $\max_{1 \le l \le a} |Y_l(t+1) - Y_l(t)| \le \beta$ does not hold by at most a, hence this probability becomes $an\gamma = O(n\gamma)$ when a is constant[17]. So indeed result (b) follows. $\square$

3.3.6 Notes on proof

The following claim is a tedious. I left it out of the proof, because it just makes everything less readable.

Claim 3.3. *For B_i, B defined as in the proof of lemma 3.6, it holds that*

$$B_{i+1} = Bw\lambda + \frac{Bw^2}{n} + \frac{(Bw+n)B_i}{n}$$

[16] Chosing $a = a(n)$, lemma 4.1 also follows

[17] $a = a(n)$ implies lemma 4.1

for all $i+1 \leq i_0$, where

$$w = \left\lceil \frac{n\lambda}{\beta} \right\rceil,$$

(which is equivalent to the definition of w in equation 3.1).

Proof. By definition of B_i, it holds that

$$\begin{aligned}
& Bw\lambda + \frac{Bw^2}{n} + \frac{(Bw+n)B_i}{n} \\
&= \left(Bw\lambda + \frac{Bw^2}{n}\right) + \frac{(Bw+n)}{n}\left(Bw\lambda + \frac{Bw^2}{\lambda}\right)\left(\left(1+\tfrac{Bw}{n}\right)^i - 1\right)\frac{n}{Bw} \\
&= \left(Bw\lambda + \frac{Bw^2}{n}\right)\left(1 + \left(1+\tfrac{Bw}{n}\right)^i - 1 + \frac{n}{Bw}\left(\left(1+\frac{Bw}{n}\right)^i - 1\right)\right) \\
&= \left(Bw\lambda + \frac{Bw^2}{n}\right)\left(\left(1+\frac{Bw}{n}\right)^i\left(1+\frac{n}{Bw}\right) - \frac{n}{Bw}\right) \\
&= \left(Bw\lambda + \frac{Bw^2}{n}\right)\left(\left(1+\frac{Bw}{n}\right)^{i+1} - 1\right)\frac{n}{Bw}
\end{aligned}$$

which equals B_{i+1}. □

Claim 3.4. *Let f be analytic on at least $D \cup \partial D$, then*

$$\left|A_3 + wz'\left(\tfrac{k_i}{n}\right)\right| = O\left(\tfrac{w^2}{n}\right)$$

Proof. By the most simple formula to calculate the first derivative (see for instance Vuik et al. (2007, see theorem 3.2.1, p. 26)):

$$\left| z'\left(\tfrac{k_i}{n}\right) - \frac{z\left(\frac{k_{i+1}}{n}\right) - z\left(\frac{k_i}{n}\right)}{\frac{w}{n}} \right| = O\left(\frac{w}{n}\right) \leq \frac{Cw}{n},$$

because $\frac{k_{i+1}-k_i}{n} = \frac{w}{n}$, for some constant $C > 0$. This implies that

$$|A_3 + wz'\left(\tfrac{k_i}{n}\right)| \leq \frac{Cw^2}{n} = O\left(\frac{w^2}{n}\right).$$

This finishes the proof of this lemma. The upperbound is thus given by the maximum of $z''(t)$ for $0 \leq t \leq \sigma n$, which exists and is bounded by assumption on f. □

Another approach to the theorem

Many of the first lemma's assume that $\gamma = 0$. However, this need not be, as is shown in lemma 3.11. However, there is another approach, for which just some other restriction on γ is needed.

Theorem 3.2. *Updating assumption (i) by*

$$\mathbb{P}\left(|Y_{t+1} - Y_t| > \beta(n)\middle|H_t\right) = \gamma,$$

for each $t < T_D$, $\gamma = O\left(\frac{\exp(-\delta)}{n}\right)$, *and* $\delta(n) = n\frac{\lambda^3}{\beta^3}$ *(which is equivalent to the definition of* δ *in equation 3.4). It holds that for some* $t \geq 0$ *for which assumption (a2) holds,*

$$\mathbb{P}(|Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq wg(n) + \kappa\beta\sqrt{2w\delta}|H_t) \leq 2\exp(-\delta)$$

and thus also result (a) and (b), where (b) can be updated such that $\lambda > \lambda_1$ *is chosen, or where* $\gamma = 0$ *in result (b).*

Proof. First, let me define a few sets for readability. For $0 \leq k < w$ and particular t as assumed in the lemma,

$$B_k = \{\omega : |Y_{t+k+1} - Y_{t+k}| \leq \beta\}$$

and $B = \bigcap_{k=0}^{w-1} B_k$. Last, write

$$A = \left\{\omega : |Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq wg(n) + \kappa\beta\sqrt{2w\delta}\right\}$$

Look at B^c, the complement of B, and notice that certainly $\mathbb{P}(B^c|H_t) \leq O(\exp(-\delta))$, because

$$\mathbb{P}(B^c|H_t) \leq \sum_{k=0}^{w-1} \mathbb{P}(B_n^c|H_t)$$

and $\mathbb{P}(B_k^c|H_t) = O\left(\frac{\exp(-\delta)}{n}\right)$ by definition. Moreover, $w \leq n+1$, hence one gets that indeed, $\mathbb{P}(B^c|H_t) = O(\exp(-\delta))$. This also shows that $\mathbb{P}(B|H_t) = 1 - O(\exp(-\delta))$.

Now, write $P_{H_t}(\cdot) = \mathbb{P}(\cdot|H_t)$ for simplicity (and notice that this, on itself, is a probability measure)[18]. By a simple property of probability measures (for instance proven by Jacob and Protter (2004, see theorem 3.4, p. 17)), it holds that[19]

$$\mathbb{P}_{H_t}(A) = \mathbb{P}_{H_t}(A|B)\mathbb{P}_{H_t}(B) + \mathbb{P}_{H_t}(A|B^c)\mathbb{P}_{H_t}(B^c)$$

and since $\mathbb{P}_{H_t}(A|B^c) \leq 1$ (by the most trivial upperbound for the probability measure) and $\mathbb{P}_{H_t}(A|B) \leq 2\exp(-\delta)$ by lemma 3.5, we have that

$$\mathbb{P}_{H_t}(A) = O(\exp(-\delta))(1 - O(\exp(-\delta)) + O(\exp(-\delta)).$$

This shows exactly that $\mathbb{P}(A|H_t) = O(\exp(-\delta))$. □

[18] See lemma 2.3 for a justification.

[19] Here, it is used implicitely that $\mathbb{P}_{H_t}(B) \neq 0$ a.s. (this is trivial, because $\gamma < 1$ can be assumed) and $\mathbb{P}_{H_t}(B^c) \neq 0$ a.s. also holds, because otherwise this whole lemma can be neglected.

Hereby, I can end the section on generalizing assumption (a1), that $\gamma = 0$. This leaves the option to dive into the next part of the proof, where the concentration of $Y_{t+w} - Y_t$ is converted to the concentration of $Y_t - nz(\frac{t}{n})$. It leaves the assumption that $|Y(t)| < C_0 n$ untouched, while the result still follows.

For the original theorem of Wormald (1995), this result follows always (even for (i')), hence that theorem can always be updated.

3.4 Original version of Wormald's Theorem

As I started out proving the theorem of Wormald, I did so by it's old theorem as displayed below. This is the original theorem by Wormald Wormald (1995, Theorem 1 and 2, p. 1219). All further definitions remain the same.

Theorem 3.3 (Wormald). *Let a be fixed. For $1 \leq l \leq a$, define $y^{(l)} : S_n^+ \to \mathbb{R}$ and $f_l : \mathbb{R}^{a+1} \to \mathbb{R}$, such that for some constant C and all l, $|y_l(h_t)| < Cn$ for all $h_t \in S_n^+$ and for all n. Suppose also that for some function $m = m(n)$:*

(i) there is a constant C' such that for all $t < m$ and all l,

$$|Y_t^{(l)} - Y_t^{(l)}| < C'$$

always.

(ii) for all l and uniformly over $t < m$,

$$\mathbb{E}[Y_{t+1}^{(l)} - Y_t^{(l)} | H_t] = f_l\left(\frac{t}{n}, \frac{Y_t^{(1)}}{n}, \ldots, \frac{Y_t^{(a)}}{n}\right) + o(1)$$

always.

(iii) for each l, the function f_l is continous and satisfies a Lipschitz condition on D, where D is some bounded, connected, open set containing the intersection of $\{(t, z^{(1)}, \ldots, z^{(a)}) : t \geq 0\}$ with

$$\{(0, z^{(1)}, \ldots, z^{(a)}) : \mathbb{P}(Y_0^l = z^{(l)} n, 1 \leq l \leq a) \neq 0 \text{ for some } n\}$$

Then

(a) For $(0, \hat{z}^{(1)}, \ldots, \hat{z}^{(l)}) \in D$ the system of differential equations

$$\frac{\mathrm{d}z_l}{\mathrm{d}x} = f_l(x, z_1, \ldots, z_a), \qquad l = 1, \ldots, a$$

has a unique solution in D for $z_l : \mathbb{R} \to \mathbb{R}$ with

$$z_l(0) = \hat{z}^{(l)}, \qquad l = 1, \ldots, a,$$

and which extends to points arbitrarily close to the boundary of D

(b) *Almost surely,*

$$Y_t^{(l)} = nz_l(\tfrac{t}{n}) + o(n)$$

uniformly over $0 \le t \le \min\{\sigma n, m\}$ *and for each* l*, where* $z_l(t)$ *is a solution in (a) with* $\hat{z}_l = \frac{Y_0^{(l)}}{n}$ *and* $\sigma = \sigma(n)$ *is the supremum of those* s *to which the solution can be extended.*

Moreover, the assumption in (i') can be weakened to what is stated below. The above is written down for simplicity, but what follows is what I prove.

Theorem 3.4 (Weakening of (i)). *Theorem 3.3 also holds if condition (i) is weakened to:*

(i') *for some functions* $w = w(n)$ *and* $\lambda = \lambda(n)$ *with* $\lambda^4 \log(n) < w < \frac{n^{2/3}}{\lambda}$ *and* $\lambda \to \infty$ *as* $n \to \infty$*, for all* l *uniformly and for all* $t < m$*:*

$$\mathbb{P}\left(|Y_{t+1}^{(l)} - Y_t^{(l)}| > \frac{\sqrt{w}}{\lambda^2 \sqrt{\log(n)}} \middle| H_t \right) = o(n^{-3}),$$

always on Ω_n.

The exact proof can be found in Appendix A, which follows the same structure as the proof of theorem 3.1 to a large extend, but the details are different. Because I started out proving this theorem, and not theorem 3.1, the text is more or less the same. This is the reason it is in the appendix. Moreover, as noted in theorem 3.2, the assumption of $|Y(t)| \le C_0 n$ can be left out in this theorem, due to my way of proving it.

3.4.1 Difference between the two versions of Wormald's theorem

The difference between this original version (Wormald, 1995) and the general version (Wormald, 1997) are found in the first assumption (assumption (i) or (i')). In the original version (1995), the difference in the processes $Y_t^{(l)}$ are restricted for each l to a specific bound, with fixed asymptical probability. In the generalised version, the maximal difference in the processes $Y_t^{(l)}$ is upperbounded by something that is at least 1 ($\beta(n)$) with probability $\gamma(n)$, where β and γ have very little restrictions.

This echoes in the result of the second part of the theorem, that tells how close $Y_t^{(l)}$ lies to $nz_l(\frac{t}{n})$ and with what probability. In the original version (1995), the probability of that $Y_t^{(l)}$ equals $nz_l(\frac{t}{n}) + o(n)$ is given by $1 - o(1)$. In the generalised version, this becomes $Y_t^{(l)}$ equals $nz_l(\frac{t}{n}) + O(\lambda n)$ with a probability that depends on the particular choice of λ and on γ and β.

Chapter 4

A Balls and Bins Model of Trade

Balls and bins can be seen as the "classical methafor for the multinomial distribution" (Corrado, 2011, p. 349). In the most simple model, I look at n bins of equal size, which are empty at $t = 0$. At each (discrete) step in time of length one, a ball is thrown into one of the bins - either with equal probability or in the general case proportional to the size of the bin.

This model can, for instance, describe trade flow from one country to all others. In the model, there are n country-trade pairs, which are represented by the bins. Each ball represents a unit of trade. When more balls are thrown, this models the event that there is more trade.

The natural questions that arize are the margins within this model.

- How many bins are empty after m balls are thrown?
- What is the maximum amount of balls in a bin?
- How many bins are there of any particular amount of balls?

All questions above are random variables defined on a histories. The sections below show that each of them satisfies the assumptions in theorem 3.3 or theorem 3.1.

It also raises the question of how the balls and bins model can be described mathematically. To do so, I present two different approaches. The first looks at the number of bins with a certain amount of balls. The second looks at the number of balls in a particular bin.

For the everything that follows, another lemma - an extension of Wormald's theorem - is needed.

Lemma 4.1. *(as the note on theorem 5.1 by Wormald (1997)) Let all assumptions and definitions in theorem 3.1 hold, but change $a \in \mathbb{N}$ to a is a function in n, then result (a) remains the same and result (b) must be updated to:*

Let $\lambda > \lambda_1 + C_0 n\gamma$ with $\lambda = o(1)$. For a sufficiently large C, with probability

$$1 - O\left(an\gamma + \frac{a\beta}{\lambda}\exp\left(-n\frac{\lambda^3}{\beta^3}\right)\right),$$

$Y_t^{(l)} = nz_l(\frac{t}{n}) + O(\lambda n)$, uniformly for $0 \le t \le \sigma n$ and for each $1 \le l \le a(n)$, where $z_l(x)$ is the solution in (a) of theorem 3.1 with $\hat{z}_l = \frac{1}{n}Y_0^{(l)}$ and $\sigma = \sigma(n)$ is the supremum of those x to which the solution can be extended before reaching within l^∞ distance $C\lambda$ of the boundry of D.

Proof. See lemma 3.12. □

4.1 Ball's point of view

This model describes trade flow from one country to all others, based on one very unrealistic assumption: trade between the country and each other country has equal probability.

As in the section above, the choice of a depends on n. This time, denote $Y_l(t)$ the amount of bins with l balls. If there is a maximum amount of balls thrown (say cn for some[1] $c > 0$), then $0 \le l \le a(n) = cn$. Moreover, $|Y_l(t+1) - Y_l(t)| \le 1$, as a ball can drop

(i) into a bin with l balls and $Y_l(t+1) - Y_l(t) = -1$,

(ii) into a bin with $l-1$ balls, giving $Y_l(t+1) - Y_l(t) = 1$ or

(iii) into another bin, offering $Y_l(t+1) = Y_l(t)$

The expected difference can be denoted as

$$\mathbb{E}[Y_l(t+1) - Y_l(t)|H_t] = -\frac{Y_l(t)}{n} + \frac{Y_{l-1}(t)}{n},$$

for $l \ge 1$ and as $\mathbb{E}[Y_0(t+1) - Y_0(t)|H_t] = -\frac{Y_0(t)}{n}$. This is equivalent to defining

$$f_l\left(\frac{t}{n}, \frac{Y_0(t)}{n}, \dots, \frac{Y_{cn}(t)}{n}\right) = -\frac{Y_l(t)}{n} + \frac{Y_{l-1}(t)}{n},$$

for $1 \le l \le cn$ and $f_0(\frac{t}{n}, \frac{Y_0(t)}{n}, \dots, \frac{Y_{cn}(t)}{n}) = \frac{Y_0(t)}{n}$. This offers already every requirement for Wormald's theorem[2]. To see so, define $\beta(n) = 1$ and $\gamma(n) = 0$, $\lambda_1(n) = 0 = o(1)$. Then $\max_{0\le l\le a(n)} |Y_l(t+1) - Y_l(t)| \le \beta$ with probability $1-\gamma$ and

$$\left|\mathbb{E}[Y_l(t+1) - Y_l(t)|H_t] - f_l\left(\frac{t}{n}, \frac{Y_0(t)}{n}, \dots, \frac{Y_{cn}(t)}{n}\right)\right| \le \lambda_1$$

[1] Take $\lfloor cn \rfloor$ in case $cn \notin \mathbb{N}$ in all that follows.

[2] See, again, theorem 3.1

at least as long as $t \leq T_D$, with $D(n)$ defined as

$$D(n) = \left\{ \mathbf{x} \in \mathbb{R}^{cn+2} : -\epsilon < x_i < c + \epsilon, \text{ for } 0 \leq i \leq cn \right\},$$

for some $\epsilon > 0$, where $\mathbf{x} = (x_0, x_1, \dots, x_{cn+1})$.[3]

4.1.1 Requirements for Wormald's theorem

Wormald's theorem rests on solving the system of differential equations given by

$$\frac{\mathrm{d}z_l(x)}{\mathrm{d}x} = f_l(x, z_0, z_1, \dots, z_{cn}),$$

for each $0 \leq l \leq cn$, on the open, connected set D, passing through $z_l(0) = \frac{1}{n} Y_l(0)$. This is equivalent to solving

$$\begin{aligned}
\frac{\mathrm{d}z_0(x)}{\mathrm{d}x} &= -z_0(x) \\
\frac{\mathrm{d}z_1(x)}{\mathrm{d}x} &= z_0(x) - z_1(x) \\
&\vdots \\
\frac{\mathrm{d}z_{cn}(x)}{\mathrm{d}x} &= z_{cn-1}(x) - z_{cn}(x)
\end{aligned} \tag{4.1}$$

passing through $z_0(0) = 1$ and $z_l(0) = 0$ for $1 \leq l \leq cn$. Notice that if $Y_{cn}(t) > 0$ for some t, then there are cn balls in some bin and thus there must be cn balls thrown, implying that one might change the last line in (4.1) to $\frac{\mathrm{d}z_{cn}(x)}{\mathrm{d}x} = z_{cn-1}(x)$. I do not do this.

The solutions to (4.1) are given by $z_l(x) = \frac{x^l}{l!} e^{-x}$ for each $0 \leq l \leq cn$. This offers the following corollary.

Corollary 4.1. *For each $\lambda(n) > 0$ with $\lambda = o(1)$, it holds that*

$$Y_l(t) = \frac{t^l}{n^{l-1} l!} e^{-\frac{t}{n}} + O(\lambda n)$$

with probability

$$1 - O\left(\frac{n}{\lambda} \exp\left(-n\lambda^3\right)\right),$$

uniformly for $0 \leq t \leq \sigma n$ and for each $0 \leq l \leq cn$ where $\sigma(n)$ as before.

Proof. This follows from everything that is stated above, theorem 3.1 and lemma 4.1. □

[3] In case $c > 0$ but not in $\mathbb{N}$, just take $\lfloor cn \rfloor$ and look at the case where $\lfloor cn \rfloor$ balls are thrown, hence $D(n)$ is defined for elements in $\mathbb{R}^{\lfloor cn \rfloor} + 2$.

This corollary tells for instance that if one choses $\lambda = \left(\frac{2\log(n)}{n}\right)^{1/3}$, then there exists some $C_0 > 0$ such that

$$\left| Y_l(t) - \frac{t^l}{n^{l-1} l!} e^{-\frac{t}{n}} \right| \leq C_0 \log(n)^{1/3} n^{2/3}$$

with probability

$$1 - O\left(\frac{1}{n^{2/3}\log(n)^{1/3}}\right).$$

Corollary 4.2. *Let $Y_0(t)$ denote the amount of empty bins at time t. There exists a $C_0 > 0$ such that*

$$\mathbb{P}\left(\left|Y_0(t) - ne^{-\frac{t}{n}}\right| \leq C_0 \log(n)^{1/3} n^{2/3}\right) = 1 - O\left(\frac{1}{n^{2/3}\log(n)^{1/3}}\right)$$

Take for example $c = 4$ and $n = 500$. This yields the following plot (figure 4.1) for $nz_l(c)$, where l is the variable.

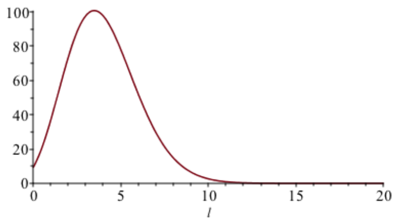


Figure 4.1: The solutions to the differential equations yielded by Wormald's theorem for different values of l, in $x = c$.

Another example can be found in figure 4.2, where different values for n are displayed, given that $C_0 = 0.01, c = 30$. This figure shows that

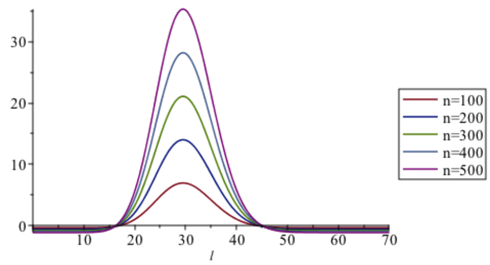


Figure 4.2: The solutions to the differential equations yielded by Wormald's theorem for different values of l, in $x = c$ and plotted for different n, plus the respective "error" for each n.

.

there is some area about which the theorem does tell a lot (or at least something), but also a vast area for which it is hard to say things about the maximum or minimum. With high probability, the maximum does not lay within the bigger bulge, in the example the bulge ends at approximately $l = 40$, hence the maximum should be bigger than 40. However, it is hard to differentiate between for instance $l = 50$ and $l = 60$, because the solution to the differential equations yielded by Wormald's theorem are close to zero and the "error" interval remains the same over all l.

Two problems arise. First, what is C_0? This is very important, because if for instance $C_0 = 1$ and $c = 30$, the moment that the solution to the differential equations minus the error term becomes bigger than zero, lies somewhere between $n = 10^6$ and $n = 10^7$, while for $C_0 = 0.01$ and $c = 30$, this happens for all $n \in \mathbb{N}$. Second, what determines the spread, hence which values determine lowerbounds to the maximum and upperbounds to the minimum?

4.1.2 Bounds for the maximum and minimum

The above offers upper bounds in probability for the amount of bins with l balls. With this information, it is possible to find bounds for the maximum amount of balls in a bin. Define M_n as the maximum amount of balls when there are n bins and cn balls thrown. Notice that, for $k = 0, 1 \dots, cn$

$$\mathbb{P}(M_n \geq k) = \mathbb{P}(Y_l(cn) \geq 1, \text{ for some } k \leq l \leq cn).$$

See also that the event $\bigcup_{i=k}^{cn} \{Y_i \geq 1\}$ is not necesarrily a union of independent sets, but the basic probability laws (Jacob & Protter, 2004, p. 8) do offer the inequality

$$\mathbb{P}(M_n \geq k) \geq \mathbb{P}(Y_k(cn) \geq 1).$$

Figure 4.3 shows the upper- and lowerbound (that has the probability as in Wormald's theorem) for $Y_l(cn)$ at different values of l. This figure, together

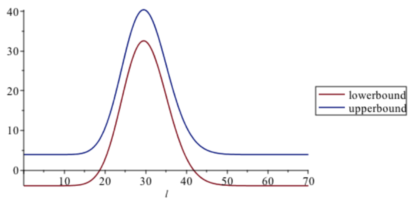


Figure 4.3: Upper- and lowerbounds given by Wormald's theorem for $Y_l(cn)$ at different values of l, given $n = 500$, $c = 30$ and $C_0 = 0.03$.

.

with figure 4.2 suggests that with high probability, the theorem of Wormald offers upper and lowerbounds for the minimum and maximum respectively.

Theorem 4.1. *Let $c > 0$, then $M_n \geq h(n)$ with probability $1 - O\left(\frac{1}{n^{2/3}\log(n)^{1/3}}\right)$, where $h(n)$ is the greatest integer solution in l to*

$$\frac{c^l}{l!} \geq \frac{(1 + C_0 \log(n)^{1/3} n^{2/3})e^c}{n},$$

if such a solution exists, and $h(n) = 0$ otherwise. Moreover, for each $l \in \mathbb{N}$ there exists an $N_l \in \mathbb{N}$ such that $h(n) \geq l$ for each $n \geq N$.

Proof. Let $c > 0$. Chose again $\lambda = \left(\frac{2\log(n)}{n}\right)^{1/3}$. Let $C_0 > 0$ such that

$$Y_l(cn) \in \left[\frac{c^l n}{l!}e^{-c} - C_0 \log(n)^{1/3} n^{2/3}, \frac{c^l n}{l!}e^{-c} + C_0 \log(n)^{1/3} n^{2/3}\right]$$

with probability at least $1 - \frac{C_1}{n^{2/3}\log(n)^{1/3}}$, for each $1 \leq l \leq cn$ and some $C_1 > 0$. This is possible by corollary 4.1. Define $g(n) = n - \log(n)^{1/3}n^{2/3}$, then $g(n) \to \infty$ as $n \to \infty$. This implies that there exists an $N_l \in \mathbb{N}$, such that for each $n \geq N_l$

$$\frac{c^l}{l!}e^{-c}n - C_0 \log(n)^{1/3} n^{2/3} \geq 1,$$

hence for each $n \geq N_l$, the probability that the maximum is greater than l has at least probability

$$1 - \frac{C_1}{n^{2/3}\log(n)^{1/3}}.$$

For each l, this is dependent on c. Hence indeed, $M_n \geq h(n)$ with the asked probability, where $h(n)$ is the either 0 or the greatest integer solution in l to

$$\frac{c^l}{l!} > \frac{(1 + C_0 \log(n)^{1/3} n^{2/3})e^c}{n}.$$

That there exists an $N_l \in \mathbb{N}$ for each l such that $h(n) > l$ follows from the choice of N_l above. This finishes the proof of the theorem. □

Notice that theorem 4.1 also holds for the minimum, although it must be slightly adjusted. This is formulated in the following corollary.

Corollary 4.3. *Take $c > 0$. Let m_n define the minimum amount of balls in any bin after cn balls are thrown, then $m_n \leq \hat{h}(n)$ with probability $1 - O\left(\frac{1}{n^{2/3}\log(n)^{1/3}}\right)$, where $\hat{h}(n)$ is the smallest integer solution in l (greater or equal to 0) to*

$$\frac{c^l}{l!} > \frac{(1 + C_0 \log(n)^{1/3} n^{2/3})e^c}{n},$$

or $\hat{h}(n) = cn$ if no solution exists. Moreover, for each $l \in \mathbb{N}_{\geq 0}$ there exists an $N_l \in \mathbb{N}$ such that $\hat{h}(n) \leq l$ for all $n \geq N_l$.

Proof. The proof is completely analogous to the proof of theorem 4.1. □

Another remark on the theorem is about some knowledge on $Y_l(cn)$. Because, for each t, $Y_l(t)$ maps in $\mathbb{N}$, the theorem can actually be updated to l being the greatest integer solution to

$$\frac{c^l}{l!} > \frac{C_0 \log(n)^{1/3} n^{2/3} e^c}{n},$$

because if one knows that $Y_l(cn) > 0$ with probability p, then $Y_l(cn) \geq 1$, with probability p.

4.1.3 The Range Distribution

Next to the maximum and the minimum alone, there is also the range distribution. The range in n, after cn balls are thrown, is defined as $M_n - m_n$, hence one is interested in $\mathbb{P}(M_n - m_n \geq k)$ for each $1 \leq k \leq cn$. It is helpfull to rewrite $k = l_1 - (l_1 - k)$, for some $k \leq l_1 \leq cn$. Then, one might look at the events $\{M_n \geq l_1\}$ and $\{m_n \leq l_1 - k\}$. Last, notice that

$$\begin{aligned} \mathbb{P}(M_n - m_n \geq k) &= \mathbb{P}\left(\bigcup_{i=k}^{cn} \{M_n \geq i\} \cap \{m_n \leq i - k\} \right) \\ &\geq \mathbb{P}\big(\{M_n \geq l_1\} \cap \{m_n \leq l_1 - k\}\big). \end{aligned}$$

Hence, finding an appropriate l_1 may be enough to find good bounds for the range distribution. If the events $\{M_n \geq l_1\}$ and $\{m_n \leq l_1 - k\}$ are almost independent or can be expressed in terms of each other, the last part of the equation can be solved to find viable bounds.

4.1.4 Estimating C_0

In the following, notice that the bound C_0 can be chosen independently of λ. Hence, a rescaling of λ by some function λ_0 (and the probability accordingly) can always result in $\lambda_0(n) = \frac{C_0'}{C_0}\lambda(n)$. In what follows, when I say "chose C_0 equal to x" I mean: chose $C_0' = x$ and take some rescaled $\lambda_0(n)$ such that for the $\lambda(n)$ chosen:

$$C_0 \lambda_0 = C_0' \lambda(n).$$

It is thus necesarry to know what C_0 is, because the last part of asymptotic in the probability is rescaled by the power of $1/C_0^3$.

To do so, notice that the function f as used for this model is

$$f_l(x_0, x_1, \ldots, x_{cn}) = x_{l-1} - x_l,$$

for $2 \leq l \leq cn$ and $f_1(x_0, x_1, \ldots, x_{cn}) = -x_1$. This implies that f_l is Lipschitz with constant $L = 1$.

Lemma 4.2. *The value of B as in lemma 3.6 is (for this model) the maximum of*

$$\max\{4\sqrt{2}+1, 2c+2\epsilon+1\}$$

Proof. See lemma 3.1 for $g(n) = O(\lambda) \leq G\lambda(n)$. The particular bound on G is given by: $G = L = 1$, because $\lambda_1 = 0$ and thus

$$\left| f\left(\frac{t+k}{n}, \frac{Y(t+k)}{n}\right) - f\left(\frac{t}{k}, \frac{Y(t)}{n}\right) \right| \leq L\frac{w\beta}{n} \leq L\lambda(n),$$

because $w = \lceil \frac{n\lambda}{\beta} \rceil$ and $\lambda(n) < 1$ can be assumed. In this particular model, $L = 1$. Hence, $g(n)$ as in lemma 3.1 has the property $g(n) \leq L\lambda(n)$. This means that, as in lemma 3.5,

$$\begin{aligned} wg(n) + \kappa\beta\sqrt{2w\alpha} &\leq Lw\lambda(n) + \kappa\beta\sqrt{2w\frac{n\lambda^3}{\beta^3}} \\ &\leq Lw\lambda(n) + \kappa\beta\sqrt{2w\left\lceil\frac{n\lambda}{\beta}\right\rceil\frac{\lambda^2}{\beta^2}} \\ &= (L+\sqrt{2}\kappa)w\lambda(n). \end{aligned}$$

Moreover, an upperbound on κ is determied by (see lemma 3.4):

$$\begin{aligned} \beta + \left| f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + g(n) \right| &\leq \beta + (L+1)\lambda + \beta \\ &\leq (3+L)\beta \end{aligned}$$

This means that $B' = L + \sqrt{2}(3+L) = 3\sqrt{2} + L(\sqrt{2}+1)$. In this particular model, that is equivalent to $B' = 4\sqrt{2}+1$, which finishes part 1.

For the second part, notice that each element in D is upperbounded by $c+\epsilon$ for some $\epsilon > 0$, hence by definition of $\sigma(n)$ it must hold that $(t, z_1(t), z_2(t), \ldots, z_{cn}(t)) \in D$, which implies that

$$|f_l(t, z_1(t), \ldots, z_n(t))| = |-z_{l-1}(t) + z_l(t)| \leq 2c + 2\epsilon,$$

hence the maximum of $|f_l|$ for $t \leq \sup_n \sigma(n)$ is also upperbounded by this. Thus indeed

$$B = \max\{4\sqrt{2}+1, 2c+2\epsilon+1\},$$

by definition of B. □

Lemma 4.3. *Suppose*[4] *$\sigma(n) = O(1)$ and C_0 as in 4.3 is upperbounded universally over all $t \leq \sigma n$ by*

$$C_0 \leq 3\sup_n \left\{ \left(1 + \frac{Bw}{n}\right)^{\lfloor \frac{\sigma n}{w} \rfloor} \right\} - 1 \leq 3e^{A \cdot B} - 1,$$

for some $A > 0$, which is well-defined.

[4] That $\sigma(n)$ is $O(1)$ follows from the fact that D is bounded in each element.

Proof. I will first show that the bound is well-defined and then that it is in fact a bound. Notice that, rewriting $a_n = \frac{1}{B}\frac{n}{w}$, it holds that

$$\left(1+\frac{Bw}{n}\right)^{\lfloor\frac{\sigma n}{w}\rfloor} \leq \left(1+\frac{1}{a_n}\right)^{\sigma B a_n}$$

Moreover, there exists some $A > 0$ such that $\sigma(n) \leq A$ for all n. Thus, also,

$$\left(1+\frac{Bw}{n}\right)^{\lfloor\frac{\sigma n}{w}\rfloor} \leq \left(\left(1+\frac{1}{a_n}\right)^{a_n}\right)^{A\cdot B}.$$

Notice that the limit of the inner part is e as $a_n \to \infty$ as $n \to \infty$. Thus, indeed, the supremum exists.

For the second part, take $i = \lfloor\frac{t}{w}\rfloor$ and rewrite

$$B_i = (\lambda n + w)\left(\left(1+\frac{Bw}{n}\right)^i - 1\right).$$

Notice that $w \leq 2\lambda n$ by definition of w and assumption that $w \geq n^{2/3}$ (else the probability in the theorem is unbounded, hence always satisfied). So, indeed,

$$B_i \leq 3\sup_n\left\{\left(1+\frac{Bw}{n}\right)^i - 1\right\}\lambda n,$$

by part one of this proof. That the bound for C_0 holds, is because the change from k_i to t the change in Y and z is at most $w\beta \leq 2\lambda n$, hence, indeed

$$C_0 \leq 3e^{A\cdot B} - 1.$$

□

Corollary 4.4. *Because D is upperbounded in its first element by $c+\epsilon$, it follows that*

$$C_0 \leq 3e^{cB+\epsilon B} - 1,$$

with $B = \max\{4\sqrt{2}+1, 2c+2\epsilon+1\}$.

Figure 4.4 shows how C_0 evolves as c grows.

4.1.5 Results

Let n be the number of bins, cn the number of balls throw into the bins and $Y_l(t)$ is the amount of bins with l balls. C_0' is a bound chosen, such that

$$\left|Y_l(cn) - \frac{c^l}{l!}e^{-c}\right| \leq C_0' \log(n)1/3n^{2/3}, \tag{4.2}$$

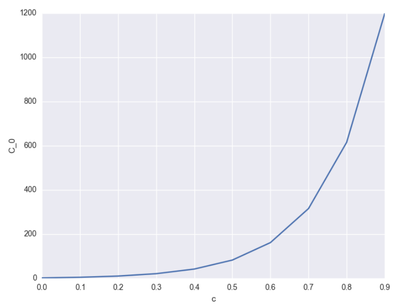


Figure 4.4: Upperbound on C_0 for different c.

See figure 4.6 for realisations of the model for different n and some fixed c, where the bound as in 4.2 is plotted for $C_0' = 0.15$. Figure 4.5 shows histograms for the number of bins with 5 balls to get a basic understanding of the model. The combination of both figures shows that the bound as in 4.3 can probably be made better, because the amount of realisations outside of the bound (as plotted) was less than 1 percent. Moreover, corollary 4.1 offers that for very large n, the probability is bounded below by something that gets arbitrary close to 1 as n grows, for any C_0'.

Maximum

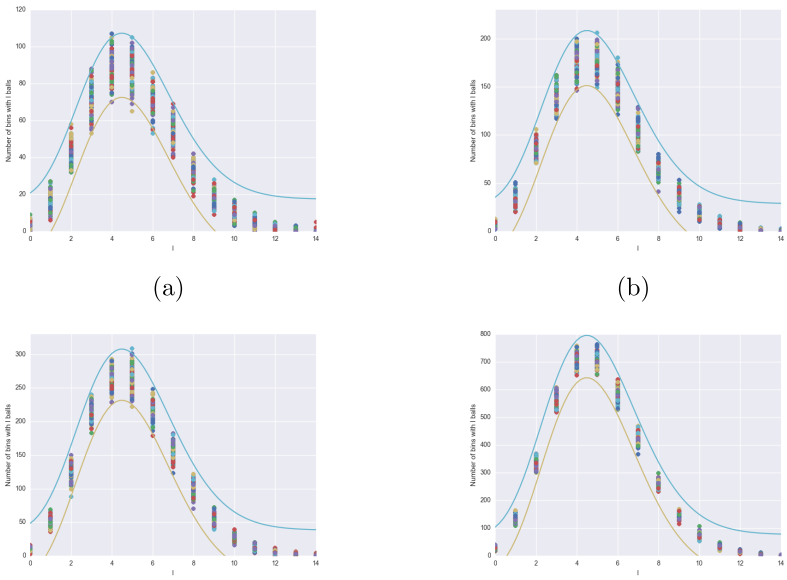
For the maximum, the probability (in cn) that it is bigger than $h(n)$ as in theorem 4.1 has the same probability as $Y_l - \frac{c^l}{l!}e^{-c}$ falling outside the bound $C_0' \log(n)^{1/3} n^{2/3}$, as seen in figures 4.5 and 4.6. See figure 4.7 that plots realisations of the maximum for different values of c, together with the lowerbound $h(n)$. Moreover, the probability that $M_t \geq h(n)$ grows arbitrary close to 1 as n grows by theorem 4.1.

Range

For the range, notice that the probability that $Y_0(t) > 0$ is very small, because the probability that one particular bin gets hit with no ball at a step in time is $1 - \frac{1}{n}$. Thys the probability that it gets no balls after cn throws becomes

$$\left(1 - \frac{1}{n}\right)^{cn} \to \frac{1}{e^c}$$

as $n \to \infty$. Moreover, by corollary 4.3, there exists some $N_l \in \mathbb{N}$ such that $\hat{h}(n) \leq l$ for each $0 \leq l \leq cn$ for $n \geq N_l$, hence the minimum becomes upperbounded by 0. This implies that, although for some c large and n small, the minimum is lowerbounded by something more than 0, but when



(a) (b)

(c) (d)

Figure 4.5: Realisations (100 in each sub-figure) of an equiprobable model, with $n = 500, 1000, 1500, 4000$ respectively, with (a continuous extensions of the) upper- and lowerbound plotted. In all figures, $C_0' = 0.15$ and $c = 5$.

n grows, it is often upperbounded by 0 with high probability. This implies that the range distribution is often equivalent to the maximum distribution.

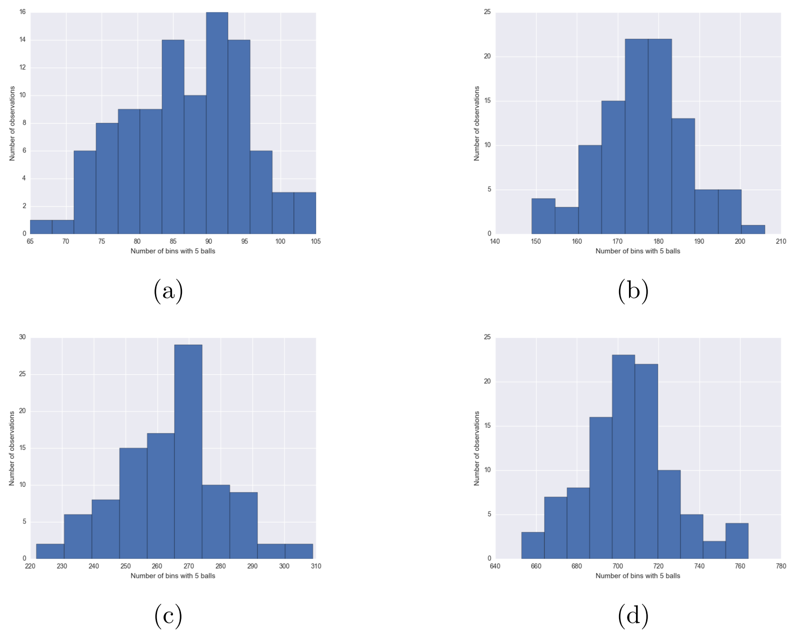


Figure 4.6: Histograms of the number of bins with 5 balls (100 in each figure) of an equiprobable model, with $n = 500, 1000, 1500, 4000$ respectively, here $c = 5$.

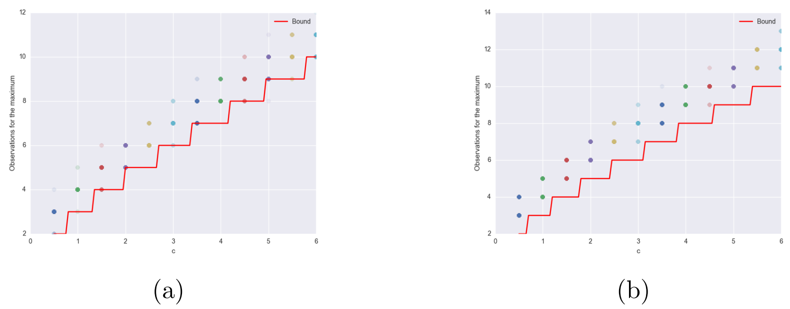


Figure 4.7: Realisations (100 for each value of c) of an equiprobable model, with $n = 500, 1500$ respectively, with (a continuous extensions of the) lower-bound plotted. In all figures, $C_0 = 0.15$.

4.2 Model with different bin sizes: a growing number of balls

A more sophisticated version of the balls and bins model, is one where the probability of a ball hitting two particular bins has different probability. Denote $\tilde{Y}_l(t)$ the amount of balls in bin l at time t. Again, it is asked to model this. Before, I modelled this directly. This time, I shift the process of balls entering bins, as is written down below. At the end, I shift it back.

First, take a distribution $(\pi_1, \dots, \pi_n) \in \mathbb{R}^n$ such that $\pi_i > 0$ for each $1 \leq i \leq n$ and $\sum_{i=1}^n \pi_i = 1$. Second, take some function $N(n)$.[5]. Third, define a random process $Y_l(t)$ for $1 \leq l \leq n$, with $Y_l(0) = N\pi_l$, with the property

$$\mathbb{E}[Y_l(t+1) - Y_l(t)|H_t] = \frac{Y_l(t)}{n}.$$

In this model, t is linked to the amount of balls thrown, but not equal to it. This is no problem, because the distribution after cn balls are thrown is asked. This can be achieved by defining D properly, hence only looking at $\sum_{l=1}^n Y_l(t) \leq cn + N$, or at $Y_l(t) \leq cn$ for each l and t[6]. I model $\tilde{Y}_l(t) = Y_l(t) - N\pi_l$. The results using this model are shown in the sections below.

4.2.1 Wormald's theorem applied

This section contains the results of Wormald's theorem, where I chose some elements specifically. These results are stated as a theorem, proven afterwards. The section below compares it with reality.

Theorem 4.2. *Let $\tilde{Y}_l(t)$ be defined as above and take some $c > 0$, $\lambda(n) > 0$ with $\lambda = o(1)$ and $\beta(n)$ such that*

$$\max_{1 \leq l \leq n} |Y_l(t+1) - Y_l(t)| \leq \beta(n)$$

with probability 1, then it holds that.

$$\mathbb{P}\left(\left|\tilde{Y}_l(t) - N\pi_l(e^{t/n} - 1)\right| = O\left(n\lambda(n)\right)\right) = 1 - O\left(\frac{n}{\lambda}\exp\left(-\frac{n\lambda(n)^3}{\beta(n)^3}\right)\right),$$

Proof. First, take $f_l\left(\frac{t}{n}, \frac{Y_1(t)}{n}, \dots, \frac{Y_n(t)}{n}\right) = \frac{Y_l(t)}{n}$, $\lambda_1(n) = 0$, $\gamma(n) = 0$ and $a(n) = n$. Let $\beta(n) \geq 1$ undefined, as the particular process will define what

$$\max_{1 \leq l \leq n} |Y_l(t+1) - Y_l(t)|$$

[5] For example: $N(n) = cn$

[6] In this case, the restriction on the total amount of balls can be looked at later, specifically because it does not necesarrily happen that there are exactly N balls thrown after some t. Which bin get's balls then?

is exactly. Clearly, it holds that

$$\left|\mathbb{E}[Y_l(t+1) - Y_l(t)|H_t] - f_l\left(\frac{t}{n}, \frac{Y_1(t)}{n}, \dots, \frac{Y_n(t)}{n}\right)\right| \leq \lambda_1(n)$$

with probability $1 - \gamma(n)$. Secondly, the solutions to

$$\frac{\mathrm{d}z_l(x)}{\mathrm{d}x} = z_l(x),$$

going through $z_l(0) = \frac{1}{n}N\pi_l$ are given by $z_l(x) = \frac{1}{n}N\pi_l e^x$. From Wormald's theorem and lemma 4.1, it follows that

$$\mathbb{P}\left(\left|Y_l(t) - N\pi_l e^{t/n}\right| = O\left(n\lambda(n)\right)\right) = 1 - O\left(\frac{n}{\lambda(n)}\exp\left(-\frac{n\lambda(n)^3}{\beta(n)^3}\right)\right).$$

Changing $Y_l(t) = \tilde{Y}_l(t) + N\pi_l$, the theorem follows. □

Notice that the choice of λ determines largely what is in both asymptotics $O(\cdot)$. Moreover, $N\pi_l$ becomes a multiplier of the amount of balls in each bin as time passes. When many balls are thrown, $\tilde{Y}_l(t)$ comes relatively close to $Y_l(t)$.

Moreover, a definition of D is needed. Let $\epsilon > 0$. Define $D(n)$ as

$$\left\{\mathbf{x} \in \mathbb{R}^n : -\epsilon < x_0 < c\left(1 + \frac{N(n)}{cn}\right) + \epsilon, -\epsilon < x_i < c\left(1 + \frac{N(n)}{cn}\right) + \epsilon\right\},$$

where $\mathbf{x} = (x_0, x_1, \dots, x_n)$, then D is clearly bounded in each element as long as $N(n) = O(n)$.

The Model

This section is paritioned in three parts. First, I present a way to find a process that follows all assumptions above. Second, I use this process to show what the theorem tells. Third, I compare these results to realizations of a balls and bins model with different bin sizes.

Claim 4.1. *Let $g : \mathbb{N} \to [0, \infty)$ a function in n. Any process $Y_l(t)$ ($t = 0, 1, 2, \dots$), defined such that (almost surely)*

$$\mathbb{P}(Y_l(t+1) - Y_l(t) = g\left(n\right)|H_t) = \frac{Y_l(t)}{ng\left(n\right)}$$

and

$$\mathbb{P}(Y_l(t+1) - Y_l(t) = 0|H_t) = 1 - \frac{Y_l(t)}{ng\left(n\right)}$$

with the property that $\frac{Y_l(t)}{ng(n)} \in [0,1]$ and such that $Y_l(0) = \pi_l$ is a process that is suitable for theorem 4.2, with $\beta(n) = \max\{g(n), 1\}$.

Proof. Let $Y_l(t)$ as in the claim. Notice that

$$\mathbb{E}[Y_l(t+1) - Y_l(t)|H_t] = \frac{Y_l(t)}{ng(n)} g(n) = \frac{Y_l(t)}{n}.$$

Moreover, $Y_l(t)$ is positive (it can only grow as t grows) and $Y_l(0) = N(n)\pi_l$ by definition. The implication on $\beta(n)$ follows directly from the assumptions in theorem 4.2. □

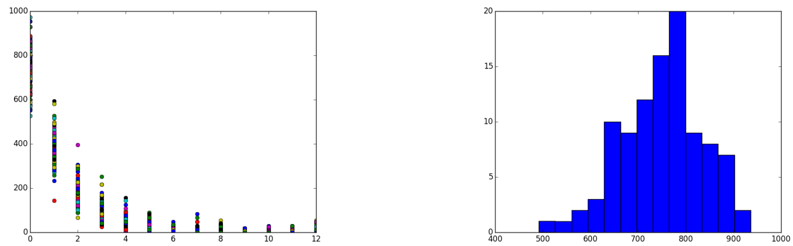
Corollary 4.5. *The process $Y_l(t)$ defined by $Y_l(0) = N(n)\pi_l$ and with*

$$\mathbb{P}\left(Y_l(t+1) - Y_l(t) = 2c \middle| H_t\right) = \frac{Y_l(t)}{2cn},$$

or $Y_l(t+1) - Y_l(t) = 0$ otherwise can be used for theorem 4.2, under the assumption that $Y_l(t) \leq cn + N(n)$ for all l and all relevant t. This implies that $\beta(n) = \max\{2c, 1\}$.

The model as presented in the corollary above is the one I will use in what follows. A different choice of g would, after all, not make any difference on the results, since Wormald's theorem does not use the specific distribution of change in the process.

See figures 4.8 and 4.9 for specific examples. Here, I chose for the first eight bins to have $\pi_i = \frac{1}{2^{i+1}}$ $(1 \leq i \leq 8)$ and all other bins equal probability, hence $\pi_i = \frac{1}{n-8} - \frac{1}{n-8}\sum_{j=1}^{8} \pi_j$, $9 \leq i \leq n$. Last, I took $N = cn$.

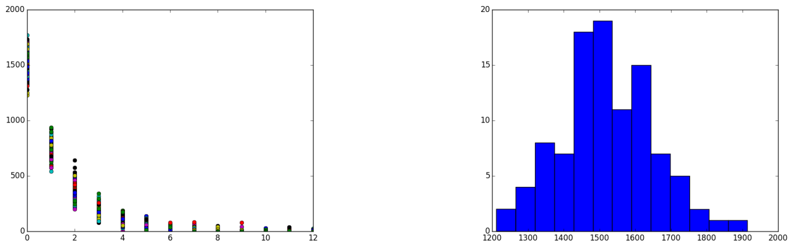


(a) (b)

Figure 4.8: Realisations (100 in each sub-figure) of the model with $n = 500$ and $c = 3$. The left figure is the amount of balls per bin for the first 13 bins, the right picture is a histogram of the amount of balls in the first bin (with $\pi_1 = 1/4$).

4.2.2 Bounds for the maximum and minimum

Contrary to the equiprobable model, the model for different sizes looks at the amount of balls in the bins - not the amount of bins with a certain



(a) (b)

Figure 4.9: Realisations (100 in each sub-figure) of the model with $n = 1000$ and $c = 3$. The left figure is the amount of balls per bin for the first 13 bins, the right picture is a histogram of the amount of balls in the first bin (with $\pi_1 = 1/4$).

collection of balls. Define $M_t = \max_{1 \le l \le n} Y_l(t)$ and $m_t = \min_{1 \le l \le n} Y_l(t)$, then

$$\mathbb{P}(M_t \ge k) = \mathbb{P}\left(\bigcup_{l=1}^{n} \{Y_l \ge k\}\right) \quad \text{and} \quad \mathbb{P}(m_t \le k) = \mathbb{P}\left(\bigcup_{l=1}^{n} \{Y_l(t) \le k\}\right).$$

All that follows, is done for the maximum. Clearly, the same holds for the minimum.

Claim 4.2. *Under the assumption that $Y_i(t)$'s (in i) are independent for each t, it holds that*

$$1 - \prod_{l=1}^{n} \mathbb{P}(Y_l(t) \le k) = \mathbb{P}(M_t > k) \le \sum_{l=1}^{n} \mathbb{P}(Y_l(t) > k),$$

for each k.

Proof. The second inequality holds by the subadditivity of any probability measure, when noticing that $\mathbb{P}(M_t > k) = \mathbb{P}\left(\bigcup_{l=1}^{n} \{Y_l > k\}\right)$. This also implies that

$$\mathbb{P}(M_t > k) = 1 - \mathbb{P}\left(\bigcap_{l=1}^{n} \{Y_l \le k\}\right),$$

hence by independence of Y_l's, the claim follows. □

Explicit forms

All following results are based on two assumptions. The first is that, for all relevant t and k, the sets $\{Y_1(t) \le k\}, \{Y_2(t) \le k\}, \ldots, \{Y_n(t) \le k\}$ are independent. Another result is that $t_f = \log(2)n$ lets $z(t_f/n)$ be an element of D.[7] The theorem below is a general one, which can be applied to find different kind of bounds.

Theorem 4.3. *Take $n_0 \in \mathbb{N}$ and $k > 0$. Let $N(n)$ be a function such that $Y_l(0) = N(n)\pi_l$ for each $1 \leq l \leq n$. Take $I \subset \{1, 2, \ldots, n\}$ and find, for each $l \in I$, λ_l such that*

$$N(n)\pi_l e^{t/n} - C_0 n\lambda_l(n) > k,$$

for each $n \geq n_0$. Then

$$\mathbb{P}(M_t > k) \geq 1 - \prod_{l \in I} C_1 \left(\frac{n\beta(n)}{\lambda_l(n)} \exp\left(-\frac{n\lambda_l(n)^3}{8c^3} \right) \right),$$

for some $C_0, C_1 > 0$.

Another variation of this theorem is presented below, which offers worse bounds for the maximum distribution but 'costs' less to calculate (because for each $1 \leq l \leq n$, or a subset thereof, λ_l must be determined, however more l can be found for which the bound holds). The proof for both theorems is almost identical.

Theorem 4.4. *Let $I \subset \{1, 2, \ldots, n\}$, then*

$$\mathbb{P}\left(M_t > \min_{l \in I} \left\{ N(n)\pi_l e^{\frac{t}{n}} \right\} - C_0 n\lambda(n) \right) = 1 - O\left(\left(\frac{n\beta(n)}{\lambda(n)} \exp\left(-n\frac{\lambda^3}{\beta^3} \right) \right)^{|I|} \right),$$

for each $\lambda(n) = o(1)$, $\lambda > 0$ and $t \leq \sigma n$

Proof. Let $\lambda = o(1)$, $t \leq \sigma n$. For each $l \in I$, it holds by theorem 4.2 that

$$\mathbb{P}\left(\left| Y_l(t) - N(n)\pi_l e^{t/n} \right| \leq C_0\lambda(n) \right) = 1 - O\left(\frac{n\beta(n)}{\lambda(n)} \exp\left(-n\frac{\lambda^3}{\beta^3} \right) \right),$$

which implies that

$$\mathbb{P}\left(Y_l(t) \geq N(n)\pi_l e^{t/n} - C_0\lambda(n) \right) = 1 - O\left(\frac{n\beta(n)}{\lambda(n)} \exp\left(-n\frac{\lambda^3}{\beta^3} \right) \right),$$

Notice that $\min_{l \in I} \left\{ N(n)\pi_l e^{t/n} \right\} \leq N(n)\pi_l e^{t/n}$. The result follows by the fact that $\mathbb{P}(B) + \mathbb{P}(\Omega \setminus B) = 1$, by independence (assumption) of the $Y_l(t)$'s and by claim 4.2. □

[7] This is due to the fact that $z(\frac{t_f}{n}) = c\pi_l e^{\log(2)n/n} = c\pi_l 2$ hence it is away from the boundry of D

4.2.3 The range distribution

Next to a maximum and minimum distribution, there is also the range distribution. Define R_t the range at time t. Notice that, for each $k > 0$, $\{R_t > k\} = \{M_t - m_t > k\}$. There exists some $k \leq l_1 \leq N(n) + cn$ such that[8]

$$\mathbb{P}(R_t > k) \geq \mathbb{P}(\{M_t > l_1\} \cap \{m_t \leq l_1 - k\}) \approx \mathbb{P}(M_t > l_1)\mathbb{P}(m_t \leq l_1 - k).$$

Finding an appripriate l_1 can be done by using theorem 4.4 to lowerbound probabilities on the events $\{M_t > l_1\}$ and $\{m_t < l_1 - k\}$ for multiple l_1 (or all, if needed). Then, the last step is to just take the maximal multiplied probability as the lowerbound. This is not the best possible bound, but just *a* bound presented here as an example.

4.2.4 Bounding C_0.

Rephrasing theorem 4.2 to: there exists some $C_0 > 0$ such that

$$\mathbb{P}\left(\left|Y_l(t) - N(n)\pi_l e^{t/n}\right| \leq C_0\lambda(n)\right) = 1 - O\left(\frac{n\beta}{\lambda}\exp\left(-n\frac{\lambda^3}{\beta^3}\right)\right) \quad (4.3)$$

.

As by theorem 3.1, but mostly the proof of it, C_0 can be determined (or some bound on C_0). When looking at the function defined by

$$f_i(x_1, x_2, \ldots, x_{n+1}) = x_{i+1},$$

for $1 \leq i \leq n$, it clearly holds that

$$|f_i(x_1, \ldots, x_{n+1}) - f_i(y_1, \ldots, y_{n+1})| = |x_{i+1} - y_{i+1}| \leq 1 \cdot \max_{1 \leq i \leq n+1} |x_i - y_i|,$$

hence f_i is lipschitz continuous[9] with lipschitz constant $L = 1$ (however D is defined) and as long as D is bounded.

See lemma ... for the fact that B'' can be determined by L, because f is clearly differentiable in this case.

Lemma 4.4. *The value of B as in lemma 3.6, is the maximum of*

$$\max\{3\sqrt{2} + L(\sqrt{2} + 1), L(3c + 1)\}$$

Proof. For the first element of the maximum, see lemma 4.2.

[8] That the approximation holds is not trivial, as the maximum and minimum are clearly dependent.

[9] In the l^∞ space, as in (Wormald, 1997, p. 34), but also in the l^1 space with the same constant.

A the maximum on D, for $z_l(t)$ (and thus $f_l(t, z_1(t), \dots, z_n(t))$) is given by

$$z_l(t) = \frac{N(n)}{n}\pi_l e^{\frac{t}{n}} \leq 3c,$$

because $z_l(t)$ is at least l^∞ distance $C\lambda$ away from the boundry and t is upperbounded by $\sigma(n)$ such that $z_l(t)$ is away from that boundry by assumption. Hence, indeed,

$$B = \max\{2\sqrt{2} + L(2\sqrt{2}+1), L(3c+1)\},$$

because of the definition of B. □

Corollary 4.6. *For model 1, it follows, by lemma 4.3, that because D is bounded in its first element by $A = 2c$,*

$$C_0 \leq e^{2cB} - 1,$$

with $B = \max\{4\sqrt{2} + 1, 3c + 1\}$.

A better result than the corollary above is also possible, if more is known about the particular choice of λ, B and β. Moreover, if $C_0(t)$ is a function in t, the bound can also be optimized (for t much smaller than $\sigma(n)n$).

4.2.5 Results

It is important to see that C_0 exists for arbitrary $\lambda > 0$. Hence, it is possible to just rescale λ by $1/C_0$ and get the results. It, of course, does change the probability in (b) correspondingly. I do not do reshifts to $\tilde{Y}$ here, because it does not alter the results (besides some basic shift).

Bins

For the following results, chose λ such that $\lambda(n)C_0 = \frac{c^{0.8}}{n^{0.3}}$. Furthermore, let $\pi_i = \frac{1}{2^{i+1}}$ for $1 \leq i \leq 8$ and $\pi_i = \frac{1}{n-8} - \frac{1}{n-8}\sum_{i=1}^{8}\pi_i$. For the amount of balls in particular bins - the random variables for which the process is defined - I offer figures 4.10 and 4.11 The first shows realisations of the model for different n and c fixed (and shows the result for the first four bins). The second shows realisations of the model for different c and fixed n. Each also displayes the solution of the function $z(\frac{t}{n})n$ and the error-bound around this defined by λ. Last, t is chosen to be $\log(2)n$, as for this t it is sure that each $z_l(\frac{t}{n}) = c\pi_l 2 \leq 2c$.

The probabilities that Y_l lays within the bound are of the order given by theorem 3.1. The exact bound, however, depends on C_0 and C_1. It is possible to find universal bounds; bounds that are the same for all bins. However, it is also possible to look at the theorem defined for one bin at a time. This

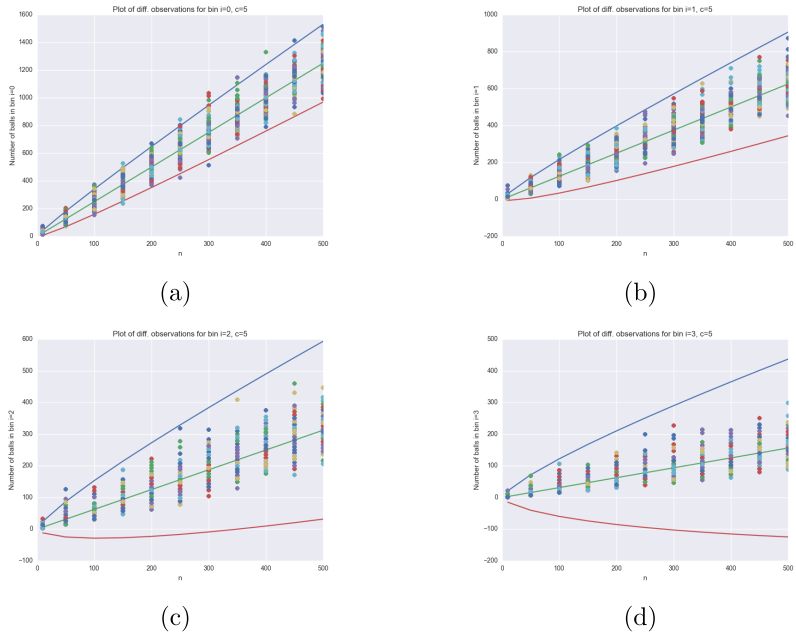


Figure 4.10: Realisations (100 for each value of n) of the model, for $c = 5$ and the first 4 bins.

might offer better bounds for bins that are smaller. For larger bins, the expected difference and the upperbound on the difference are (certainly near the end of the process) quite large.

Moreover for small bins, γ could be chosen differently. If, for instance, $\pi_l(n) = O(1/n^{2+\epsilon})$ for some $\epsilon > 0$, the upperbound $\beta(n) = 2c$ can be replaced by $\beta(n) = 1$, which might offer a better probability (certainly for small n).

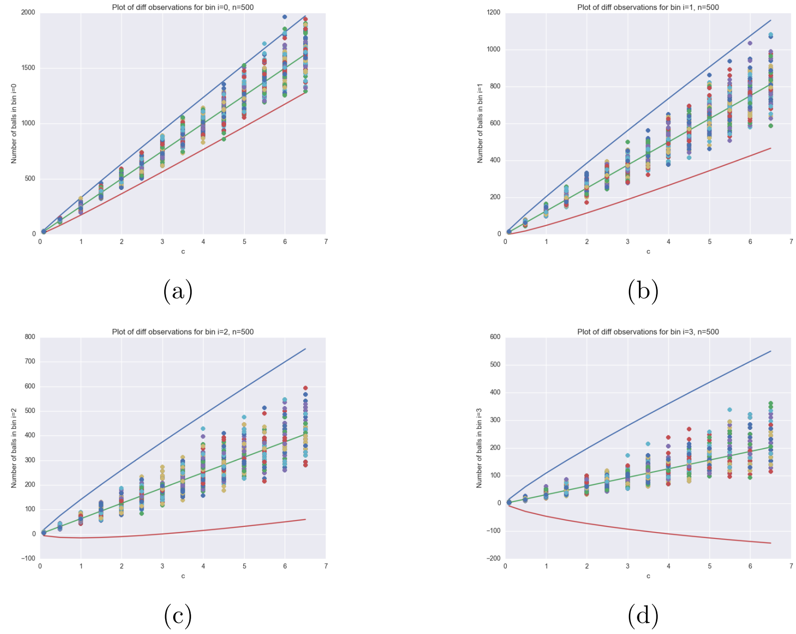
Maximum

The method as described above (with $N(n) = cn$) for arbitrary bin-size distributions works fine to describe how the maximum behaves as long as there exist bins for which $\frac{\pi_l(n)}{\lambda(n)} \to \infty$ as $n \to \infty$ because in this case

$$N(n)\pi_l e^{t/n} - C_0 n\lambda(n) \to \infty, \text{ as } n \to \infty,$$

Hence, for each $k > 0$, there exists some n for which the maximum grows above k with probability $1 - O\left(\frac{n\beta}{\lambda}\exp\left(-n\frac{\lambda^3}{\beta^3}\right)\right)$. Moreover, for some arbitrary $\pi_l(n)$, define

$$N(n)\pi_l e^{t/n} - C_0 n\lambda(n) = k(n),$$



(a) (b) (c) (d)

Figure 4.11: Realisations (100 for each value of c) of the model, for $n = 500$ and the first 4 bins.

then $M_t > k(n)$ with probability $1 - O\left(\frac{n\beta}{\lambda}\exp\left(-n\frac{\lambda^3}{\beta^3}\right)\right)$. This is a little less restricted than the above, because $k(n)$ might have some finite limit as $n \to \infty$.

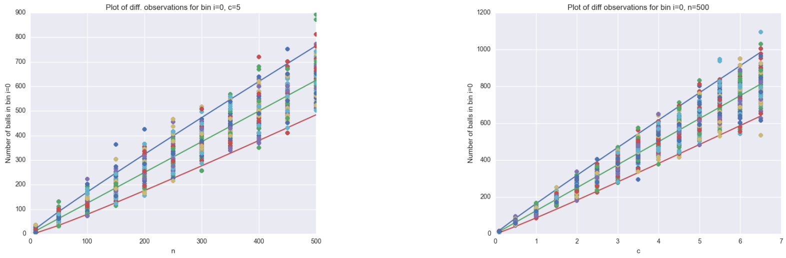
For the maximum, the distribution as used for the bins above is not very interesting, as it is clear that (from a certain point forward) the maximum is going to be determined only by the first bin with very high probability. For instance, when $n = 300$ or larger, this seems to happen (see figure 4.10).

I show it for another distribution, namely one that has $\pi_i = \frac{1}{8}$ for $1 \leq i \leq 4$ and is equiprobable for the remaining part. See figure 4.12 for realisations of the first bin (for different n and c), with $\lambda(n) = \frac{c^{0.8}}{2C_0 n^{0.3}}$.

It is, however, nice to notice that one might define $\lambda_1(n) = \frac{1}{4^{1/3}\lambda(n)}$ and get the same (order of) probability for the maximum being larger than $\frac{cn}{2} - \lambda_1(n)n$. See figure 4.13 for some realisations showing this bound.

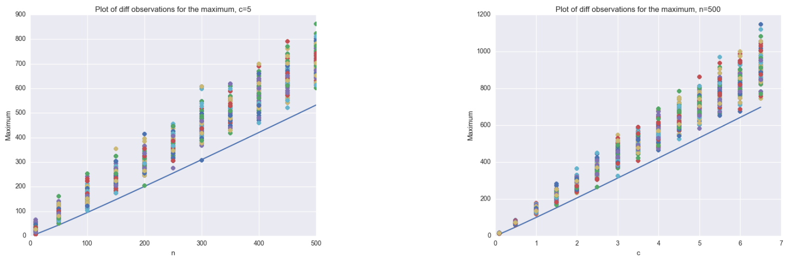
Range

When there exist bins with the property that π_l is lowerbounded in n, then the maximum is determined by those l, as can be seen in theorem 4.4. Moreover, as n gets larger, there must always exist bins for which $\pi_l(n) \to 0$ as $n \to \infty$. The bins with the smallest π_l determine the minimum. Suppose



(a) (b)

Figure 4.12: Realisations (100 for each value of n, c respectively) of the first bin.



(a) (b)

Figure 4.13: Realisations (100 for each value of n, c respectively) of the maximum.

that there are bins that have

$$\mathbb{P}(Y_l(1) - Y_l(0) = 2c | H_0) = \frac{1}{n}\frac{cn}{2cn} = \frac{1}{2n},$$

which offers that, when $\log(2)n$ steps in time are taken, the probability that Y_l gets no balls at all becomes

$$\left(1 - \frac{1}{2n}\right)^{\log(2)n} \to \frac{1}{\sqrt{2}},$$

as $n \to \infty$. In such cases, with other words, the probability that a particular bin gets no balls becomes very large. Of course, one might argue a rescaling of the time in terms of c must happen, for intsance through $t = \log(2c)n$. In this case C_0 would have to change because z_l must be somewhere inside D for this t, which it is not (yet). However, the probability of the bin not being hit would still be $\frac{1}{\sqrt{2c}}$ as $n \to \infty$. Rescaling t in n is not possible, because that would violate the boundedness of D.

Thus, the minimum is (for c not too large, and n not too small) with high probability determined by $N(n) = cn$ in the models above. This implies that, certainly in the models above, the range is determined by the maximum.

In case of small n, the model described by Wormald does not offer good bounds. In case of large c, it also does not, because this influences the bound of D to a large extend (therefore) also C_0 (hence the probability). Of course, this could be fixed by taking n (very) large as well.

Chapter 5

Conclusions and Future Research

Let $a, n \in \mathbb{N}$ arbitrary and if $\{Y_t^{(l)}\}_{t=0}^{\infty}$ is a random process (dependent on n) for each $1 \leq l \leq a$, theorem 3.1 as by Wormald (1997) is used to show that if the difference of a random process in time, scaled by $\frac{1}{n}$, can be described by sufficiently smooth function f_l, the value of the process at time t is close to something determined only by the system of differential equations implied by the functions $f_1, \ldots, f_a$, with a probability that grows as n grows and can get arbitrary close to 1 as n grows for some processes.

I offered a proof of the theorem that follows the lines set out by Wormald (1995), but his proof is short and, to my belief, incomplete. I made it specific and proved his statements my own way, providing particular bounds that are used to find results later on.

Moreover, I updated the original version of his theorem (Wormald, 1995), and later also the generelized one (Wormald, 1997), by proving one part in a different way. This changes the assumptions in the generalised version and leaves one assumption unused in the original version, without changing the result, see theorem 3.2 and lemma A.4.

I applied the theorem on two different examples, both variations of the balls and bins model.

Let n denote the amount of bins and take cn as the amount of balls thrown into those bins, where one ball is thrown at a step in time, hitting either of the n bins with equal probability. If one descibes the amount of bins with l balls for $1 \leq l \leq cn$ as $Y_t^{(l)}$, this is a random process dependent on n that satisfies the assumptions in theorem 3.1. See corollary 4.1 that shows how $Y_{cn}^{(l)}$ is concentrated. This can be used to show that the maximum amount of balls in any particular bin after cn balls are thrown is bigger than a particular function in n, with a probability that can get arbitrary close to 1 as n grows, see theorem 4.1. Moreover, I conclude that (as long as c is relatively small compared to n), the range is determined by the maximum with high probability and the minimum amount of balls in any bin is 0 with high probability.

Let n again be the amount of bins, but at each time t, bin $1 \leq i \leq n$ gets hit with c balls with probability $\frac{Y_t^{(l)}}{2cn}$, where $Y_t^{(l)}$ is the amount of balls in bin l at time t, with a starting distribution given by $Y_0^{(l)} = \pi_l(n)$, and $\sum_{i=1}^{n} \pi_i(n) = 1$, $\pi_i(n) \geq 0$ for each $1 \leq i \leq n$ and each n. In this case, again $Y_t^{(l)}$ is a random variable (for each $t \geq 0$) that satisfies the assumptions in Wormald's theorem (theorem 3.1). Again, I use this to describe the distribution of the maximum amount of balls in any particular bin after $t_f \in \mathbb{N}$ steps in time are taken and show that the maximum is greater than a particular function in n with a probability that gets arbitrary close to 1 as n grows, for some starting distributions $\pi_l(n)$. I note, again, that the range is determined by the maximum (for t_f not too large), because the minimum amount of balls in a bin is 0 with high probability.

5.1 Future Research

5.1.1 Equiprobable model

There are different ways to get better results. I present one possibility. Look at the variables $\hat{Y}_l(t)$ that are defined by

$$\hat{Y}_l(t) = \frac{Y_l(t)}{c},$$

This already implies that the upperbound for $Y_l(t) \leq cn$ can be updated to be $\hat{Y}_l(t) \leq n$. Hence, the upperbound on D for all variables that are not t can be modified to be upperbounded by $1 + \epsilon$ instead of $c + \epsilon$. The rest of the process remains

$$\mathbb{E}(\hat{Y}_l(t+1) - \hat{Y}_l(t)|H_t) = \tfrac{\hat{Y}_{l-1}(t)}{n} + \tfrac{\hat{Y}_l(t)}{n}$$

and the function f remains the same. This results in the fact that C_0 can be upperbounded by $e^{2B}-1$, where[1] $B = 4\sqrt{2}+1$, because also $t \leq n$ instead of $t \leq cn$ (however, this implies that for the original process, some information gets lost). Moreover, $e^{2B} - 1$ is still somewhere between 10^5 and 10^6, but it does not grow as c grows. For a slightly better bound see the section below.

5.1.2 Regular model

In case of small $\pi_l(n)$ in the model for different bin-sizes, the reason the upperbound on C_0 is very unfortunate (see figure 4.4, noting that C_0 in this model is even "worse") because it grows very quick in c. This is possibly because, certainly near the end of the process, "the expected changes

[1] See lemma 4.2 and 4.3

are much smaller than the [...] upperbound on their maximum change" (Wormald, 1997, p.66). Although Wormald notes this for a deletion operation in graphs, it also holds for this process. Thus, there might be some rescaling possible, or one could "use different martingale inequality", replacing the one by Azuma (Wormald, 1997).

Similar to the equiprobable model, I present one possibility. Look at the variables $\hat{Y}_l(t)$ that are defined by

$$\hat{Y}_l(t) = \frac{Y_l(t)}{c},$$

Again, C_0 can be upperbounded by $e^{2B} - 1$, where $B = 4\sqrt{2} + 1$, this is still quite large.

This bound is partly implied by the Lipschitz constant for f. If the process is updated, this can be changed. Look at

$$\mathbb{P}(\hat{Y}_l(t+1) - \hat{Y}_l(t) = 2\kappa) = \frac{\hat{Y}_l(t)}{2n},$$

which offers

$$\mathbb{E}[\hat{Y}_l(t+1) - \hat{Y_l(t)} | H_t] = \kappa \frac{\hat{Y}_l(t)}{n},$$

hence an update in the function f_l, being now

$$f_l(x_0, x_1, \ldots, x_n) = \kappa x_l.$$

This shows that the lipschitz constant of f_l becomes κ. However, before approximately the same amount of balls are thrown, the time also has to be rescaled; by $\frac{1}{\kappa}$.

The bound for B can be updated using the knowledge that

$$\beta + \left| f_l\left(\frac{t}{n}, \frac{\hat{Y}_1(t)}{n}, \ldots, \frac{\hat{Y}_n(t)}{n}\right)\right| \leq \beta + \kappa(1 + \tfrac{w}{n}),$$

by definition of f_l and defining $g(n) = \kappa \frac{w}{n}$. In part one of lemma 4.2, this implies that B' can be updated to equal $\sqrt{2} + (2 + \sqrt{2})\kappa$. Noticing that $z(t) = \pi_l(n) e^{\kappa t}$ must be upperbounded by $2 + \epsilon$ to remain in D at all, it holds that t must be upperbounded by at least

$$t \leq \inf_{n \geq n_0} \left\{ \min_{1 \leq l \leq n} \left\{ \log\left(\frac{2+\epsilon}{\pi_l(n)}\right)\right\}\right\} \frac{1}{\kappa},$$

for some n_0. The exact result different for each distribution π. Defining A as the bound above, multiplied by κ, the total bound becomes

$$C_0 \leq 3e^{\frac{1}{\kappa} A \cdot B} - 1 = 3\exp\left(\frac{1}{\kappa} A\left(\sqrt{2} + (\sqrt{2}+2)\kappa\right)\right) - 1 = 3\exp\left(\frac{A}{\kappa} + 2 + \sqrt{2}\right) - 1.$$

If κ gets bigger, then the lipschitz constant gets worse, but the effect is (apparently) less than the rescaling of time.

References

Aliprantis, C. D., & Burkinshaw, O. (1998). *Principles of Real Analysis* (Third ed.). San Diego: Academic Press.

Armenter, R., & Koren, M. (2014). A Balls-and-Bins Model of Trade. *The American Economic Review*, *104*(7), pp. 2127-2151.

Azuma, K. (1967). Weighted Sums of Certain Dependent Random Variables. *Tohoku Math*, *19*(3), pp.357-367.

Corrado, C. J. (2011). The Exact Distribution of the Maximum, Minimum and the Range of Multinomial/Dirichlet and Multivariate Hypergeometric Frequencies. *Statistics and Computing*, *21*(3), pp. 349-359.

Heinonen, J. (2005). *Lectures on Lipshitz Analysis* [Lecuture Notes]. University of Jyväskylä.

Hurewiz, W. (1958). *Lectures on Ordinary Differential Equations*.

Jacob, J., & Protter, P. (2004). *Probability Essentials* (Second ed.). Berlin Heidelberg: Springer-Verlag.

Ogay, A. (2016). *Exact Distributions of the Multinomial Order Statistics* (Master's Thesis). Delft University of Technology.

Rosenthal, J. S. (2006). *A First Look at Rigorous Probability Theory* (Second ed.). Singapore: World Scientific Publishing Co. Pte. Ltd.

Shamir, E., & Spencer, C. (1987). Sharp Concentration of the Chromatic Number of Random Graphs $G_{n,p}$. *Combinatorica*, *7*(1), pp.121-129.

Vuik, C., van Beek, P., Vermolen, F., & van Kan, J. (2007). *Numerical Methods for Ordinary Differential Equations* (Second ed.). Delft: VSSD.

Wormald, N. C. (1995). Differential Equations for Random Processes and Random Graphs. *The Annals of Applied Probability*, *5*(4), pp.1217-1235.

Wormald, N. C. (1997). *The Differential Equations Method for Random Graph Processes and Greedy Algorithms* [Lecuture Notes]. University of Melbourne.

Appendix A

Proof original theorem of Wormald

The theorem of Wormald is stated below, but proven in a vast amount of different lemma's for readability purposes. Here, I follow the proof by Wormald (1995).

A.1 Proof of Wormald's theorem

I will prove the theorem by Wormald (1995) using different lemma's, finally adding up to the full theorem. I must add that in the following, all three main assumptions (i'), (ii) and (iii) of central theorem hold, but I will explicitely mention when they are used. Assume the result in (a) holds (given the assumptions in the theorem, that is), which is a well-known result for ordinary differential equations, as noted by Wormald (1995), see for instance Hurewiz (1958, p.32, theorem 11 and 12). As a last remark before I start the proof, I need to say that at first, I assume $a = 1$ and $l = 1$ for that matter.

A.1.1 Transformation to Martingale

Before I can start, the following claim - albeit it is almost trivial - is needed.

Claim A.1. *Let each* $(s, z) \in \mathbb{R}^2$ *with* $\mathbb{P}(Y_{sn} = zn) \neq 0$ *for some* $sn = 0, 1, 2, \ldots, m(n)$ *in* D*, then* $\left(\frac{t+k}{n}, \frac{Y_{t+k}}{n}\right) \in D$ *for each* $0 \leq k \leq w(n)$ *and* $0 \leq t \leq m(n) - w(n)$*, where* $w(n)$ *as in assumption (i') and* $m(n)$ *as in the general assumptions.*

Proof. Let $0 \leq k \leq w(n)$ and $0 \leq t \leq m(n) - w(n)$. Take $s = \frac{t+k}{n}$ and $z = \frac{Y_{t+k}}{n}$. Notice that $sn = t + k \leq m(n) - w(n) + w(n) = m(n)$. Then $\mathbb{P}(Y_{t+k} = Y_{t+k}) \neq 0$ trivially, hence $\left(\frac{t+k}{n}, \frac{Y_{t+k}}{n}\right) \in D$ on the whole space (i.e. for each ω fixed). □

One can aruge that the assumptions in this claim are largely overdone. I still use them, because it is used in the original proof of the main theorem (Wormald, 1995) and since it will become handy when getting deeper in the proof. Now, it is time to state the first lemma.

Lemma A.1. *Assume that, always,*

$$|Y_{t+k+1} - Y_{t+k}| \leq \frac{\sqrt{w}}{\lambda^2 \log(n)}$$

for all k (I call this assumption (a1)) and each $(s,z) \in \mathbb{R}^2$ with $\mathbb{P}(Y_{sn} = zn) \neq 0$ for some $sn = 0, 1, 2, \ldots, m(n)$ is in D (I call this assumption (a2)) and $0 \leq t \leq m - w$. Then there exists a function $g(n) = o(1)$ such that

$$M_k = Y_{t+k} - Y_t - kf\left(\frac{t}{n}, \frac{Y_t}{n}\right) - kg(n)$$

is a sumpermartingale with respect to the σ-algebra's generated by $H_t, \ldots, H_{t+w}$.

Notice that the map $\omega \mapsto f\left(\frac{t}{n}, \frac{Y_t(\omega)}{n}\right)$ is H_t-measurable (see also claim 2.1), which will be used in the proof of this lemma. Last remark on the lemma, redefining $\mathcal{F}_i = \sigma(H_{t+i})$ would fit the definition of martingales, however it would make one lose a feeling of what is actually being using.

Proof. First, I show that, for $0 \leq k < w$

$$\begin{aligned}
\mathbb{E}[Y_{t+k+1} - Y_{t+k} | H_{t+k}] &= f\left(\frac{t+k}{n}, \frac{Y_{t+k}}{n}\right) + o(1) \\
&= f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + o(1). \qquad \text{(A.1)}
\end{aligned}$$

The first equality follows directly from assumption (ii). For the second equality, notice that $k = o(n)$ (which trivially holds) as $k < w \leq \frac{n^{2/3}}{\lambda}$ and $\lambda \to \infty$ as $n \to \infty$. Therefore,

$$\left| \frac{t+k}{n} - \frac{t}{n} \right| = o(1).$$

Moreover, by the assumption in Lemma A.1, one gets

$$\begin{aligned}
|Y_{t+k} - Y_t| &\leq k \frac{\sqrt{w}}{\lambda^2 \sqrt{\log(n)}} \\
&\leq \frac{n^{2/3} n^{1/3}}{\lambda^{7/2} \sqrt{\log(n)}} \\
&= \frac{n}{\lambda^{7/2} \sqrt{\log(n)}} = o(n),
\end{aligned}$$

since $\lambda \to \infty$ as $n \to \infty$ (thus also $\lambda^{7/2}\sqrt{\log(n)} \to \infty$). Without further ado, it must now hold that:

$$\left|\frac{Y_{t+k}}{n} - \frac{Y_t}{n}\right| = o(1)$$

and thus, by assumption (iii) - that f is Lipschitz-continous on D[1],

$$\left|f\left(\frac{t+k}{n}, \frac{Y_{t+k}}{n}\right) - f\left(\frac{t}{n}, \frac{Y_t}{n}\right)\right| = o(1)$$

Thus indeed, equation (A.1) holds.

This finishes the first part. In the second part, I prove that M_k is indeed a supermartingale. Notice, at first, that the existence of a function $g(n) = o(1)$ follows directly from part one of this proof, as $g(n)$ can be taken such that

$$g(n) = \max_{0 \leq k < w(n)} \left\{ \left| \mathbb{E}[Y_{t+k+1} - Y_{t+k} | H_{t+k}] - f\left(\frac{t}{n}, \frac{Y_t}{n}\right) \right| \right\}$$

Secondly, it follows that

$$\begin{aligned}
\mathbb{E}[M_{k+1}|H_{t+k}] &= \mathbb{E}\left[Y_{t+k+1} - Y_t - (k+1)f\left(\frac{t}{n}, \frac{Y_t}{n}\right) - (k+1)g(n)\middle| H_{t+k}\right] \\
&= \mathbb{E}\left[Y_{t+k+1} - Y_{t+k}\middle| H_{t+k}\right] + Y_{t+k} - Y_t \\
&\quad - (k+1)f\left(\frac{t}{n}, \frac{Y_t}{n}\right) - (k+1)g(n) \\
&\leq Y_{t+k} - Y_t - kf\left(\frac{t}{n}, \frac{Y_t}{n}\right) - kg(n) \\
&= M_k,
\end{aligned}$$

hence M_k is a super-martingale. The second equality here holds because Y_t, Y_{t+k} and $f\left(\frac{t}{n}, \frac{Y_t}{n}\right)$ are H_{t+k}-measurable[2] and the inequality follows from part one of the proof and the choice of $g(n)$. This shows the lemma. □

The choice of $g(n)$ here is very helpfull in a later part of the proof. Of course, chosing

$$g'(n) = \max_{0 \leq k < w(n)} \left\{ \mathbb{E}[Y_{t+k+1} - Y_{t+k} | H_{t+k}] - f\left(\frac{t}{n}, \frac{Y_t}{n}\right) \right\}$$

would suffice in the proof of this lemma. The particular $g(n)$ - as in the proof of the lemma - has two nice properties, which is the reason for chosing it here already: it is positive and it also lowerbounds $\mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}]$. This last property guarantees that one can also generate a submartingale that looks a lot like the supermartingale M_k, which happens in lemma A.3.

[1] That these elements are indeed in D follows from claim A.1

[2] This is made more specific by claim 2.1

A.1.2 Azuma's inequality

The statements on Azum'as lemma are identical to the proof of theorem 3.1.

A.1.3 Concentration of Y_{t+w} and Y_t

Now that Azuma's lemma is derived and I have shown that a variation on assumption (i'), (ii) and (iii) implicitly transform differences in Y_t into a martingale through a 'clever' trick, I can start showing concentration of Y_t. To do so, I state the following lemma - something that follows almost immediately from Azuma's lemma.

Within, conditional probability is used - which is not the case for Azuma's lemma. However, H_t needs to be 'known' for the creation of the martingale $M_0, M_1, \ldots, M_w$ as it is a martingale with respect to $H_t, H_{t+1}, \ldots, H_{t+w}$. From this point on, I look at the space conditioned on the history up to time t; the space conditioned on H_t.

Lemma A.2. *The assumptions of lemma A.1 hold.* $M_0, M_1, \ldots, M_w$ *as in the proof of lemma A.1, then*

$$\mathbb{P}\left(Y_{t+w} - Y_t - wf\left(\frac{t}{n}, \frac{Y_t}{n}\right) \geq wg(n) + \frac{C_0\sqrt{2}w\gamma}{\lambda^2\sqrt{\log(n)}}\Big| H_t\right) \leq \exp(-\gamma^2) \tag{A.2}$$

for each $\gamma > 0$ *and some* $C_0 > 0$.

Proof. This proof first shows that the supermartingale $M_0, M_1, \ldots, M_w$ suffies the assumptions in Azuma's lemma and then shows the actual result in the lemma. To do so, notice that

$$\begin{aligned}|M_{k+1} - M_k| &= \left|Y_{t+k+1} - Y_{t+k} - f\left(\frac{t}{n}, \frac{Y_t}{n}\right) - g(n)\right| \\ &\leq |Y_{t+k+1} - Y_{t+k}| + \left|f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + g(n)\right| \\ &\leq \frac{\sqrt{w}}{\lambda^2\sqrt{\log(n)}} + \left|f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + g(n)\right|\end{aligned}$$

for each $k \in \{0, 1, \ldots, w-1\}$, which follows from the definition of M_k and because of assumption (a1) in lemma A.1. Furthermore, by assumption (i'), it is known that

$$\frac{\sqrt{w}}{\lambda^2\sqrt{\log(n)}} \geq \frac{\lambda^2\sqrt{\log(n)}}{\lambda^2\sqrt{\log(n)}} = 1$$

Since $\left|\mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}] - f\left(\frac{t}{n}, \frac{Y_t}{n}\right)\right| = o(1)$ by the proof in lemma A.1.

By choice of $g(n) = o(1)$, it thus follows that[3]

$$\left| f\left(\frac{t}{n}, \frac{Y_t}{n}\right) + g(n) \right| \leq \frac{(C_0 - 1)\sqrt{w}}{\lambda^2 \sqrt{\log(n)}}$$

for some $C_0 > 1$. Hence, one can deduce that

$$|M_{k+1} - M_k| \leq \frac{C_0\sqrt{w}}{\lambda^2\sqrt{\log(n)}},$$

or in words that the super-martingale differences are bounded above uniformly over $k = 0, 1, \ldots, w-1$. Last, one must see that

$$\mathbb{E}[M_0] = \mathbb{E}[Y_t - Y_t - 0 \cdot f(\tfrac{t}{n}, \tfrac{Y_t}{n}) - 0 \cdot g(n)] = 0.$$

This concludes part one of this proof, as the assumptions in lemma 3.2 are met.

Now take $c = \frac{C_0\sqrt{w}}{\lambda\sqrt{\log(n)}}$ and $\alpha = \gamma\sqrt{2w}$. From lemma 3.2 we get that

$$\mathbb{P}(M_w \geq \alpha c | H_t) \leq \exp\left(-\frac{\alpha^2}{2w}\right)$$

which is equivalent to

$$\mathbb{P}\left(M_w \geq \gamma\sqrt{2w}\frac{C_0\sqrt{w}}{\lambda^2\sqrt{\log(n)}}\middle| H_t\right) \leq \exp\left(\frac{-\gamma^2 2w}{2w}\right) = \exp(-\gamma^2).$$

To finish the proof, notice that $M_w = Y_{t+w} - Y_t - wf(\frac{t}{n}, \frac{Y_t}{n}) - wg(n)$, hence

$$\mathbb{P}\left(M_w \geq \alpha c\right) = \mathbb{P}\left(Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n}) \geq wg(n) + \alpha c | H_t\right).$$

□

There are a few things left to do to derive at a main result on concentration. First, I state another lemma which tells something about the concentration of $Y_{t+w} - Y_t - wf(\frac{t}{n}, \frac{Y_t}{n})$.

Lemma A.3. *Suppose again all assumptions in lemma A.1 hold. Then*

$$\mathbb{P}\left(|Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq w(g(n) + \tfrac{1}{\lambda})\middle| H_t\right) = o(n^{-1})$$

[3] The particular choice of C_0 is not very important, as it is just a constant, while the lemma says something that holds for every γ. Moreover, it holds because by the statement above, $\frac{\sqrt{w}}{\lambda^2\sqrt{\log(n)}}$ is not in $o(1)$.

Proof. The first part of this proof creates a submartingale $(K_k)_{k=0}^{w(n)}$. Here, I use the particular choice of $g(n)$. Define

$$K_k = Y_{t+k} - Y_t - kf(\tfrac{t}{n}, \tfrac{Y_t}{n}) + kg(n).$$

This is a submartingale. To see so, notice that

$$\begin{aligned}\mathbb{E}[K_{k+1}|H_{t+k}] &= \mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}] - (k+1)f(\tfrac{t}{n}, \tfrac{Y_t}{n}) + (k+1)g(n)\\ &\geq Y_{t+k} - Y_t - kf(\tfrac{t}{n}, \tfrac{Y_t}{n}) + kg(n),\end{aligned}$$

because $-g(n) \leq \mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}] - f(\frac{t}{n}, \frac{Y_t}{n}) \leq g(n)$ by definition of $g(n)$. Therefore $\mathbb{E}[Y_{t+k+1} - Y_{t+k}|H_{t+k}] \geq f(\frac{t}{n}, \frac{Y_t}{n}) - g(n)$. This finishes part one of the proof.

Now, it follows that $-K_k$ is a supermartingale. To see that the differences are bounded by the same bound as for the supermartingale $(M_n)_{n=0}^{w}$, notice that

$$\begin{aligned}|-K_{k+1} + K_k| = |K_{k+1} - K_k| &\leq |Y_{t+k+1} - Y_{t+k} - f(\tfrac{t}{n}, \tfrac{Y_t}{n}) + g(n)|\\ &\leq |Y_{t+k+1} - Y_{t+k}| + |f(\tfrac{t}{n}, \tfrac{Y_t}{n}) - g(n)|\\ &\leq \frac{\sqrt{w}}{\lambda^2\sqrt{\log(n)}} + |f(\tfrac{t}{n}, \tfrac{Y_t}{n}) - g(n)|.\end{aligned}$$

By the proof of lemme A.2, this is bounded above by[4]

$$|f(\tfrac{t}{n}, \tfrac{Y_t}{n}) - g(n)| \leq \frac{(C_0-1)\sqrt{w}}{\lambda^2\sqrt{\log(n)}}$$

hence

$$|-K_{k+1} + K_k| \leq \frac{C_0\sqrt{w}}{\lambda^2\sqrt{\log(n)}}$$

By lemma 3.2 it follows, taking $\alpha = \gamma\sqrt{2\omega}$ and $c = \frac{C_0\sqrt{w}}{\lambda^2\sqrt{\log(n)}}$ that[5]

$$\mathbb{P}(K_w \leq -\alpha c|H_t) = \mathbb{P}(-K_w > \alpha c|H_t) \leq \mathbb{P}(-K_w \geq \alpha c|H_t) \leq \exp(-\gamma^2).$$

This equivalent to

$$\mathbb{P}\left(Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n}) + wg(n) \leq -\alpha c\;|H_t\right) \leq \exp(-\gamma^2).$$

And clearly, it holds that

$$\mathbb{P}(Y_{t+w} - Y_t + wf(\tfrac{t}{n}, \tfrac{Y_t}{n}) \leq -wg(n) - \alpha c|H_t) \leq \exp(-\gamma^2)$$

[4] Here, it is possible to redefine the C_0 in lemma A.2.

[5] See, for instance, lemma A.2, or its proof.

Which implies that also:

$$\mathbb{P}(|Y_{t+w} - Y_t + wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq wg(n) + \alpha c | H_t) \leq 2\exp(-\gamma^2).$$

Now, the proof is almost finished. Take $\gamma = \frac{\lambda\sqrt{\log(n)}}{C_0\sqrt{2}}$, then

$$\exp(-\gamma^2) = \exp\left(-\frac{\lambda^2\log(n)}{2C_0^2}\right) = \exp(\log(n^{-1}))^{\frac{\lambda^2}{2C_0^2}}$$

hence $\exp(-\gamma^2) = (\frac{1}{n})^{\frac{\lambda^2}{2C_0^2}}$. Because $\lambda \to \infty$ as $n \to \infty$, one also gets $\lambda^2 \to \infty$ as $n \to \infty$, so indeed $\exp(-\gamma^2) = o(n^{-1})$. I can thus conclude that, indeed,

$$\mathbb{P}(|Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq w(g(n) + \tfrac{1}{\lambda})|H_t) = o(n^{-1}),$$

because by the choice of γ, it holds that

$$\alpha c = \gamma\sqrt{2w}\frac{C_0\sqrt{w}}{\lambda^2\sqrt{\log(n)}} = \frac{w}{\lambda}.$$

This finishes the proof. □

A.1.4 Generalizing the assumptions

Thus far, all proofs rest on one main assuption (a1): that is $|Y_{t+k+1} - Y_{t+k}|$ is bounded for all k. However, this is not in line with assumption (i') of the main theorem. Therefore, I have to show that the lemma's above, mainly lemma A.2 and A.3, all hold even while assuming something over all "relevant k" (Wormald, 1995, p. 1222).

This - I believe - comes down to very 'basic' probability theory. One can condition the events before on the k's for which assumption (a1) and then look at the probability that a k is indeed such a k. It turns out this does not influence the asymptotic order of the probability I am looking for, which is shown in the lemma below.

Lemma A.4. *Given assumption (i'), i.e.*

$$\mathbb{P}\left(|Y_{t+1} - Y_t| > \frac{\sqrt{w}}{\lambda^2\sqrt{\log(n)}}\Big|H_t\right) = o(n^{-3}),$$

for each $t < m$, always on Ω_n, it (still) holds that, for $t \leq m - w$,

$$\mathbb{P}(|Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq w(g(n) + \tfrac{1}{\lambda})|H_t) = o(n^{-1}).$$

Proof. First, let me define a few sets for readability. For $0 \leq k < w$ and $t \leq m - w$, write

$$B_k = \left\{\omega : |Y_{t+k+1} - Y_{t+k}| \leq \frac{\sqrt{w}}{\lambda^2\sqrt{\log(n)}}\right\}$$

and $B = \bigcap_{k=0}^{w-1} B_k$. Last, write

$$A = \left\{ \omega : |Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq w(g(n) + \tfrac{1}{\lambda}) \right\}$$

Look at B^c, the complement of B, and notice that certainly $\mathbb{P}(B^c|H_t) = o(n^{-2})$, because

$$\mathbb{P}(B^c|H_t) \leq \sum_{k=0}^{w-1} \mathbb{P}(B_n^c|H_t)$$

and $\mathbb{P}(B_k^c|H_t) = o(n^{-3})$ by definition. Moreover, $w \leq \frac{n^{2/3}}{\lambda}$, hence one gets that indeed, $\mathbb{P}(B^c|H_t) = o(n^{-2})$. This also shows that $\mathbb{P}(B|H_t) = 1 - o(n^{-2})$.

Now, write $P_{H_t}(\cdot) = \mathbb{P}(\cdot|H_t)$ for simplicity (and notice that this, on itself, is a probability measure)[6]. By a simple property of probability measures (for instance proven by Jacob and Protter (2004, see theorem 3.4, p. 17)), it holds that

$$\mathbb{P}_{H_t}(A) = \mathbb{P}_{H_t}(A|B)\mathbb{P}_{H_t}(B) + \mathbb{P}_{H_t}(A|B^c)\mathbb{P}_{H_t}(B^c)$$

and since $\mathbb{P}_{H_t}(A|B^c) \leq 1$ (by the most trivial upperbound for the probability measure) and $\mathbb{P}_{H_t}(A|B) = o(n^{-1})$ by lemma A.3, we have that

$$\mathbb{P}_{H_t}(A) = o(n^{-1})(1 - o(n^{-2})) + o(n^{-2}).$$

This shows exactly that $\mathbb{P}(A|H_t) = o(n^{-1})$. □

Hereby, I can end the section on generalizing assumption (a1). This leaves the option to dive into the next part of the proof, where the concentration of $Y_{t+w} - Y_t$ is converted to the concentration of $Y_t - nz(\frac{t}{n})$.

A.1.5 Concentration of $Y_t - nz(\frac{t}{n})$

The concentration of $Y_t - nz(\frac{t}{n})$ - that what the theorem is all about - I show in this section. I use a proof by induction, and to do so, I break this part down into three lemma's, of which I first state the 'biggest', but I prove this last. The reason lies in the definitions and assumptions within the biggest lemma.

Lemma A.5. *Define $k_i = iw$ for $i = 0, 1, \ldots, i_0$, with $i_0 = \min\{\lfloor \frac{m}{w} \rfloor, \lfloor \frac{\sigma n}{w} \rfloor\}$, where m, w, σ as in the assumptions of the theorem ((i'), (ii), (iii)). Then, for some function $\lambda_1 = \lambda_1(n) \to \infty$ as $n \to \infty$, it holds that*

$$\mathbb{P}(|Y_{k_i} - z(\tfrac{k_i}{n})n| \geq B_i) = o(\tfrac{i}{n}). \tag{A.3}$$

with

$$B_i = \frac{\left(\frac{w}{\lambda_1} + \frac{Bw^2}{n}\right)\left((1 + \frac{Bw}{n})^i - 1\right)n}{Bw},$$

for some $B > 0$.

[6] See lemma 2.3

I will prove this lemma by induction. To do so, let me introduce some definitions for readability.

$$\begin{aligned} A_1 &= Y_{k_i} - z(\tfrac{k_i}{n})n \\ A_2 &= Y_{k_{i+1}} - Y_{k_i} \\ A_3 &= z(\tfrac{k_i}{n})n - z(\tfrac{k_{i+1}}{n})n \end{aligned}$$

Now that I have introduced the framework for inductive proof, it is time to state two other lemma's that will help me prove main lemma (A.5) of this section.

Lemma A.6. *There exists a function $\lambda_1(n)$ with $\lambda_1 \to \infty$ as $n \to \infty$, such that*

$$\left| A_2 - wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)\right| < \frac{w}{\lambda_1}$$

with probability $1 - o(n^{-1})$.

Proof. Assume the induction hypothesis, equation A.3. Notice that $g(n) \to 0$ as $n \to \infty$ by choice of $g(n) = o(1)$, because the latter implies that $\lim_{n\to\infty} \frac{g(n)}{1} = 0$. This makes the choice of λ_1 possible, as letting it be defined

$$\lambda_1(n) = \frac{1}{g(n) + \frac{1}{\lambda}},$$

where λ as before. This grows to ∞ because $g(n) \to 0$ and $\frac{1}{\lambda} \to 0$ as $n \to \infty$. Now see that

$$\begin{aligned} &\mathbb{P}\left(\left| A_2 - wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)\right| \geq \frac{w}{\lambda_1} \middle| H_{k_i}\right) \\ &= \mathbb{P}\left(|Y_{k_{i+1}} - Y_{k_i} - wf(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n})| \geq w(g(n) + \tfrac{1}{\lambda}) \middle| H_{k_i}\right). \end{aligned}$$

Taking $t = k_i$, and noticing that $k_{i+1} - k_i = w$, one gets that this is equivalent to

$$\mathbb{P}\left(|Y_{t+w} - Y_t - wf(\tfrac{t}{n}, \tfrac{Y_t}{n})| \geq w(g(n) + \tfrac{1}{\lambda}) \middle| H_t\right).$$

By lemma A.3 (because $t = k_i \leq i_0 w \leq m$) this is $o(n^{-1})$. Lemma 2.5 tells that, because $Y = o(n^{-1})$ implies that $\mathbb{E}[Y] = o(n^{-1})$, the above also holds in the whole probability space, not just conditioned on H_t. This shows that indeed,

$$\left| A_2 - wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right)\right| < \frac{w}{\lambda_1}$$

has probability $1 - o(n^{-1})$. □

Lemma A.7. *Let z be a solution in (a), then*

$$\left| A_3 + wz'\left(\tfrac{k_i}{n}\right)\right| = O\left(\frac{w^2}{n}\right)$$

Proof. By the most simple formula to calculate the first derivative (see for instance Vuik et al. (2007, see theorem 3.2.1, p. 26)):

$$\left| z'\left(\frac{k_i}{n}\right) - \frac{z\left(\frac{k_{i+1}}{n}\right) - z\left(\frac{k_i}{n}\right)}{\frac{w}{n}} \right| = O\left(\frac{w}{n}\right) \leq \frac{Cw}{n},$$

because $\frac{k_{i+1}-k_i}{n} = \frac{w}{n}$, for some constant $C > 0$. This implies that

$$|A_3 + wz'\left(\tfrac{k_i}{n}\right)| \leq \frac{Cw^2}{n} = O\left(\frac{w^2}{n}\right).$$

This finishes the proof of this lemma.[7]

A general approach would be to notice that, because z is continuously differentiable (on the desired domain), there exists a $x \in \left[\frac{k_i}{n}, \frac{k_{i+1}}{n}\right]$ such that

$$z'(x) = \frac{z\left(\frac{k_{i+1}}{n}\right) - z\left(\frac{k_i}{n}\right)}{\frac{w}{n}},$$

because again $\frac{k_{i+1}}{n} - \frac{k_i}{n} = \frac{w}{n}$ by the Mean Value Theorem (Vuik et al., 2007). Hence,

$$\begin{aligned}
\left| z'\left(\frac{k_i}{n}\right) - \frac{z\left(\frac{k_{i+1}}{n}\right) - z\left(\frac{k_i}{n}\right)}{\frac{w}{n}} \right| &= \left| z'\left(\frac{k_i}{n}\right) - z'(x) \right| \\
&= \left| f\left(\frac{k_i}{n}, z\left(\frac{k_i}{n}\right)\right) - f(x, z(x)) \right| \\
&\leq L\left(\left| \frac{k_i}{n} - x \right| + \left| z\left(\frac{k_i}{n}\right) - z(x) \right| \right),
\end{aligned}$$

because f is lipschitz with constant L. This is, however, in case of the l^1 norm lipschitz assumption, for the l^∞, just take the max of the two. Next, notice that

$$z(x) - z\left(\frac{k_i}{n}\right) = \int_{\frac{k_i}{n}}^{x} f(t, z(t))\mathrm{d}t \leq \left(x - \frac{k_i}{n}\right) \max_{t\in[0,\sigma(n)n]} f(t, z(t)),$$

which is a value that f takes on the closed interval because it is continuous by assumption[8] and notice too that $x - \frac{k_i}{n} \leq \frac{w}{n}$ by choice of x. Hence,

[7] Or, in fact, it does not. It must be noted that $z(x)$ is differentiable almost everywhere (with respect to the lebesque measure, that is) because the $D \subset \mathbb{R}^{n+1}$ is open and $f : D \to \mathbb{R}$ is Lipschitz, see for instance Heinonen (2005, p. 18). This also offers the bound for C; it is upperbounded by L, the lipschitz constant of f. Thus, under assumption that f is analytic, the proof up to this point holds.

[8] Thus so is $z(x)$. However, this bound seems to be quite abusive.

indeed,

$$\left| z'\left(\frac{k_i}{n}\right) - \frac{z\left(\frac{k_{i+1}}{n}\right) - z\left(\frac{k_i}{n}\right)}{\frac{w}{n}} \right| = O\left(\frac{w}{n}\right) = O\left(\frac{w}{n}\right),$$

where the exact bound depends on the assumption on f. □

Now, it is time to prove lemma A.5, as all the building blocks are present to do so.

Proof of lemma A.5. First, notice that $z(0) = \frac{Y_0}{n}$, hence the induction hypothesis (equation A.3) holds for $n = 0$. Second, it is helpful to see that

$$\left| A_3 + wf(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}) \right| = \left| A_3 + wz'(\tfrac{k_i}{n}) - wz'(\tfrac{k_i}{n}) + wf(\tfrac{k_i}{n}) \right|.$$

From lemma A.7 and the fact that z is a solution to (a) (i.e. that $z'(x) = f(x, z(x))$ on D), it is known that[9]

$$\left| A_3 + wf(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}) \right| \leq \frac{Cw^2}{n} + \left| wf\left(\tfrac{k_i}{n}, z(\tfrac{k_i}{n})\right) - wf\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right) \right|,$$

for some (constant) $C \geq 0$. Now it is time to use - once more - the fact that f is Lipschitz continuous. This implies that

$$\left| f\left(\tfrac{k_i}{n}, z(\tfrac{k_i}{n})\right) - f\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right) \right| \leq L \left| z(\tfrac{k_i}{n}) - \frac{Y_{k_i}}{n} \right|,$$

for some constant $L \geq 0$. Furthermore, the induction hypothesis (equation A.3) offers that $|Y_{k_i} - z(\frac{k_i}{n})n| < B_i$ with probability $1 - o(\frac{i}{n})$. This shows the following inequality holds,

$$\left| f\left(\tfrac{k_i}{n}, z(\tfrac{k_i}{n})\right) - f\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right) \right| \leq L\frac{B_i}{n},$$

with probability $1 - o(\frac{i}{n})$. Hence I can deduce that

$$\left| A_3 + wf\left(\frac{k_i}{n}, \frac{Y_{k_i}}{n}\right) \right| \leq \frac{Bw^2 + BwB_i}{n},$$

for n sufficiently large, with probability $1 - o(\frac{i}{n})$ and some $B > 0$ constant (this is the part where I take B).

Now, let me move to the final part. To do so, notice that

$$\begin{aligned} |A_1 + A_2 + A_3| &= \left| Y_{k_i} - z(\tfrac{k_i}{n})n + Y_{k_{i+1}} - Y_{k_i} + z(\tfrac{k_i}{n})n - z(\tfrac{k_{i+1}}{n})n \right| \\ &= \left| Y_{k_{i+1}} - z(\tfrac{k_{i+1}}{n}) \right|. \end{aligned}$$

[9] Here, I use that $\frac{k_i}{n} \leq \sigma$ by choice of i_0.

Without further ado, lemma A.6 and A.7 offer the following upperbound[10]

$$\begin{aligned}
|A_1 + A_2 + A_3| &\le |A_1| + \left|A_2 - wf\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right)\right| + \left|A_3 + wf\left(\tfrac{k_i}{n}, \tfrac{Y_{k_i}}{n}\right)\right| \\
&\le B_i + \frac{w}{\lambda_1} + \frac{Bw^2 + BwB_i}{n}, \\
&= \frac{w}{\lambda_1} + \frac{Bw^2}{n} + \frac{(Bw+n)B_i}{n} \\
&= B_{i+1}
\end{aligned}$$

with probability at most $(1 - o(\frac{i}{n}))(1 - o(\frac{1}{n})) = 1 - o(\frac{i+1}{n})$. This finishes the proof, as it shows the induction hypothesis, equation A.3. □

As this lemma is proven, I am ready to show that - given assumption (a2) - the theorem of Wormald holds. This is exactly what the lemma below tells.

Lemma A.8. *Given assumption (a2) and $a = 1$, (b) in theorem 3.3 holds.*

Proof. By lemma A.5, it is know that for $t = k_i$ and $i = 0, \dots, i_0$ with i_0 as before, (b) is satisfied almost surely, because $|Y_{k_i} - nz(\frac{k_i}{n})| < B_i$ almost surely in Ω_n, because

$$\mathbb{P}(|Y_{k_i} - z(\tfrac{k_i}{n})n| < B_i) = 1 - o(\tfrac{i}{n})$$

and $i \le i_0 < n$, hence

$$\mathbb{P}(|Y_{k_i} - z(\tfrac{k_i}{n})n| < B_i) = 1 - o(1).$$

Furthermore, it is usefull to see that

$$\frac{B_i}{n} = \left(\frac{1}{\lambda_1 B} + \frac{w}{n}\right)\left(\frac{1}{Bw}\left(1 + \frac{Bw}{n}\right)^i + \frac{1}{Bw}\right) \to 0$$

as $n \to \infty$ because $w \le \frac{n^{2/3}}{\lambda}$, which implies that $B_i = o(n)$.

Last, it is known by (i') that with high probability $(1 - o(n^{-3}))$, the changes in Y_t between k_i and k_{i+1} are at most

$$|Y_{k_{i+1}} - Y_{k_i}| \le \frac{w\sqrt{w}}{\lambda^2\sqrt{\log(n)}} \le \frac{n}{\lambda^{2/7}\sqrt{\log(n)}} = o(n)$$

So indeed, almost surely,

$$Y_t = z(\tfrac{t}{n})n + o(n),$$

which finishes the proof of this lemma and - only assuming (a2) - almost the proof of theorem 3.3. □

[10] For a more heavy 'calculation', showing the last equality, see claim 3.3

A.1.6 Final generalization

The final generalization - that the assumption on the points that D contains is too strong - makes the proof complete. This shows that the full theorem of Wormald (1995) holds. To see this, I present the following lemma.

Lemma A.9. *Suppose assumption (a2) does not hold, then one still gets (b) in theorem 3.3 for* $a = 1$.

Proof. Take $\epsilon > 0$ and define $D'(\epsilon)$ as the set that contains all points $(s, z) \in D$ that have at least ϵ distance to the boundry of D, denoted as ∂D, in the z-direction, i.e.

$$D'(\epsilon) = \{(s, z) \in D : A_\epsilon(s, z) \cap \partial D = \emptyset\},$$

where $A_\epsilon(s, z) = \{(s, x) \in D : z - \epsilon < x < z + \epsilon\}$ Redefine σ' as σ, but now for D'.

Let $n \in \mathbb{N}$ and take $\epsilon > \frac{1}{\lambda^{7/2}\sqrt{\log(n)}}$. In the induction hypothesis of lemma A.5, one can add that $\left(\frac{Y_{k_i}}{n}, \frac{k_i}{n}\right)$ is in $D'(\epsilon)$. Assuming (a1) also provides that

$$\left|\frac{Y_{k_{i+1}} - Y_{k_i}}{n}\right| \leq \frac{1}{\lambda^{7/2}\sqrt{\log(n)}}$$

Moreover, $\frac{k_{i+1}}{n} \leq \frac{i_0 w}{n} \leq \sigma'$, hence indeed $\left(\frac{k_{i+1}}{n}, \frac{Y_{i+1}}{n}\right)$ does not leave $D'(\epsilon)$, given (a1).

Moreover, the conditioning on relevant k (i.e. conditioning on (a1)), as in lemma **??** gives the same probability, hence the result in that lemma remains when not assuming (a2).

Last, notice that $(0, \frac{Y_0}{n}) \in D'(\epsilon)$ for some n large enough, because $\frac{Y_0}{n} = \hat{z}_0$ and D is open, hence the induction here and in lemma A.5 hold, even for this $D'(\epsilon)$.

Notice, furthermore, that ϵ can be chosen as close to 0 as one wants (for n large enough, that is), so one can get arbitrarily close to σn (as $\sigma' n$ goes to σn as n grows)[11]. One can conclude that (b) holds, even without assuming (a2). □

The last thing that is left to be shown is that the proof works exactly the same for $a \neq 1$. To see this, notice that the probability of

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i),$$

by basic probability law and thus the probability that any of the events $\{|Y_l(k_i) - z(k_i/n)n| > B_i\}$ occur (in l) is upperbounded by the sum of the probability of either event occuring, hence by $o(a\frac{i}{n})$ in lemma A.5.

[11] This shows that, the theorem holds for $0 \leq t \leq \min\{\sigma n, m\}$ and not only for $0 \leq t \leq \min\{\sigma' n, m\}$.

Appendix B

Python Code

B.1 Simple Model

```python
import random
import matplotlib.pyplot as plt
from math import factorial, exp, log, gamma
import numpy as np
import seaborn as sns


def get_max_bound(n, c_range, C0):
    ans_list = [0]*len(c_range)
    for i in range(len(c_range)):
        for l in range(int(c_range[i]), int(n*c_range[
            i])):
            if c_range[i]**l/gamma(l + 1) > (C0*log(n)
                **(1./3)*n**(2./3))*exp(c_range[i])/n:
                ans_list[i] = l
            else:
                break

    return ans_list


def drop_ball(l, n):
    l[random.randint(0, n - 1)] += 1
    return l


def f(x, c, n):
    ans = []
```

```
    for i in x:
        ans.append(float(c)**i/(gamma(i + 1))*exp(-c)*
           n)

    return ans


def bound_probability(n, c, C0):
    N = 100
    count_out = [0]*N
    x = range(1, int(c*n) + 1)
    z = f(x, c, n)
    lower = [i - C0*log(n)**(1./3)*(n**(2./3)) for i
       in z]
    upper = [i + C0*log(n)**(1./3)*(n**(2./3)) for i
       in z]

    for i in range(N):
        l = [0]*n
        for j in range(1, int(c*n) + 1):
            l = drop_ball(l, n)

        print len(l)
        for k in range(len(l)):
            if l[k] > upper[k] or l[k] < lower[k]:
                count_out[i] += 1

    return float(sum(count_out))/len(count_out)


def max_distribution(n, c, C0):
    m = [0]*100
    for j in range(100):
        l = [0]*n
        rtot = [0]*int(c*n)
        for i in range(1, int(c*n)+1):
            l = drop_ball(l, n)

        for i in range(0, len(l)):

            rtot[l[i]] += 1

        # finding the maximum
```

```
        for i in range(len(rtot)):
            if rtot[i] > m[j]:
                m[j] = i

    return m


def equiprobable(n, c, C0):

    sns.set(style="darkgrid")
    l_5 = []

    for j in range(100):
        l = [0]*n
        rtot = [0]*int(c*n)
        for i in range(1, int(c*n)+1):
            l = drop_ball(l, n)

        for i in range(0, len(l)):
            rtot[l[i]] += 1

        l_5.append(rtot[5])
        plt.plot(rtot, 'ro', alpha=0.3)

    x = np.arange(0, 80, 0.2)
    z = f(x, c, n)
    z_lower = [i - C0*log(n)**(1./3)*n**(2./3) for i
       in z]
    z_upper = [i + C0*log(n)**(1./3)*n**(2./3) for i
       in z]

    plt.plot(x, z_lower)
    plt.plot(x, z_upper)
    plt.xlabel("l")
    plt.ylabel("Number of bins with l balls")
    plt.axis([0, 14, 0, 230])
    plt.show()

    plt.ylabel("Number of observations")
    plt.xlabel("Number of bins with 5 balls")
    plt.hist(l_5)
    plt.show()
```

```
def plot_max(n, C0):
    sns.set(style='darkgrid')
    for c in [0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5,
        5.5, 6]:
        plt.plot([c]*100, max_distribution(n, c, C0),
            'o', alpha=0.1)

    c = [0.5 + x/20. for x in range(2*55 + 1)]
    bound = get_max_bound(n, c, C0)
    plt.xlabel('c')
    plt.ylabel('Observations for the maximum')
    plt.plot(c, bound, 'r', label='Bound')
    plt.legend()

    plt.show()


def main():
    n = 1500
    c = 5
    C0 = 0.15
    plot_max(n, C0)
    # equiprobable(n, c, C0)
    # plt.hist(max_distribution(n, c, C0))
    plt.show()


if __name__ == '__main__':
    main()
```

B.2 Different bin-size model

```
from math import log, exp, sqrt
import random


def distribute_balls(begins, n, c, max_time, time=
    False):
    bins = begins[:]
    total = 0
    t = 0
    while t < max_time:
        t += 1
        for i in range(len(bins)):
```

```
            x = random.random()
            if x < float(bins[i])/(2*c*n):
                bins[i] += 2*c
                total += 2*c

    if not time:
        return bins
    else:
        return bins, time


def bin_i_distribution(beginds, n, c, max_time,
   rel_num, i):
    """returns the amounts of balls in bin i, given
       begin distribution (beginds), n, c and the
       number of realisations
    (rel_num)"""
    bin_i = []
    times = []
    t = 0
    if i > n:
        return "Range␣Error␣in␣'bin_i_distribution':␣i
            ␣=␣{0}␣>␣n".format(i)

    for j in range(rel_num):
        b, t = distribute_balls(beginds, n, c,
            max_time, time=True)
        bin_i.append(b[i])
        times.append(t)

    return bin_i, t


def distribution(beginds, n, c, rel_num, **options):
    """returns for each element in which (max, min,
       range, zero) the given values in a list of
       rel_num realisations.
    Give type and it returns max for max, min for min,
        range for range and max, min, range for all (
       or leaving it out)."""
    maxs = []
    mins = []
    ranges = []
    for j in range(rel_num):
```

```
            max_time = log(2)*n
            dist = distribute_balls(beginds, n, c,
                max_time, time=False)
            maxs.append(max(dist))
            mins.append(min(dist))
            ranges.append(max(dist) - min(dist))

        if options.get('type') == 'max':
            return maxs
        if options.get('type') == 'min':
            return mins
        if options.get('type') == 'range':
            return ranges
        if options.get('type') == 'all':
            return maxs, mins, ranges

    return maxs, mins, ranges


def equiprobable_begin_ds(n, c):
    return [float(c)]*n


def special_ds(n, c):
    over = c*n - (c*n/4. + c*n/8. + c*n/16. + c*n/32.
        + c*n/64. + c*n/128. + c*n/264. + c*n/512.)
    return [c*n/4., c*n/8., c*n/16., c*n/32., c*n/64.,
         c*n/128., c*n/264., c*n/512.] + [over/(n - 8)
        ]*(n - 8)


def special_ds_2(n, c):
    over = c*n - (c*n/2.)
    return [c*n/8., c*n/8., c*n/8., c*n/8.] + [over/(n
         - 4)]*(n - 4)


def bound_constant(c):
    return c**(0.8)/2


def sol(n, c, i):
    return special_ds_2(n, c)[i]*2
```

```python
def sol_max(n, c, I):
    min_val = c*n
    for i in I:
        if special_ds_2(n, c)[i] < min_val:
            min_val = special_ds_2(n, c)[i]

    return min_val*2


def lamb(n):
    return 1/(n**0.3)
```