

*$I^2SDS$*   
*The Institute for Integrating Statistics in Decision Sciences*

*Technical Report TR-2010-7*  
**April 26, 2010**

**Phases in the Diffusion of Gases  
via the Ehrenfest Urn Model**

Srinivasan Balaji  
*Department of Statistics*  
*The George Washington University*

Hosam Mahmoud  
*Department of Statistics*  
*The George Washington University*

Zhang Tong  
*Department of Statistics*  
*The George Washington University*

# Phases in the Diffusion of Gases via the Ehrenfest Urn Model

Srinivasan Balaji<sup>1</sup>

Hosam Mahmoud<sup>1</sup>

Zhang Tong<sup>1</sup>

April 26, 2010

## Abstract

The Ehrenfest urn is a model for the diffusion of gases between two chambers. Classic research deals with this system as a Markovian model with a fixed number of balls, and derives the steady-state behavior as a binomial distribution (which can be approximated by a normal distribution). We study the gradual change for an urn containing  $n$  (a very large number) balls from the initial condition to the steady state. We look at the status of the urn after  $k_n$  draws. We identify three phases of  $k_n$ : The growing sublinear, the linear, and the superlinear. In the growing sublinear phase the amount of gas in either chamber is normally distributed, with parameters that are influenced by the initial conditions. In the linear phase a different normal distribution applies, in which the influence of the initial conditions is attenuated. The steady state is not a good approximation until a certain superlinear amount of time has elapsed. At the superlinear stage the mix is nearly perfect, with a nearly perfect symmetrical normal distribution in which the effect of the initial conditions is completely washed away. We give interpretations for how the results in different phases conjoin at the “seam lines.” In fact, these Gaussian phases are all manifestations of one master theorem. The results are obtained via martingale theory.

**AMS subject classifications:** Primary: 60C05, 60F05, 05A05; secondary: 60G42

**Keywords:** Urn model, random structure, martingale, central limit theorem, diffusion of gases.

---

<sup>1</sup>Department of Statistics, The George Washington University, Washington, D.C. 20052, U.S.A.

# 1 The Ehrenfest urn as a model for gas diffusion

The Ehrenfest urn was first proposed as a model for the diffusion of nonreacting gases (Ehrenfest and Ehrenfest, 1907). We deal here with the speed of this diffusion across time phases. The model is for two chambers (say A and B) containing gases (possibly the same). The two chambers are connected through a pipe controlled by a valve. The valve is opened at time 0 and the diffusion proceeds over epochs of time, which we can take as the unity. In each time unit (diffusion step) one molecule of gas randomly chosen from the population of molecules in both chambers jumps from its chamber to the other one. This continual switching of sides affects a gradual diffusion; inducing change in the amount of gas in each chamber. It is of interest to know the amount of gas (number of molecules) in chamber A after a certain period of time.

This physical model of gas diffusion can be visualized in terms of a scheme of drawing balls from an urn. We can think of the molecules in chamber A as balls of a certain color (say white) and those in chamber B as balls of an antithetical color (say red). The gas model with  $n$  molecules can then be viewed as  $n$  balls of two colors all residing in one urn, which evolves in the following manner. At each discrete point in time, we pick a ball at random from the urn. We paint that ball with the opposite color and put it back in the urn. In this equivalent model, the interest is to know the number of white balls (the amount of gas in Chamber A) after a certain period of time.

The classic research deals with this system as a Markovian model with a fixed number of balls, and derives the steady-state behavior as a binomial distribution; see Bellman and Harris (1951), Blom (1989) and Karlin and McGregor (1965); for an overview see Mahmoud (2008).

Antognini (2005) looks at the speed of diffusion for a fixed number of particles. In a physical system the number of gas molecules is very large, we shall take it to be  $n$ , and is apportioned as  $\lfloor \alpha n \rfloor \sim \alpha n$  in chamber A and  $n - \lfloor \alpha n \rfloor \sim (1 - \alpha)n$  in Chamber B, for some  $\alpha \in (0, 1)$ . One is interested in knowing the behavior of the gases after a certain finite interval of time. So, the question is *How many white balls are in the urn after  $k = k_n$  draws, for functions  $k_n$  of various growth rates?* The study of the evolution of urns through various stages prior to the steady state is a topic of recent

interest, see for example, Mikhailov (1977, 1980), Vatutin and Mikhailov (1982), Mahmoud (2010+) and Smythe (2009).

## 2 Scope

We shall identify three phases of  $k_n$ :

- (a) The sublinear phase, when  $k_n = o(n)$ ;
- (b) The linear phase, when  $k_n = \lambda_n n$ , for some  $\lambda_n > 0$  of a magnitude separated from zero and infinity;
- (c) The superlinear phase, when  $n = o(k_n)$ .

We shall prove the following general trends. Trivially, at the very low end of the sublinear phase, when  $k_n = O(1)$ , as  $n \rightarrow \infty$ , there is not much change in the content of the two chambers, only a finite perturbation on the initial conditions can be felt. Changes begin to happen, when  $k_n$  grows to infinity. In what follows the normally distributed random variate with mean 0 and variance  $\nu^2$  is denoted by  $\mathcal{N}(0, \nu^2)$ , and  $\xrightarrow{\mathcal{D}}$  denotes convergence in distribution.

**Theorem 1** *Let  $W_{k_n}$  be the number of white balls in the Ehrenfest urn (molecules in Chamber A) after  $k_n$  draws (gas diffusion steps) from an urn with  $n$  balls, of which initially the number of white balls is  $W_0(n) = \lfloor \alpha n \rfloor$ , where  $k_n \rightarrow \infty$ , in a sublinear, linear or superlinear fashion. Then,*

$$\frac{W_{k_n} - \left( \frac{n}{2} + \left( W_0(n) - \frac{n}{2} \right) \left( \frac{n-2}{n} \right)^{k_n} \right)}{\sqrt{\frac{1}{4}n + \left( \left( \frac{n}{2} - W_0(n) \right)^2 - \frac{n}{4} \right) \left( \frac{n-4}{n} \right)^{k_n} - \left( \frac{n}{2} - W_0(n) \right)^2 \left( \frac{n-2}{n} \right)^{2k_n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

The shift and scale are the mean and variance of  $W_{k_n}$ . This theorem has the following manifestations in various phases. (We used three different approximation techniques in each of the three phases.) When  $k_n$  grows sublinearly to  $\infty$ , one sees normal behavior in the amount of gas in each

chamber, even for a fairly slowly growing function  $k_n$ . We call the phase when  $k_n$  grows sublinearly to  $\infty$ , the *growing sublinear phase*. Functions that are asymptotically as small as  $\frac{1}{20} \ln \ln n$ , for example, are sufficient to give a normally distributed mix in each chamber. For the sublinear phase, the initial conditions persist, and the asymptotic normal result in this case contains the initial condition  $\alpha$ . The Gaussian law in Theorem 1 takes the form

$$\frac{W_{k_n} - n\left(\frac{1}{2} + \left(\alpha - \frac{1}{2}\right)\left(\frac{n-2}{n}\right)^{k_n}\right)}{\sqrt{k_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4\alpha(1-\alpha)).$$

Normality continues to hold in the linear and superlinear phases. However, in each phase we get a different normal distribution; the mean and scale factors are essentially different. In the linear phase a different normal distribution (in the usual style of central limit theorems) is in effect, and the parameters of the distribution depend on both the initial condition  $\alpha$  and the coefficient of linearity.

A typical instance of the linear phase is when  $k_n = cn + o(\sqrt{n})$ , for a positive constant  $c$ , in which case the Gaussian law in Theorem 1 takes the form

$$\frac{W_{k_n} - \left(\left(\alpha - \frac{1}{2}\right)e^{-2c} + \frac{1}{2}\right)n}{\sqrt{\frac{e^{4c} - 1 - 4c(2\alpha - 1)^2}{4e^{4c}}n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Note how the influence of the initial conditions is attenuated as we get deeper in the linear phase.

As one might expect, after a very long period of time, as in the superlinear case, the diffusion is nearly complete. In the superlinear phase, and if additionally  $k_n = \frac{1}{4}n \ln n + g_n$ , for any function  $g_n$  such that  $g_n/n \rightarrow \infty$ , Theorem 1 takes the symmetric form

$$\frac{W_{k_n} - \frac{1}{2}n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4}\right),$$

which is the usual approximation of the binomial distribution by the normal. Note also how the effect of any initial conditions is washed away.

Diaconis (1996) takes a different view and discusses the “cutoff phenomenon” in the Ehrenfest urn model, where the total variation distance to the stationary distribution experiences a sharp decline after a large number of draws (while keeping the size of the urn fixed).

The problem can also be viewed as an allocation scheme of balls in urns, where  $k_n$  balls are dropped randomly in  $n$  urns. A ball in the Ehrenfest urn is represented by an urn in the allocation scheme. As a ball changes color during the history of the Ehrenfest process, the number of balls in the corresponding urn in the allocation model changes parity. More specifically, suppose the balls in the Ehrenfest urn are labeled  $1, \dots, n$ . We label the urns in the allocation scheme with  $1, \dots, n$ , say from left to right, and the  $i$ th urn represents the  $i$ th ball in the Ehrenfest urn. The urns labeled  $1, \dots, W_0(n)$  are the initial white balls in the Ehrenfest model, and the urns labeled  $W_0(n) + 1, \dots, n$  are the initial red balls in the Ehrenfest model. After  $k_n$  ball drops, an urn among the  $W_0(n)$  leftmost urns containing an even number of balls indicates that the corresponding ball in the Ehrenfest urn (initially white) has been drawn an even number of times, and it is now white; let the number of such urns be  $L_{k_n}$ . Likewise, an urn among the  $n - W_0(n)$  rightmost urns containing an odd number of balls indicates that the corresponding ball in the Ehrenfest urn (initially red) has been drawn an odd number of times, and it is now white; let the number of such urns be  $R_{k_n}$ . Then,

$$W_{k_n} = L_{k_n} + R_{k_n}.$$

Allocation scheme formulations, such as this one, received quite a bit of attention, see for example Kolchin, Sevastyanov, and Chistyakov (1976), where the schemes were handled by the method of moments.

### 3 Organization

The rest of this paper has the following organization. In Section 4 we set up a stochastic recurrence and discuss moments. In Section 5 we derive the underlying martingale. In Section 6 we discuss the three phases: the growing sublinear, the linear, and the superlinear, where a subsection is devoted for each phase. In these subsections we prove the announced results. Detailed proofs are relegated to Appendices A and B.

Throughout, we shall use the following standard probability notation. We shall use the symbol  $\xrightarrow{\mathcal{P}}$  for convergence in probability. The notation  $o_{\mathcal{L}_1}(g(n))$  will stand for a sequence of random variables that is  $o(g(n))$  in the  $\mathcal{L}_1$  norm, that is, when we describe a sequence of random variables  $X_n$  to be  $o_{\mathcal{L}_1}(g(n))$ , we mean that  $\mathbf{E}[|X_n|]/|g(n)| \rightarrow 0$ , as  $n \rightarrow \infty$ . We let  $\mathcal{F}_j$  be the sigma field generated by the first  $j$  draws.

Unless otherwise stated, all asymptotics will mean asymptotic equivalents and bounds as  $n \rightarrow \infty$ . The number  $n/(n-2)$  will appear often, and we shall give it the designation  $\rho_n$ . We shall repeatedly use well-known facts about  $\rho_n^{ym}$ , for  $y > 0$ , such as the fact that  $\rho_n^{ym}$  is asymptotically  $e^{2y} + O(1/n)$ .

We shall also need the backward difference operator  $\nabla$ , which when applied to a function  $h(i)$ , with integer argument  $i$ , gives the difference between two successive steps; that is,  $\nabla h(i) = h(i) - h(i-1)$ .

## 4 Exact moments

Let  $W_j = W_j(n)$  be the number of white balls (molecules in Chamber A) after  $j$  such draws (diffusion steps). Let  $I_n^W$  and  $I_n^R$  respectively be the indicators of picking a white or a red ball in the  $n$ th step. Because of their mutual exclusion, we have  $I_n^R = 1 - I_n^W$ . There is stochastic dependence between  $W_{j-1}$  and  $W_j$ . After  $j-1$  draws, the number of white balls in the urn is  $W_{j-1}$ , and the number of white balls will increase by 1 after one drawing, if a red ball is picked, but will decrease by 1, if a white ball is picked. And so,

$$W_j = W_{j-1} + I_n^R - I_n^W = W_{j-1} + 1 - 2I_n^W. \quad (1)$$

A recurrence for the mean follows from the expectation of the stochastic recurrence (1) and a recurrence for the variance follows from expectation of its square. Solving these recurrences we obtain

$$\mathbf{E}[W_j] = \frac{1}{2}n + \left(W_0(n) - \frac{n}{2}\right) \left(\frac{n-2}{n}\right)^j; \quad (2)$$

$$\begin{aligned} \mathbf{Var}[W_j] &= \frac{1}{4}n + \left(\left(\frac{n}{2} - W_0(n)\right)^2 - \frac{n}{4}\right) \left(\frac{n-4}{n}\right)^j \\ &\quad - \left(\frac{n}{2} - W_0(n)\right)^2 \left(\frac{n-2}{n}\right)^{2j}. \end{aligned} \quad (3)$$

(The special case of  $W_0(2n) = n$  is developed and solved in Antognini, 2005.)

Note that under the assumption that  $W_0(n) = \lfloor \alpha n \rfloor \sim \alpha n$ , the mean after  $k_n$  diffusion steps experiences phases according to how fast  $k_n$  grows. For the growing sublinear, linear and superlinear phases we have the mean asymptotics

$$\mathbf{E}[W_{k_n}] \sim \begin{cases} \alpha n, & \text{for } k_n = o(n); \\ \left( \left( \alpha - \frac{1}{2} \right) e^{-2\lambda_n} + \frac{1}{2} \right) n, & \text{for } k_n = \lambda_n n; \\ \frac{1}{2} n, & \text{for } n = o(k_n). \end{cases}$$

Like the mean, under the assumption that  $W_0(n) = \lfloor \alpha n \rfloor \sim \alpha n$ , the variance of the amount of gas in Chamber A after  $k_n$  diffusion steps experiences phases according to how fast  $k_n$  grows. For the growing sublinear, linear and superlinear phases we have the variance asymptotics

$$\mathbf{Var}[W_{k_n}] \sim \begin{cases} 4\alpha(1 - \alpha) k_n, & \text{for } k_n = o(n); \\ \frac{e^{4\lambda_n} - 1 - 4\lambda_n(2\alpha - 1)^2}{4e^{4\lambda_n}} n, & \text{for } k_n = \lambda_n n; \\ \frac{1}{4} n, & \text{for } n = o(k_n). \end{cases}$$

Observe how the average and variance of the three phases meet at the seam lines. The linear phase with  $\lambda_n = 0$  gives the result in the growing sublinear phase, and with  $\lambda_n = \infty$  gives the result of the superlinear phase.

## 5 A martingale underlying gas diffusion

Conditioning the recurrence (1) on the content of the sigma field  $\mathcal{F}_{j-1}$ , one gets

$$\mathbf{E}[W_j | \mathcal{F}_{j-1}] = \left(1 - \frac{2}{n}\right) W_{j-1} + 1. \quad (4)$$

There is an associated martingale as in the following lemma.

**Lemma 1** For  $j = 0, 1, \dots$ ,

$$M_j := \rho_n^j W_j - \frac{\rho_n^{j+1} - \rho_n}{\rho_n - 1}$$



is a martingale, where  $\rho_n = n/(n-2)$ .

*Proof.* Introduce the transformation

$$M_j = a_j W_j + b_j.$$

We wish to turn  $M_j$  into a martingale with suitable choices of deterministic sequences  $a_j$  and  $b_j$ . So,  $M_j$  must satisfy

$$\mathbf{E}[M_j | \mathcal{F}_{j-1}] = M_{j-1} = a_{j-1} W_{j-1} + b_{j-1}. \quad (5)$$

We compute

$$\begin{aligned} \mathbf{E}[M_j | \mathcal{F}_{j-1}] &= \mathbf{E}[a_j W_j + b_j | \mathcal{F}_{j-1}] \\ &= a_j \mathbf{E}[W_j | \mathcal{F}_{j-1}] + b_j. \end{aligned}$$

From (4) we proceed with

$$\mathbf{E}[M_j | \mathcal{F}_{j-1}] = a_j \left(1 - \frac{2}{n}\right) W_{j-1} + a_j + b_j.$$

Matching the coefficients of this equality with those in (5), we arrive at recurrences for  $a_j$  and  $b_j$ . We have  $a_j = \rho_n a_{j-1}$ . This recurrence unfolds easily to give  $a_j = \rho_n^j a_0$ , for any arbitrary value of  $a_0$ ; we take  $a_0 = 1$ .

We also have the recurrence  $b_j = b_{j-1} - a_j$ , which unwinds into

$$b_j = b_0 - \sum_{k=1}^j \rho_n^k,$$

for arbitrary  $b_0$ ; we take  $b_0 = 0$  and simplify the sum to

$$b_j = -\frac{\rho_n^{j+1} - \rho_n}{\rho_n - 1}. \quad \square$$

The fact that  $M_j$  is a martingale is key to proving Gaussian limits in all the phases. We shall deal with the centered martingale

$$\tilde{M}_j = M_j - W_0(n)$$

(which has mean 0) to employ the martingale central limit theorem, which requires calculations on a zero-mean martingale. Sufficient conditions for the central limit theorem for a zero-mean martingale  $X_{j,n}$  are conditional Lindeberg's condition and the conditional variance condition on the martingale differences  $\nabla X_{j,k_n} = X_{j,k_n} - X_{j-1,k_n}$ ; (see Theorem 3.2 and Corollary 3.1, P. 58 in Hall and Heyde, 1980).

Specifically in our case, conditional Lindeberg's condition requires that, for some positive increasing sequence  $\xi_n$ , and for all  $\varepsilon > 0$ ,

$$U_n := \sum_{j=1}^{k_n} \mathbf{E} \left[ \left( \frac{\nabla \tilde{M}_j}{\xi_n} \right)^2 \mathbf{1}_{\left\{ \left| \frac{\nabla \tilde{M}_j}{\xi_n} \right| > \varepsilon \right\}} \mid \mathcal{F}_{j-1} \right] \xrightarrow{\mathcal{P}} 0,$$

where the indicator  $\mathbf{1}_{\mathcal{E}}$  is a function of a sample space that assumes the value 1 if  $\mathcal{E}$  occurs, and otherwise it assumes the value 0, and, for a constant  $c$ , a  $c$ -conditional variance condition requires that

$$V_n := \sum_{j=1}^{k_n} \mathbf{E} \left[ \left( \frac{\nabla \tilde{M}_j}{\xi_n} \right)^2 \mid \mathcal{F}_{j-1} \right] \xrightarrow{\mathcal{P}} c. \quad (6)$$

When both conditions hold, the sum  $\sum_{j=1}^{k_n} \nabla \tilde{M}_j / \xi_n = (M_{k_n} - M_0) / \xi_n = (M_{k_n} - W_0(n)) / \xi_n$  converges to the normally distributed random variable  $\mathcal{N}(0, c^2)$ .

In all the phases we take  $\xi_n = \rho_n^{k_n} \sqrt{\mathbf{Var}[W_{k_n}]}$ . For calculations involved in conditional Lindeberg's condition the following uniform bound is helpful in all the phases.

**Lemma 2**

$$\left| \frac{\nabla \tilde{M}_j}{\rho_n^j} \right| \leq 4.$$

*Proof.* With the help of (1) write the absolute differences as

$$\begin{aligned} |\nabla \tilde{M}_j| &= |(M_j - W_0(n)) - (M_{j-1} - W_0(n))| \\ &= |(\rho_n^j W_j + b_j) - (\rho_n^{j-1} W_{j-1} + b_{j-1})| \\ &= |(\rho_n^j (W_{j-1} + I_j^R - I_j^W) + b_j) - (\rho_n^{j-1} W_{j-1} + b_{j-1})| \\ &\leq \rho_n^{j-1} ((\rho_n - 1) W_{j-1} + \rho_n |I_j^R - I_j^W| + \rho_n) \\ &\leq \rho_n^{j-1} \left( \frac{2}{n-2} W_{j-1} + \frac{2n}{n-2} \right). \end{aligned}$$

The number of white balls at any stage is at most  $n$ , and the lemma follows.  $\square$

Lemma 2 enables us to verify conditional Lindeberg's condition in all the phases.

**Lemma 3**

$$U_n = \sum_{j=1}^{k_n} \mathbf{E} \left[ \left( \frac{\nabla \tilde{M}_j}{\rho_n^j \sqrt{\mathbf{Var}[W_{k_n}]}} \right)^2 \mathbf{1}_{\left\{ \left| \frac{\nabla \tilde{M}_j}{\rho_n^j \sqrt{\mathbf{Var}[W_{k_n}]}} \right| > \varepsilon \right\}} \mid \mathcal{F}_{j-1} \right] \xrightarrow{\mathcal{P}} 0.$$

*Proof.* In all the growing phases, the variance grows with  $n$ . Therefore, for any given  $\varepsilon > 0$ , the uniform bound in Lemma 2 asserts that the sets  $\{|\nabla \tilde{M}_j| > \varepsilon \rho_n^j \sqrt{\mathbf{Var}[W_{k_n}]}\}$  are all empty, for all  $n$  greater than some positive integer  $n_0(\varepsilon)$ . For large  $n$  we have

$$\begin{aligned} U_n &= \sum_{j=1}^{n_0(\varepsilon)} \mathbf{E} \left[ \left( \frac{\nabla \tilde{M}_j}{\rho_n^j \sqrt{\mathbf{Var}[W_{k_n}]}} \right)^2 \mathbf{1}_{\left\{ \left| \frac{\nabla \tilde{M}_j}{\rho_n^j \sqrt{\mathbf{Var}[W_{k_n}]}} \right| > \varepsilon \right\}} \mid \mathcal{F}_{j-1} \right] \\ &\leq \frac{1}{\mathbf{Var}[W_{k_n}]} \sum_{j=1}^{n_0(\varepsilon)} \mathbf{E} \left[ \left( \frac{\nabla \tilde{M}_j}{\rho_n^j} \right)^2 \mid \mathcal{F}_{j-1} \right] \\ &\leq \frac{16n_0(\varepsilon)}{\mathbf{Var}[W_{k_n}]} \\ &\rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad \square \end{aligned}$$

For calculations involved in conditional Lindeberg's condition we need  $\mathbf{E}[(\nabla \tilde{M}_j)^2 \mid \mathcal{F}_{j-1}]$  (see the definition of  $V_n$  in (6)), which is

$$\begin{aligned} \mathbf{E}[(\nabla \tilde{M}_j)^2 \mid \mathcal{F}_{j-1}] &= \mathbf{E}[M_j^2 \mid \mathcal{F}_{j-1}] - M_{j-1}^2 \\ &= \mathbf{E}[(\rho_n^j W_j + b_j)^2 \mid \mathcal{F}_{j-1}] - (\rho_n^{j-1} W_{j-1} + b_{j-1})^2 \\ &= \mathbf{E}[(\rho_n^j (W_{j-1} + 1 - 2I_j^W) + b_j)^2 \mid \mathcal{F}_{j-1}] \\ &\quad - (\rho_n^{j-1} W_{j-1} + b_{j-1})^2. \end{aligned}$$

After some laborious but straightforward calculation involving (1) we get

$$\begin{aligned} \mathbf{E}\left[(\nabla \tilde{M}_j)^2 \mid \mathcal{F}_{j-1}\right] &= \left(\rho_n^{2j} - \rho_n^{2j-2} - \frac{4}{n}\rho_n^{2j}\right)W_{j-1}^2 \\ &\quad + \left(2\rho_n^{2j} + 2b_j\rho_n^j - \frac{4}{n}b_j\rho_n^j - 2\rho_n^{j-1}b_{j-1}\right)W_{j-1} \\ &\quad + 2b_j\rho_n^j + \rho_n^{2j} + b_j^2 - b_{j-1}^2. \end{aligned}$$

This further simplifies to

$$\mathbf{E}\left[(\nabla \tilde{M}_j)^2 \mid \mathcal{F}_{j-1}\right] = -\frac{4}{n^2}\rho_n^{2j}W_{j-1}^2 + \frac{4}{n}\rho_n^{2j}W_{j-1}.$$

Summing we reconstruct  $V_n$  as

$$V_n = \frac{1}{\rho_n^{2k_n} \mathbf{Var}[W_{k_n}]} \left( -\frac{4}{n^2} \sum_{j=1}^{k_n} \rho_n^{2j} W_{j-1}^2 + \frac{4}{n} \sum_{j=1}^{k_n} \rho_n^{2j} W_{j-1} \right). \quad (7)$$

## 6 Phases during long-term drawing

Suppose the gas diffusion process is perpetuated indefinitely. We shall see that as the ball drawing continues from the Ehrenfest urn the process experiences different phases.

### 6.1 The growing sublinear phase

Let  $k_n$  be in the growing sublinear phase ( $k_n$  grows to  $\infty$ , and  $k_n = o(n)$ ). The number of white balls after  $0 \leq j \leq k_n$  draws has obvious bounds—if all the draws are of red balls an increase by  $j$  goes in favor of the number of white balls over their initial number, and if all the draws are of white balls, a deficit of  $j$  occurs against the initial number of white balls. We have the inequalities

$$W_0(n) - j \leq W_j \leq W_0(n) + j.$$

We can ascertain that

$$W_j = \alpha n + O(k_n), \quad (8)$$

for all  $0 \leq j \leq k_n$ .

*Proof of Theorem 1 in the sublinear phase.* In Lemma 3 conditional Lindeberg's condition has been verified throughout the growing sublinear phase. The proof will be complete if we show that  $V_n$  converges to a constant in probability.

In (7) replace  $W_{j-1}$  by the asymptotic equivalent in (8) to get

$$\begin{aligned} V_n &= \frac{1}{\rho_n^{2k_n} \mathbf{Var}[W_{k_n}]} \left( -\frac{4}{n^2} \sum_{j=1}^{k_n} \rho_n^{2j} (\alpha n + O(k_n))^2 \right. \\ &\quad \left. + \frac{4}{n} \sum_{j=1}^{k_n} \rho_n^{2j} (\alpha n + O(k_n)) \right) \\ &= \frac{1}{4\alpha(1-\alpha)k_n(1+o(1))} \left( 4\alpha(1-\alpha) + O\left(\frac{k_n}{n}\right) + O\left(\frac{k_n^2}{n^2}\right) \right) \sum_{j=1}^{k_n} \rho_n^{2j}. \end{aligned}$$

Recall that  $\rho_n = n/(n-2)$ , and we can bound the remaining sum asymptotically:

$$k_n \leq \sum_{j=1}^{k_n} \rho_n^{2j} \leq k_n \rho_n^{2k_n} = k_n(1+o(1)).$$

And so, we have

$$\begin{aligned} V_n &= \frac{1}{k_n(1+o(1))} \left( 1 + O\left(\frac{k_n}{n}\right) + O\left(\frac{k_n^2}{n^2}\right) \right) (k_n + o(k_n)) \\ &\rightarrow 1. \end{aligned}$$

The 1–conditional variance condition has been verified in the growing sublinear phase.

With both conditions checked, the martingale central limit theorem gives

$$\sum_{j=1}^{k_n} \frac{\nabla \tilde{M}_j}{\rho_n^{k_n} \sqrt{\mathbf{Var}[W_{k_n}]}} = \frac{M_{k_n} - W_0(n)}{\rho_n^{k_n} \sqrt{\mathbf{Var}[W_{k_n}]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Subsequently, we write

$$\frac{\rho_n^{k_n} W_{k_n} - \frac{\rho_n^{k_n+1} - \rho_n}{\rho_n - 1} - W_0(n)}{\rho_n^{k_n} \sqrt{\mathbf{Var}[W_{k_n}]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

which after reorganization is the statement of the theorem.  $\square$

## 6.2 The linear phase

In the linear phase  $k_n = \lambda_n n$ , for some  $\lambda_n > 0$  of a magnitude uniformly bounded from above and below, that is, for two positive constants,  $S_1$  and  $S_2$ , and for all  $n$ ,

$$S_1 \leq \lambda_n \leq S_2.$$

At this phase of the gas diffusion we have the following asymptotic equivalents (as  $n \rightarrow \infty$ ), following from (2) and (3):

$$\mathbf{E}[W_{k_n}] = \mu_n n + o(n), \quad (9)$$

and

$$\mathbf{Var}[W_{k_n}] = v_n n + o(n), \quad (10)$$

where

$$\mu_n = \left(\alpha - \frac{1}{2}\right) e^{-2\lambda_n} + \frac{1}{2},$$

and

$$v_n = \frac{e^{4\lambda_n} - 1 - 4\lambda_n(1 - 2\alpha)^2}{4e^{4\lambda_n}} = O(1).$$

We start with a first-order result for  $W_{k_n}$ .

**Theorem 2** *For  $k_n = \lambda_n n$ , for some  $\lambda_n > 0$  of a magnitude separated from zero and infinity,*

$$\frac{W_{k_n}}{\left(\left(\alpha - \frac{1}{2}\right) e^{-2\lambda_n} + \frac{1}{2}\right) n} \xrightarrow{\mathcal{P}} 1.$$

*Proof.* By Chebyshev's inequality,

$$\begin{aligned} \mathbf{Prob}(|W_{k_n} - \mathbf{E}[W_{k_n}]| \geq \varepsilon \mathbf{E}[W_{k_n}]) &\leq \frac{\mathbf{Var}[W_{k_n}]}{\varepsilon^2 (\mathbf{E}[W_{k_n}])^2} \\ &\sim \frac{v_n n}{\varepsilon^2 \mu_n^2 n^2} \\ &\rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Hence,

$$\frac{W_{k_n}}{\mathbf{E}[W_{k_n}]} \xrightarrow{\mathcal{P}} 1.$$

From the convergence  $\mathbf{E}[W_{k_n}]/(\mu_n n) \rightarrow 1$ , and Slutsky's Theorem in its multiplicative form (cf. Karr, 1993, P. 147), we obtain

$$\frac{W_{k_n}}{\mu_n n} \xrightarrow{\mathcal{P}} 1. \quad \square$$

Before we dwell on the proof of a central limit theorem for the amount of gas in Chamber A by the end of some linear phase, we need a technical lemma, which shows that  $W_{k_n}$  grows linearly with  $n$ , like its mean, with correction terms that are  $o_{\mathcal{L}_1}(n)$ . The purpose of this calculation is for later summation to verify conditional Lindeberg's condition.

**Lemma 4** *Let  $W_{k_n}$  be the number of white balls in the urn after  $k_n$  draws, where  $k_n = \lambda_n n$ , for some  $\lambda_n$ , such that  $0 < S_1 \leq \lambda_n \leq S_2 < \infty$ . Then*

$$W_{k_n} = \mu_n n + o_{\mathcal{L}_1}(n),$$

and

$$W_{k_n}^2 = \mu_n^2 n^2 + o_{\mathcal{L}_1}(n^2),$$

*Proof.* From the asymptotics of the mean and variance, as given in (9) and (10), for large  $n$  we have

$$\begin{aligned} \mathbf{E}[(W_{k_n} - \mu_n n)^2] &= \mathbf{Var}[W_{k_n}] + (\mathbf{E}[W_{k_n}] - \mu_n n)^2 \\ &= v_n n + o(n^2) \\ &= o(n^2). \end{aligned} \tag{11}$$

So, by Jensen's inequality

$$\mathbf{E}[|W_{k_n} - \mu_n n|] \leq \sqrt{\mathbf{E}[(W_{k_n} - \mu_n n)^2]} = o(n),$$

which implies

$$W_{k_n} = \mu_n n + o_{\mathcal{L}_1}(n).$$

Moreover, by the Cauchy-Schwarz inequality we have

$$\begin{aligned} \mathbf{E}[|W_{k_n}^2 - \mu_n^2 n^2|] &= \mathbf{E}[|W_{k_n} + \mu_n n| |W_{k_n} - \mu_n n|] \\ &\leq \sqrt{\mathbf{E}[(W_{k_n} + \mu_n n)^2] \mathbf{E}[(W_{k_n} - \mu_n n)^2]}. \end{aligned}$$

Obviously,  $W_{k_n} + \mu_n n \leq n + e^{-2S_1} n + o(n) = O(n)$ . We employ (11) to bound  $\sqrt{\mathbf{E}[(W_{k_n} - \mu_n n)^2]}$  by  $o(n)$ . Subsequently, we obtain

$$\mathbf{E}\left[|W_{k_n}^2 - \mu_n^2 n^2|\right] = o(n^2),$$

which, according to our definition, is the second statement of the lemma.  $\square$

*Proof of Theorem 1 in the linear phase.* In Lemma 3 conditional Lindeberg's condition has been verified throughout the linear phase. It remains to verify the conditional variance condition.

Recall the expressions for  $V_n$  (cf. (6)). In this phase the asymptotic equivalents in Lemma 4 apply only in the linear phase. However, before the linear phase the obvious bound  $n$  on  $W_{j-1}$  is sufficient for our purpose. A precise execution of this line to extract the asymptotics is in Appendix A, where it is shown that  $V_n \xrightarrow{\mathcal{P}} 1$ . The 1–conditional variance condition has been verified in the linear phase.

According to the martingale central limit theorem

$$\sum_{j=1}^{k_n} \frac{\nabla \tilde{M}_j}{\rho_n^{k_n} \sqrt{\mathbf{Var}[W_{k_n}]}} = \frac{M_{k_n} - M_0}{\rho_n^{k_n} \sqrt{\mathbf{Var}[W_{k_n}]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Subsequently, we write

$$\frac{\rho_n^{k_n} W_{k_n} - \frac{\rho_n^{k_n+1} - \rho_n}{\rho_n - 1} - W_0(n)}{\rho_n^{k_n} \sqrt{\mathbf{Var}[W_{k_n}]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \quad \square$$

### 6.3 The superlinear phase

Suppose that the gas diffusion continued for a long period of time. As seen from the behavior of the average, the initial conditions are attenuated through the linear phase and the fixed average component  $\frac{1}{2}n$  becomes more pronounced and eventually dominates in the superlinear phase. Many of the principles of the proof for the linear phase apply within the superlinear phase, so we shall be a bit brief in presenting an adjustment of these proofs. For instance, via the asymptotic equivalents of the mean and variance in the



superlinear phase, we can mimic the proof of Theorem 2, and get a similar result. Namely, when  $n = o(k_n)$ , we have

$$\frac{W_{k_n}}{n} \xrightarrow{\mathcal{P}} \frac{1}{2}.$$

Also, in view of the mean and variance asymptotics we can replicate the result of Lemma 4. We only need to replace  $\mu_n$  by  $\frac{1}{2}$ , and the proof goes through verbatim to obtain

$$W_{k_n} = \frac{1}{2}n + o_{\mathcal{L}_1}(n),$$

and

$$W_{k_n}^2 = \frac{1}{4}n^2 + o_{\mathcal{L}_1}(n^2).$$

*Proof of Theorem 1 in the superlinear phase.* In Lemma 3 conditional Lindeberg's condition has been verified throughout the superlinear phase. The 1-conditional variance condition is checked in Appendix B.

According to the martingale central limit theorem

$$\frac{\rho_n^{k_n} W_{k_n} - \frac{\rho_n^{k_n+1} - \rho_n}{\rho_n - 1} - W_0(n)}{\rho_n^{k_n} \sqrt{\mathbf{Var}[W_{k_n}]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \quad \square$$

## Acknowledgment

The authors wish to thank the Institute for Integrating Statistics in Decision Sciences at The George Washington University for supporting this research. They would like to also thank an anonymous referee for comments that improved the exposition.

## Appendix A

### Verification of the conditional variance in the linear phase

To asymptotically handle the sums in conditional Lindeberg's condition (going over the range of indexes 1 to  $k_n = \lambda_n n$ ) let us break them up at some point near the beginning of the linear phase. Choose a small positive  $\epsilon < S_1$  and break up the sums in  $V_n$  into sums going from 1 to  $\lfloor \epsilon n \rfloor - 1$  and sums starting at  $\lfloor \epsilon n \rfloor$  and ending at  $k_n$ . Applying the asymptotics of Lemma 4, we write (7) in the form

$$\begin{aligned} V_n &= \frac{1}{\rho_n^{2k_n}(v_n n + o(n))} \left( -\frac{4}{n^2} \sum_{j=1}^{\lfloor \epsilon n \rfloor - 1} \rho_n^{2j} W_{j-1}^2 + \frac{4}{n} \sum_{j=1}^{\lfloor \epsilon n \rfloor - 1} \rho_n^{2j} W_{j-1} \right. \\ &\quad - \frac{4}{n^2} \sum_{j=\lfloor \epsilon n \rfloor}^{k_n} \rho_n^{2j} \left( \left( \left( \alpha - \frac{1}{2} \right) e^{-2j/n+o(1)} + \frac{1}{2} \right)^2 n^2 + o_{\mathcal{L}_1}(n^2) \right) \\ &\quad \left. + \frac{4}{n \rho_n^{2k_n}} \sum_{j=\lfloor \epsilon n \rfloor}^{k_n} \rho_n^{2j} \left( \left( \left( \alpha - \frac{1}{2} \right) e^{-2j/n+o(1)} + \frac{1}{2} \right) n + o_{\mathcal{L}_1}(n) \right) \right) \\ &= \frac{1}{1 + o(1)} (C_n + C'_n + D_n + H_n), \end{aligned}$$

where

$$\begin{aligned} C_n &:= -\frac{4}{n^3 v_n e^{4\lambda_n}} \sum_{j=1}^{\lfloor \epsilon n \rfloor - 1} \rho_n^{2j} W_{j-1}^2, \\ C'_n &:= \frac{4}{n^2 v_n e^{4\lambda_n}} \sum_{j=1}^{\lfloor \epsilon n \rfloor - 1} \rho_n^{2j} W_{j-1}, \\ D_n &:= -\frac{4}{n^3 v_n e^{4\lambda_n}} \sum_{j=\lfloor \epsilon n \rfloor}^{k_n} \rho_n^{2j} \left( \left( \left( \alpha - \frac{1}{2} \right) e^{-2j/n+o(1)} + \frac{1}{2} \right)^2 n^2 + o_{\mathcal{L}_1}(n^2) \right), \end{aligned}$$

and

$$H_n := \frac{4}{n^2 v_n e^{4\lambda_n}} \sum_{j=\lfloor \epsilon n \rfloor}^{k_n} \rho_n^{2j} \left( \left( \left( \alpha - \frac{1}{2} \right) e^{-2j/n+o(1)} + \frac{1}{2} \right) n + o_{\mathcal{L}_1}(n) \right).$$

For large  $n$ , we have

$$|C_n| \leq \frac{4}{n^3 v_n e^{4\lambda_n}} \sum_{j=1}^{\lfloor \epsilon n \rfloor - 1} \rho_n^{2j} n^2$$

$$\begin{aligned}
&\leq \frac{4}{nv_n e^{4\lambda_n}} \sum_{j=1}^{\lfloor \varepsilon n \rfloor} 2e^{4S_2} \\
&= O(\varepsilon), \quad \text{as } \varepsilon \rightarrow 0.
\end{aligned}$$

Likewise, we have

$$|C'_n| = O(\varepsilon), \quad \text{as } \varepsilon \rightarrow 0.$$

The formula for  $D_n$  reduces to

$$\begin{aligned}
D_n &= -\frac{4}{nv_n e^{4\lambda_n}} \sum_{j=\lfloor \varepsilon n \rfloor}^{k_n} \rho_n^{2j} \left( \left( \alpha - \frac{1}{2} \right) e^{-2j/n+o(1)} + \frac{1}{2} \right)^2 + o_{\mathcal{L}_1}(1) \\
&= -\frac{4}{nv_n e^{4\lambda_n}} \sum_{j=\lfloor \varepsilon n \rfloor}^{k_n} \rho_n^{2j} \left( \left( \alpha - \frac{1}{2} \right)^2 e^{-4j/n} + \left( \alpha - \frac{1}{2} \right) e^{-2j/n} + \frac{1}{4} \right) \\
&\quad + o_{\mathcal{L}_1}(1) \\
&= -\frac{4}{nv_n e^{4\lambda_n}} \left( \left( \alpha - \frac{1}{2} \right)^2 \left( \sum_{j=0}^{k_n} \rho_n^{2j} e^{-4j/n} - \sum_{j=0}^{\lfloor \varepsilon n \rfloor - 1} \rho_n^{2j} e^{-4j/n} \right) \right. \\
&\quad \left. + \left( \alpha - \frac{1}{2} \right) \left( \sum_{j=0}^{k_n} \rho_n^{2j} e^{-2j/n} - \sum_{j=0}^{\lfloor \varepsilon n \rfloor - 1} \rho_n^{2j} e^{-2j/n} \right) \right. \\
&\quad \left. + \frac{1}{4} \left( \sum_{j=0}^{k_n} \rho_n^{2j} - \sum_{j=0}^{\lfloor \varepsilon n \rfloor - 1} \rho_n^{2j} \right) \right) \\
&\quad + o_{\mathcal{L}_1}(1) \sum_{j=\lfloor \varepsilon n \rfloor}^{k_n} \rho_n^{2j}.
\end{aligned}$$

This calculation involves several sums of the form

$$\sum_{j=0}^{b_n-1} \rho_n^{2j} e^{-\frac{\gamma j}{n}} = \frac{\left( \frac{n}{n-2} \right)^{2b_n} e^{-\frac{\gamma b_n}{n}} - 1}{\left( \frac{n}{n-2} \right)^2 e^{-\frac{\gamma}{n}} - 1},$$

with  $b_n = \beta_n n + r_n$ , and the remainder function  $r_n$  is  $o(n)$ . Using the asymptotic relation  $(n/(n-2))^{2\beta_n n} = e^{4\beta_n} + 4\beta_n e^{4\beta_n}/n + O(1/n^2)$ , and the standard local expansion

$$e^{\frac{c}{n}} = 1 + \frac{c}{n} + \frac{c^2}{2n^2} + O\left(\frac{1}{n^3}\right),$$

we get

$$\begin{aligned}
\sum_{j=0}^{b_n-1} \rho_n^{2j} e^{-\frac{\gamma j}{n}} &= \frac{\left( \left( e^{4\beta_n} + \frac{4\beta_n e^{4\beta_n}}{n} + O\left(\frac{1}{n^2}\right) \right) \left( \frac{n}{n-2} \right)^{2r_n} e^{-\frac{\gamma\beta_n n + \gamma r_n}{n}} - 1 \right)}{\left( (4-\gamma)n + \left(\frac{\gamma^2}{2} - 4\right) + O\left(\frac{1}{n}\right) \right)} \\
&\quad \times (n-2)^2 \\
&= \frac{e^{(4-\gamma)\beta_n} \left( 1 + \frac{4\beta_n}{n} + O\left(\frac{1}{n^2}\right) \right) e^{2r_n \left( \frac{2}{n} + O\left(\frac{1}{n^2}\right) \right)} e^{-\frac{\gamma r_n}{n}} - 1}{\left( (4-\gamma)n + \left(\frac{\gamma^2}{2} - 4\right) + O\left(\frac{1}{n}\right) \right)} (n-2)^2 \\
&= \frac{e^{(4-\gamma)\beta_n} \left( 1 + \frac{4\beta_n}{n} + O\left(\frac{1}{n^2}\right) \right) \left( 1 + O\left(\frac{r_n}{n}\right) \right) - 1}{\left( (4-\gamma)n + \left(\frac{\gamma^2}{2} - 4\right) + O\left(\frac{1}{n}\right) \right)} (n-2)^2 \\
&= \begin{cases} \frac{e^{(4-\gamma)\beta_n} - 1}{4-\gamma} n + o(n), & \text{if } \gamma \neq 4; \\ \beta_n n, & \text{otherwise.} \end{cases}
\end{aligned}$$

Applying these formulas with  $\gamma = 4, 2, 0$ , we have

$$\begin{aligned}
D_n &= -\frac{1}{v_n e^{4\lambda_n}} \left[ 4 \left( \alpha - \frac{1}{2} \right)^2 \lambda_n + 2 \left( \alpha - \frac{1}{2} \right) (e^{2\lambda_n} - 1) + \frac{1}{4} (e^{4\lambda_n} - 1) \right] \\
&\quad + O(\varepsilon) + o(1) + o_{\mathcal{L}_1}(1).
\end{aligned}$$

Similarly, we have

$$H_n = \frac{1}{v_n e^{4\lambda_n}} \left( 2 \left( \alpha - \frac{1}{2} \right) (e^{2\lambda_n} - 1) + \frac{1}{2} (e^{4\lambda_n} - 1) \right) + O(\varepsilon) + o(1) + o_{\mathcal{L}_1}(1).$$

Consequently, we have

$$V_n = \frac{1}{1 + o(1)} (O(\varepsilon) + 1 + o(1) + o_{\mathcal{L}_1}(1)).$$

Taking the limit, as  $\varepsilon \rightarrow 0$ , we get

$$\lim_{\varepsilon \rightarrow 0} V_n = V_n = \frac{1}{1 + o(1)} (1 + o(1) + o_{\mathcal{L}_1}(1)).$$

Now, let  $n \rightarrow \infty$  to get

$$V_n \xrightarrow{\mathcal{P}} 1.$$

## Appendix B

### Verification of the conditional variance in the superlinear phase

For the sum in the conditional variance condition we apply the bound  $W_{j-1} \leq n$  until the superlinear phase. More precisely, to asymptotically handle the sums in conditional Lindeberg's condition (going over the range of indices 1 to  $k_n$ ) we break up the sums in  $V_n$  into sums going from 1 to  $k'_n - 1$ , which is any superlinear function of order less than  $k_n$  (giving ignorable contribution) and sums starting at  $k'_n$  and ending at  $k_n$  (most of the contribution comes near  $k_n$ ). We can take  $k'_n = \lfloor k_n / \ln(k_n/n) \rfloor$ . Then

$$\begin{aligned} V_n &= \frac{1}{\rho_n^{2k_n} \mathbf{Var}[W_{k_n}]} \left( -\frac{4}{n^2} \sum_{j=1}^{k'_n-1} \rho_n^{2j} W_{j-1}^2 + \frac{4}{n} \sum_{j=1}^{k'_n-1} \rho_n^{2j} W_{j-1} \right. \\ &\quad \left. - \frac{4}{n^2} \sum_{j=k'_n}^{k_n} \rho_n^{2j} \left( \frac{n^2}{4} + o_{\mathcal{L}_1}(n^2) \right) \right. \\ &\quad \left. + \frac{4}{n} \sum_{j=k'_n}^{k_n} \rho_n^{2j} \left( \left( \frac{n}{2} + o_{\mathcal{L}_1}(n) \right) \right) \right) \\ &= \frac{1}{1+o(1)} (\tilde{C}_n + \tilde{D}_n + \tilde{H}_n), \end{aligned}$$

where

$$\begin{aligned} \tilde{C}_n &:= -\frac{16}{n^3 \rho_n^{2k_n}} \sum_{j=1}^{k'_n-1} \rho_n^{2j} W_{j-1}^2 + \frac{16}{n^2 \rho_n^{2k_n}} \sum_{j=1}^{k'_n-1} \rho_n^{2j} W_{j-1}, \\ \tilde{D}_n &:= -\frac{16}{n^3 \rho_n^{2k_n}} \sum_{j=\lfloor \varepsilon n \rfloor}^{k_n} \rho_n^{2j} \left( \frac{n^2}{4} + o_{\mathcal{L}_1}(n^2) \right), \end{aligned}$$

and

$$\tilde{H}_n := \frac{16}{n^2 \rho_n^{2k_n}} \sum_{j=k'_n}^{k_n} \rho_n^{2j} \left( \frac{n}{2} + o_{\mathcal{L}_1}(n) \right).$$

We have

$$\begin{aligned} |\tilde{C}_n| &\leq \frac{32}{n\rho_n^{2k_n}} \sum_{j=1}^{k'_n-1} \rho_n^{2j} \\ &= O\left(\rho_n^{2k'_n-2k_n}\right). \end{aligned}$$

We also have

$$\tilde{D}_n + \tilde{H}_n = \frac{4}{n\rho_n^{2k_n}} \left( \frac{\rho_n^{2k_n+2} - \rho_n^{k'_n}}{\rho_n^2 - 1} \right) (1 + o_{\mathcal{L}_1}(1)) = 1 + o_{\mathcal{L}_1}(1).$$

Putting the terms together we see that

$$V_n \xrightarrow{\mathcal{P}} 1.$$

## References

- [1] Antognini, B. (2005). On the speed of convergence of some urn designs for the balanced allocation of two treatments. *Metrika*, **62**, 309–322.
- [2] Bellman, R. and Harris, T. (1951). Recurrence times for the Ehrenfest model. *Pacific Journal of Mathematics*, **1**, 184–188.
- [3] Blom, G. (1989). Mean transition times for the Ehrenfest urn model. *Advances in Applied Probability*, **21**, 479–480.
- [4] Diaconis, P. (1996). The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Science, USA*, **93**, 1659–1664.
- [5] Ehrenfest, P. and Ehrenfest, T. (1907). Über zwei bekannte einwände gegen das Boltzmannsche H-theorem. *Physikalische Zeitschrift*, **8**, 311–314.
- [6] Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.
- [7] Karlin, S. and McGregor, J. (1965). Ehrenfest urn models. *Journal of Applied Probability*, **2**, 352–376.

- [8] Karr, A. (1993). *Probability*. Springer, New York.
- [9] Kolchin, V., Sevastyanov, B. and Chistyakov, V. (1976). *Random Allocations*, Moscow.
- [10] Mahmoud, H. (2008). *PólyaUrn Models*. Chapman-Hall, Boca Raton, Florida.
- [11] Mahmoud, H. (2010+). Gaussian phases in generalized coupon collection. (manuscript).
- [12] Mikhailov, V. (1977). A Poisson limit theorem in the scheme of group disposal of particles. *Theory of Probability and Its Applications*, **22**, 152–156.
- [13] Mikhailov, V. (1980). Asymptotic normality of the number of empty cells for group allocation of particles. *Theory of Probability and Its Applications*, **25**, 82–90.
- [14] Vatutin, V. and Mikhailov, V. (1982). Limit theorem for the number of empty cells in an equiprobable scheme for group allocation of particles. *Theory of Probability and Its Applications*, **27**, 734–743.
- [15] Smythe, R. (2009). Phases in generalized coupon collection (private communication).