J. Appl. Prob. **47,** 841–855 (2010)
Printed in England

PHASES IN THE DIFFUSION OF GASES VIA THE EHRENFEST URN MODEL

SRINIVASAN BALAJI,*
HOSAM MAHMOUD * ** AND
ZHANG TONG,* *The George Washington University*

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 each 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.

Keywords: Urn model; random structure; martingale; central limit theorem; diffusion of gases

2010 Mathematics Subject Classification: Primary 60C05; 60F05; 05A05
Secondary 60G42

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 (see 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 chamber. 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

Received 30 November 2009; revision received 27 April 2010.
* Postal address: Department of Statistics, The George Washington University, Washington, DC 20052, USA.
** Email address: hosam@gwu.edu

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), and see Mahmoud (2008, pp. 62–67) for an overview.

Antognini (2005) looked at the speed of diffusion for a fixed number of particles. In a physical system the number of gas molecules is very large, we will 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 $\alpha \in (0, 1)$. We are 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 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 0 and ∞;

(c) the superlinear phase, when $n = o(k_n)$.

We will prove the following general trends. Trivially, at the very low end of the sublinear phase, when $k_n = O(1)$ as $n \to \infty$, 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 ∞. In what follows, the normally distributed random variate with mean 0 and variance ν^2 is denoted by $\mathcal{N}(0, \nu^2)$, and '$\xrightarrow{\mathrm{D}}$' 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 $k_n \to \infty$ in a sublinear, linear, or superlinear fashion. Then,*

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

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 ∞, we see that the amount of gas in each chamber is normally distributed, even for a fairly slow growing function k_n. We call the phase when k_n grows sublinearly to ∞ the *growing sublinear phase*. Functions that are

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

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 α. The Gaussian law in Theorem 1 takes the form

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

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 α 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

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

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 $g_n/n \to \infty$, Theorem 1 takes the symmetric form

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

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) took a different view and discussed 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 that the balls in the Ehrenfest urn are labeled $1, \ldots, n$. We label the urns in the allocation scheme with $1, \ldots, n$, say from left to right, and the ith urn represents the ith ball in the Ehrenfest urn. The urns labeled $1, \ldots, W_0(n)$ are the initial white balls in the Ehrenfest model, and the urns labeled $W_0(n) + 1, \ldots, n$ 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 *et al.* (1976), where the schemes were handled by the method of moments.

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

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, with a subsection devoted to each phase. In these subsections we prove the announced results. Detailed proofs are relegated to Appendices A and B.

Throughout, we will use the following standard probability notation. We will use the symbol '$\xrightarrow{\text{P}}$' to denote 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 $\mathrm{E}[|X_n|]/|g(n)| \to 0$ as $n \to \infty$. We let $\mathcal{F}_j$ be the sigma field generated by the first j draws.

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

We will also need the backward difference operator ∇, 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 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 nth step. Because of their mutual exclusion, we have $I_n^{\mathrm{R}} = 1 - I_n^{\mathrm{W}}$. 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 draw if a red ball is picked, but will decrease by 1 if a white ball is picked. So,

$$W_j = W_{j-1} + I_n^{\mathrm{R}} - I_n^{\mathrm{W}} = W_{j-1} + 1 - 2I_n^{\mathrm{W}}. \tag{1}$$

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

$$\mathrm{E}[W_j] = \frac{1}{2}n + \left(W_0(n) - \frac{n}{2}\right)\left(\frac{n-2}{n}\right)^j, \tag{2}$$

$$\mathrm{var}[W_j] = \frac{1}{4}n + \left(\left(\frac{n}{2} - W_0(n)\right)^2 - \frac{n}{4}\right)\left(\frac{n-4}{n}\right)^j - \left(\frac{n}{2} - W_0(n)\right)^2\left(\frac{n-2}{n}\right)^{2j}. \tag{3}$$

(The special case of $W_0(2n) = n$ was 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

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

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

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

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

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 in the superlinear phase.

5. A martingale underlying gas diffusion

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

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

There is an associated martingale as in the following lemma.

Lemma 1. *For $j = 0, 1, \ldots,$*

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

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

$$\mathrm{E}[M_j \mid \mathcal{F}_{j-1}] = M_{j-1} = a_{j-1} W_{j-1} + b_{j-1}. \tag{5}$$

We compute

$$\mathrm{E}[M_j \mid \mathcal{F}_{j-1}] = \mathrm{E}[a_j W_j + b_j \mid \mathcal{F}_{j-1}] = a_j \,\mathrm{E}[W_j \mid \mathcal{F}_{j-1}] + b_j.$$

From (4) we proceed with

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

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 $a_0 = 1$.

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

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

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

This completes the proof.

The fact that M_j is a martingale is key to proving Gaussian limits in all the phases. We will 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 the conditional Lindeberg 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 of Hall and Heyde (1980, p. 58).

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

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

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

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

When both conditions hold, the sum

$$\sum_{j=1}^{k_n} \frac{\nabla \tilde{M}_j}{\xi_n} = \frac{M_{k_n} - M_0}{\xi_n} = \frac{M_{k_n} - W_0(n)}{\xi_n}$$

converges to the normally distributed random variable $\mathcal{N}(0, c^2)$.

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

Lemma 2. *We have*

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

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

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

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

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

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

Lemma 3. *We have*

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

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 $\{|\nabla\tilde{M}_j| > \varepsilon\rho_n^j\sqrt{\mathrm{var}[W_{k_n}]}\}$ are all empty for all n greater than some positive integer $n_0(\varepsilon)$. For large n, we have

$$\begin{aligned}U_n &= \sum_{j=1}^{n_0(\varepsilon)} \mathrm{E}\left[\left(\frac{\nabla\tilde{M}_j}{\rho_n^j \sqrt{\mathrm{var}[W_{k_n}]}}\right)^2 \mathbf{1}_{\{|\nabla\tilde{M}_j/\rho_n^j\sqrt{\mathrm{var}[W_{k_n}]}|>\varepsilon\}} \;\middle|\; \mathcal{F}_{j-1}\right] \\ &\le \frac{1}{\mathrm{var}[W_{k_n}]} \sum_{j=1}^{n_0(\varepsilon)} \mathrm{E}\left[\left(\frac{\nabla\tilde{M}_j}{\rho_n^j}\right)^2 \;\middle|\; \mathcal{F}_{j-1}\right] \\ &\le \frac{16 n_0(\varepsilon)}{\mathrm{var}[W_{k_n}]} \\ &\to 0 \quad \text{as } n \to \infty.\end{aligned}$$

This completes the proof.

For calculations involved in the conditional Lindeberg condition, we need

$$\mathrm{E}[(\nabla\tilde{M}_j)^2 \mid \mathcal{F}_{j-1}]$$

(see the definition of V_n in (6)), which is

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

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

After some laborious but straightforward calculations involving (1) we obtain

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

This further simplifies to

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

Summing over j, we construct V_n as

$$V_n = \frac{1}{\rho_n^{2k_n}\,\mathrm{var}[W_{k_n}]}\left(-\frac{4}{n^2}\sum_{j=1}^{k_n}\rho_n^{2j}W_{j-1}^2 + \frac{4}{n}\sum_{j=1}^{k_n}\rho_n^{2j}W_{j-1}\right). \tag{7}$$

6. Phases during long-term drawing

Suppose that the gas diffusion process is perpetuated indefinitely. We will 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 ∞ and $k_n = o(n)$). The number of white balls after $0 \le j \le k_n$ 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

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

We can ascertain that

$$W_j = \alpha n + O(k_n) \tag{8}$$

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

Proof of Theorem 1 in the sublinear phase. In Lemma 3 the conditional Lindeberg 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 obtain

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

Recall that $\rho_n = n/(n-2)$. We can bound the remaining sum asymptotically:

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

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

So, we have

$$V_n = \frac{1}{k_n(1+o(1))}\left(1 + O\left(\frac{k_n}{n}\right) + O\left(\frac{k_n^2}{n^2}\right)\right)(k_n + o(k_n)) \to 1.$$

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

With both conditions checked, the martingale central limit theorem gives

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

Subsequently, we write

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

which after reorganization is the statement of the theorem.

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 all n,

$$S_1 \le \lambda_n \le S_2.$$

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

$$\mathrm{E}[W_{k_n}] = \mu_n n + o(n) \tag{9}$$

and

$$\mathrm{var}[W_{k_n}] = v_n n + o(n), \tag{10}$$

where

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

and

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

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 0 and ∞,*

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

Proof. By Chebyshev's inequality,

$$\begin{aligned}\mathrm{P}(|W_{k_n} - \mathrm{E}[W_{k_n}]| \ge \varepsilon\,\mathrm{E}[W_{k_n}]) &\le \frac{\mathrm{var}[W_{k_n}]}{\varepsilon^2(\mathrm{E}[W_{k_n}])^2} \\ &\sim \frac{v_n n}{\varepsilon^2 \mu_n^2 n^2} \\ &\to 0 \quad \text{as } n \to \infty.\end{aligned}$$

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

Hence,

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

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

$$\frac{W_{k_n}}{\mu_n n} \xrightarrow{\mathrm{P}} 1.$$

This completes the proof.

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 the conditional Lindeberg 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 λ_n such that $0 < S_1 \le \lambda_n \le S_2 < \infty$. 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

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

So, by Jensen's inequality,

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

which implies that

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

Moreover, by the Cauchy–Schwarz inequality we have

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

Obviously, $W_{k_n} + \mu_n n \le n + \mathrm{e}^{-2S_1} n + o(n) = O(n)$. We employ (11) to bound

$$\sqrt{\mathrm{E}[(W_{k_n} - \mu_n n)^2]}$$

by $o(n)$. Subsequently, we obtain

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

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

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

Proof of Theorem 1 in the linear phase. In Lemma 3 the conditional Lindeberg 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 given in Appendix A, where it is shown that $V_n \xrightarrow{\text{P}} 1$. The 1-conditional variance condition has been verified in the linear phase.

According to the martingale central limit theorem

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

Subsequently, we write

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

This completes the proof.

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 will be 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 obtain a similar result. Namely, when $n = o(k_n)$, we have

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

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

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

and

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

Proof of Theorem 1 in the superlinear phase. In Lemma 3 the conditional Lindeberg 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

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

This completes the proof.

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

Appendix A. Verification of the conditional variance in the linear phase

To asymptotically handle the sums in the conditional Lindeberg condition (going over the range of indices 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 $\lfloor \varepsilon n \rfloor - 1$, and sums starting at $\lfloor \varepsilon n \rfloor$ and ending at k_n. Applying the asymptotics of Lemma 4, we write (7) in the form

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

where

$$C_n := -\frac{4}{n^3 v_n \mathrm{e}^{4\lambda_n}} \sum_{j=1}^{\lfloor \varepsilon n \rfloor - 1} \rho_n^{2j} W_{j-1}^2,$$

$$C_n' := \frac{4}{n^2 v_n \mathrm{e}^{4\lambda_n}} \sum_{j=1}^{\lfloor \varepsilon n \rfloor - 1} \rho_n^{2j} W_{j-1},$$

$$D_n := -\frac{4}{n^3 v_n \mathrm{e}^{4\lambda_n}} \sum_{j=\lfloor \varepsilon n \rfloor}^{k_n} \rho_n^{2j} \left(\left(\left(\alpha - \frac{1}{2} \right) \mathrm{e}^{-2j/n + o(1)} + \frac{1}{2} \right)^2 n^2 + o_{\mathcal{L}_1}(n^2) \right),$$

and

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

For large n, we have

$$|C_n| \le \frac{4}{n^3 v_n \mathrm{e}^{4\lambda_n}} \sum_{j=1}^{\lfloor \varepsilon n \rfloor - 1} \rho_n^{2j} n^2 \le \frac{4}{n v_n \mathrm{e}^{4\lambda_n}} \sum_{j=1}^{\lfloor \varepsilon n \rfloor} 2\mathrm{e}^{4S_2} = O(\varepsilon) \quad \text{as } \varepsilon \to 0.$$

Likewise, we have

$$|C_n'| = O(\varepsilon) \quad \text{as } \varepsilon \to 0.$$

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

The formula for D_n reduces to

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

This calculation involves several sums of the form

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

with $b_n = \beta_n n + r_n$, and the remainder function r_n is $o(n)$. Using the asymptotic relation

$$\left(\frac{n}{n-2} \right)^{2\beta_n n} = e^{4\beta_n} + \frac{4\beta_n e^{4\beta_n}}{n} + O\left(\frac{1}{n^2} \right),$$

and the standard local expansion

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

we obtain

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

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

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

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

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 $\varepsilon \to 0$, we obtain

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

Now, let $n \to \infty$ to obtain

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

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 the conditional Lindeberg 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

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

where

$$\tilde{C}_n := -\frac{16}{n^3 \rho_n^{2k_n}} \sum_{j=1}^{k_n'-1} \rho_n^{2j} W_{j-1}^2 + \frac{16}{n^2 \rho_n^{2k_n}} \sum_{j=1}^{k_n'-1} \rho_n^{2j} W_{j-1},$$

$$\tilde{D}_n := -\frac{16}{n^3 \rho_n^{2k_n}} \sum_{j=\lfloor \varepsilon n \rfloor}^{k_n} \rho_n^{2j}\left(\frac{n^2}{4} + o_{\mathcal{L}_1}(n^2)\right),$$

https://doi.org/10.1239/jap/1285335413 Published online by Cambridge University Press

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

$$|\tilde{C}_n| \leq \frac{32}{n\rho_n^{2k_n}} \sum_{j=1}^{k_n'-1} \rho_n^{2j} = O(\rho_n^{2k_n'-2k_n}).$$

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

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

Acknowledgements

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.

References

Antognini, A. B. (2005). On the speed of convergence of some urn designs for the balanced allocation of two treatments. *Metrika* **62,** 309–322.

Bellman, R. and Harris, T. (1951). Recurrence times for the Ehrenfest model. *Pacific J. Math.* **1,** 179–193.

Blom, G. (1989). Mean transition times for the Ehrenfest urn model. *Adv. Appl. Prob.* **21,** 479–480.

Diaconis, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. USA* **93,** 1659–1664.

Ehrenfest, P. and Ehrenfest, T. (1907). Über zwei bekannte einwände gegen das Boltzmannsche H-theorem. *Phys. Z.* **8,** 311–314.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.

Karlin, S. and McGregor, J. (1965). Ehrenfest urn models. *J. Appl. Prob.* **2,** 352–376.

Karr, A. F. (1993). *Probability*. Springer, New York.

Kolchin, V. F., Sevastyanov, B. A. and Chistyakov, V. P. (1976). *Random Allocations*. Moscow.

Mahmoud, H. M. (2008). *Pólya Urn Models*. CRC Press, Boca Raton, FL.

Mahmoud, H. (2010). Gaussian phases in generalized coupon collection. Unpublished manuscript.

Mikhailov, V. G. (1977). A Poisson limit theorem in the scheme of group disposal of particles. *Theory Prob. Appl.* **22,** 152–156.

Mikhailov, V. G. (1980). Asymptotic normality of the number of empty cells in allocation of particles by groups. *Theory Prob. Appl.* **25,** 82–90.

Smythe, R. (2009). Phases in generalized coupon collection. Personal communication.

Vatutin, V. A. and Mikhailov, V. G. (1982). Limit theorems for the number of empty cells in an equiprobable scheme for the distribution of particles by groups. *Theory Prob. Appl.* **27,** 734–743.