# GENERAL URN MODELS WITH SEVERAL TYPES OF BALLS AND GAUSSIAN LIMITING FIELDS

MICHAEL DRMOTA* +, DANIÈLE GARDY** +, AND BERNHARD GITTENBERGER* +

ABSTRACT. We study a system of $m$ urns, where several types of balls are thrown, and an additive valuation is assigned to each urn depending on its state. Examples are the join models studied in a database context, and some models whith two types of balls. The object of our investigation is the evolution of the valuation with time, when a ball is thrown at each time unit. By means of a generating function approach we show the weak convergence of the valuation to a Gaussian field.

## 1. INTRODUCTION

Our main motivation is the analysis of some random allocation models that have been proposed to study the dynamical behaviour of relational databases. The second author introduced urn models to study the so-called *sizes of relations* obtained by projection or joins [8, 9]. The *projection* model is basically a generalization of the empty-urns model (see [15] for a detailed presentation of this last model, both for the asymptotic distribution and for the limiting process under a large set of assumptions), and we gave in [6] an analysis of the asymptotic process in a restricted dynamic case (balls are added one at a time, no deletions are allowed). The present paper has its origin in the more involved models required to deal with *joins* with insertions, or to allow for deletions in the database. These kinds of combinatorial objects have turned out to be of interest of their own; they can also be applied to completely different fields, e.g. to biological problems etc.[13]

A general formulation of the urn models for joins might be as follows. Consider a sequence of $m$ urns into which we throw different types of balls according to some rules. The balls are thrown one at a time and independently. Moreover, we assume that the balls of one type are undistinguishable. Assign to each urn $U$ containing $k_i$ balls of type $i$, $i = 1, 2, \ldots, d$, an integer valued valuation $f(k_1, k_2, \ldots, k_d) \geq 0$. We are interested in the random variable $X_m$ equal to the sum of all valuations. If we denote by $K_{ij}$ the number of balls of type $i$ in the $j$th urn, then we have

$$X_m = \sum_{j=1}^{m} f(K_{1j}, \ldots, K_{dj}).$$

This formulation allows us to present a unified treatment of several urn models :

*Date*: June 4, 2001.

* Department of Geometry, Technische Universität Wien, Wiedner Hauptstraße 8-10/113, A-1040 Wien, Austria.

** PRISM,CNRS UMR 8636 and Université de Versailles Saint-Quentin, 78035 Versailles Cedex, France.

+ We acknowledge support of the Autrian-French project AMADEUS No 02584QK.

- Semijoin and equijoin models in dynamical databases, where we have two types of balls and the valuation is the join size (see [8, 9] for the definition of the semijoin and of the equijoin, and for their modelization by random allocations of balls into urns) :

$$f(k_1, k_2) = \begin{cases} k_1 \mathbf{1}_{[k_2>0]} & \text{for the semijoin,} \\ k_1 k_2 & \text{for the equijoin.} \end{cases}$$

  A first study of the dynamic behaviour of join models under some assumptions was presented in [11], where each case required an *ad hoc* treatment.
- Urns of balance $q$ : There are again two types of balls; the balance of an urn is the relative difference between the numbers of balls of each type, and the valuation is the number of urns with the specified balance : $f(k_1, k_2) = \mathbf{1}_{[k_1=k_2+q]}$. Such models were introduced in [3] to study the behaviour of a learning process; they also appear in [6]. The model we consider in the present paper differs somewhat, in that here the number of balls of each type is known, whereas the former study assumed that only the total number of balls was known.
- It should be mentioned that the general urn model previously studied by the authors in [6] also fits into this scheme : There we (in most cases) had one type of balls and we counted the number of urns in a certain state $C$. For these urn models the function $f$ can be defined by

$$f(k) = \begin{cases} 1 & \text{if the urn is in state } C \\ 0 & \text{otherwise} \end{cases}$$

We shall prove in this paper that, when the balls are thrown at each unit time, the process associated to the number of urns with a specified number of balls of each type converges weakly towards a Gaussian field, whose covariance function can be explicitly computed.

A second contribution of our paper is the proposal of a model that allows for deletions of balls. Again we obtain the convergence of the relevant parameter towards a Gaussian process.

The plan of the paper is as follows. In Section 2 we present a very general urn model based on a generating function approach and formulate our main result : We prove that, under suitable conditions, the global valuation $X_m$ converges towards a Gaussian field. We study several examples in Section 3 (join and balanced urns models); for example the equijoin leads to variations on the Brownian sheet. Section 4 introduces a model for deletions and validates this approach on an empty-urns model. Finally Section 5 gives the proof of our theorem.

## 2. Convergence to a Gaussian Field

2.1. **Generating Functions for the Motivating Model.** First, let us consider the motivating model described in the Introduction.

We consider $d$ types of balls which are thrown into $m$ urns. Let $g_i(y) = \sum_{k\geq 0} a_{ik} y^k$, $1 \leq i \leq d$, be the generating functions enumerating the allocations of balls of type $i$ into a single urn.[1] Hence, the generating function describing the

[1] We will apply the generating function technique for combinatorial enumeration (for an introduction to this method see e.g. [7, 12]).

allocation of balls in one urn is given by

$$\phi_1(x, z_1, \dots, z_d) = \sum_{k_1,\dots,k_d \ge 0} a_{1k_1} \cdots a_{dk_d} x^{f(k_1,\dots,k_d)} z_1^{k_1} \cdots z_d^{k_d},$$

where $x$ marks the *value* of this urn. Due to the additivity property described in the Introduction, the generating function of the situation of $m$ urns is given by

$$\Phi_1(x, z_1, \dots, z_d) = \phi_1(x, z_1, \dots, z_d)^m.$$

More precisely, let $X_m(n_1, n_2, \dots, n_d)$ denote the (additive) value of these $m$ urns, where $n_i$ balls of type $i$, $1 \le i \le d$, have been thrown, i.e.

$$\mathbf{P}\{X_m(n_1, n_2, \dots, n_d) = k\} = \frac{[x^k z_1^{n_1} \cdots z_d^{n_d}]\Phi_1(x, z_1, \dots, z_d)}{[z_1^{n_1} \cdots z_d^{n_d}]\Phi_1(1, z_1, \dots, z_d)}.$$

If we define

$$\Phi_b(x_1, x_2, \dots, x_b; \mathbf{z_1}, \dots, \mathbf{z_b}) = \phi_b(x_1, x_2, \dots, x_b; \mathbf{z_1}, \dots, \mathbf{z_b})^m$$

with $\mathbf{z_j} = (z_{1j}, \dots, z_{dj})$ and

$$\begin{aligned} &\phi_b(x_1, x_2, \dots, x_b; \mathbf{z_1}, \dots, \mathbf{z_b}) \\ (2.1) \qquad &= \sum_{\substack{k_{ij} \ge 0 \\ 1 \le i \le d, 1 \le j \le b}} \prod_{j=1}^{b} \left( x_j^{f(k_{11}+\cdots+k_{1j}, k_{21}+\cdots+k_{2j}, \dots, k_{d1}+\cdots+k_{dj})} \prod_{i=1}^{d} a_{ik_{ij}} z_{ij}^{k_{ij}} \right) \end{aligned}$$

we also get the *finite dimensional* distribution, i.e.

$$\begin{aligned} &\mathbf{P}\{X_m(\mathbf{n}_1) = k_1, X_m(\mathbf{n}_1 + \mathbf{n}_2) = k_2, \dots, X_m(\mathbf{n}_1 + \mathbf{n}_2 + \cdots + \mathbf{n}_b) = k_b\} \\ (2.2) \qquad &= \frac{[x_1^{k_1} \cdots x_b^{k_b} \mathbf{z_1^{n_1}} \cdots \mathbf{z_d^{n_b}}]\Phi_b(x_1, x_2, \dots, x_b; \mathbf{z_1}, \dots, \mathbf{z_b})}{[\mathbf{z_1^{n_1}} \cdots \mathbf{z_d^{n_b}}]\Phi_b(1, 1, \dots, 1; \mathbf{z_1}, \dots, \mathbf{z_b})}. \end{aligned}$$

2.2. **Main Result.** The nature of $\Phi_1(x, z_1, \dots, z_d)$ (i.e., a $m$-th power) allows a straightforward application of the results of Bender and Richmond [1], which gives directly the convergence towards a Gaussian distribution. Especially the asymptotic mean

$$(2.3) \qquad \mathbf{E}X_m(n_1, \dots, n_d) = \frac{[z_1^{n_1} \cdots z_d^{n_d}]\frac{\partial}{\partial x}\Phi_1(1, z_1, \dots, z_d)}{[z_1^{n_1} \cdots z_d^{n_d}]\Phi_1(1, z_1, \dots, z_d)}.$$

and variance

$$\begin{aligned} \mathbf{Var}X_m(n_1, \dots, n_d) = \quad &= \frac{[z_1^{n_1} \cdots z_d^{n_d}]\frac{\partial^2}{\partial x^2}\Phi_1(1, z_1, \dots, z_d)}{[z_1^{n_1} \cdots z_d^{n_d}]\Phi_1(1, z_1, \dots, z_d)} \\ (2.4) \qquad &+\mathbf{E}X_m(n_1, \dots, n_d) - (\mathbf{E}X_m(n_1, \dots, n_d))^2 \end{aligned}$$

are both of order $m$ (as we will see in a moment). The idea is now to *approximate* $X_m(n_1, \dots, n_d)$ by

$$X_m(n_1, \dots, n_d) \approx \mathbf{E}X_m(n_1, \dots, n_d) + \sqrt{m} \cdot G(n_1/m, \dots, n_d/m),$$

where $G(t_1, \dots, t_d)$ is a proper Gaussian field.

The following theorem shows that this can be actually worked out. Note that Theorem 2.1 just refers to very general properties of generating functions and is thus applicable in more general situations.

**Theorem 2.1.** *Let $X_m(n_1,\dots,n_d)$, $m\ge 1$, $n_i\ge 0$, be a discrete process, such that there exist entire generating functions*

$$\phi_b(x_1,x_2,\dots,x_b;\mathbf{z_1},\dots,\mathbf{z_b})$$

*with $\mathbf{z_j}=(z_{1j},\dots,z_{dj})$ such that the following conditions are satisfied:*

1. *(2.2) holds with $\Phi_b=\phi_b^m$.*
2. $\phi_1(1,\mathbf{0})=1$.
3. $\dfrac{\partial\phi_1}{\partial z_i}(0,\mathbf{0})\neq 0$ *for* $i=1,\dots,d$.
4. $\phi_b(1,1,\dots,1,\mathbf{z_1},\dots,\mathbf{z_b})=\phi_1(1,\mathbf{z_1})\cdots\phi_1(1,\mathbf{z_b})$.

*Then there exists a centered and continuous Gaussian field $G(t_1,\dots,t_d)$, $t_j\ge 0$, such that the following functional limit theorem holds:*

$$Y_m(t_1,\dots,t_d):=\frac{X_m(\lfloor mt_1\rfloor,\dots,\lfloor mt_d\rfloor)-\mathbf{E}X_m(\lfloor mt_1\rfloor,\dots,\lfloor mt_2\rfloor)}{\sqrt{m}}\xrightarrow{w}G(t_1,\dots,t_d)$$

*The covariance function $B_{s_1,\dots,s_d;t_1,\dots,t_d}$ of $G(t_1,\dots,t_d)$ satisfies $B_{s_1,\dots,s_d;t_1,\dots,t_d}=B_{\min(s_1,t_1),\dots,\min(s_d,t_d);\max(s_1,t_1),\dots,\max(s_d,t_d))}$, $s_i,t_i\ge 0$, and for $s_i\le t_i$ it is given by*

$$B_{s_1,\dots,s_d;t_1,\dots,t_d}=\frac{\begin{vmatrix}\kappa_{u_1u_2} & \kappa_{u_1v_1} & \cdots & \kappa_{u_1v_d}\\ \kappa_{v_1u_2} & \kappa_{v_1v_1} & \cdots & \kappa_{v_1v_d}\\ \vdots & \vdots & & \vdots\\ \kappa_{v_du_2} & \kappa_{v_dv_1} & \cdots & \kappa_{v_dv_d}\end{vmatrix}}{\begin{vmatrix}\kappa_{v_1v_1} & \cdots & \kappa_{v_1v_d}\\ \vdots & & \vdots\\ \kappa_{v_dv_1} & \cdots & \kappa_{v_dv_d}\end{vmatrix}},\tag{2.5}$$

*where $\kappa_{yz}$ $(y,z\in\{u_1,u_2,v_1,\dots,v_d\})$ is defined by*

$$\kappa_{yz}:=\left.\frac{\partial^2(\log\phi_2(e^{u_1},e^{u_2},\rho_{11}e^{v_1},\dots,\rho_{d1}e^{v_d},\rho_{12},\dots,\rho_{d2})}{\partial y\partial z}\right|_{u_1=u_2=v_1=\cdots=v_d=0}$$

*and $\rho_{i1}=\rho_i(s_1,\dots,s_d)$ and $\rho_{i2}=\rho_i(t_1-s_1,\dots,t_d-s_d)$ $(i=1,\dots,d)$ are the saddle points $\rho_i=\rho_i(\tau_1,\dots,\tau_d)$ which are defined by the equations in $z_i$*

$$z_i\,\frac{\partial\phi_1(1,z_1,\dots,z_d)}{\partial z_i}\;=\;\tau_i\phi_1(1,z_1,\dots,z_d),\qquad i=1,\dots,d.\tag{2.6}$$

*Furthermore we have*

$$\mathbf{E}X_m(n_1,\dots,n_d)=m\,\mu_{n_1/m,\dots,n_d/m}+O\left(1\right),$$

*where*

$$\mu_{s_1,\dots,s_d}=\frac{\frac{\partial}{\partial x}\phi_1(1,\rho_1(s_1,...,s_d),\dots,\rho_d(s_1,...,s_d))}{\phi_1(1,\rho_1(s_1,...,s_d),\dots,\rho_d(s_1,...,s_d))},\tag{2.7}$$

*and*

$$\mathbf{Cov}(X_m(n_1,\dots,n_d),X_m(\tilde n_1,\dots,\tilde n_d))=m\,B_{n_1/m,\dots,n_d/m;\tilde n_1/m,\dots,\tilde n_d/m}+O\left(1\right).$$

*Remark* 1. Note that $\kappa_{yz}$ can be easily represented explicitly. We have

$$\begin{aligned}
\kappa_{u_1u_2} &= \frac{\phi_2 \frac{\partial^2\phi_2}{\partial x_1 \partial x_2} - \frac{\partial\phi_2}{\partial x_1}\frac{\partial\phi_2}{\partial x_2}}{\phi_2^2},\\
\kappa_{u_1v_i} &= \rho_{i1}\frac{\phi_2 \frac{\partial^2\phi_2}{\partial x_1 \partial z_{i1}} - \frac{\partial\phi_2}{\partial x_1}\frac{\partial\phi_2}{\partial z_{i1}}}{\phi_2^2},\\
\kappa_{v_iu_2} &= \rho_{i1}\frac{\phi_2 \frac{\partial^2\phi_2}{\partial z_{i1} \partial x_2} - \frac{\partial\phi_2}{\partial z_{i1}}\frac{\partial\phi_2}{\partial x_2}}{\phi_2^2},\\
\kappa_{v_iv_j} &= \rho_{i1}\frac{\frac{\partial\phi_2}{\partial z_{i1}}}{\phi_2}\cdot\delta_{ij} + \rho_{i1}\rho_{j1}\frac{\phi_2 \frac{\partial^2\phi_2}{\partial z_{i1} \partial z_{i2}} - \frac{\partial\phi_2}{\partial z_{i1}}\frac{\partial\phi_2}{\partial z_{i2}}}{\phi_2^2},
\end{aligned}$$

where we have to evaluate at $(x_1, x_2, \mathbf{z}_1, \mathbf{z}_2) = (1, 1, \rho_{11}, \dots, \rho_{d1}, \rho_{12}, \dots, \rho_{d2})$.

*Remark* 2. We want to point out that the assumptions for the generating function $\phi_b$ have two interesting implications which will be frequently used in the proof of the theorem. We have

$$\phi_2(x_1, x_2, \mathbf{z}_1, \mathbf{0}) = \phi_1(x_1x_2, \mathbf{z}_1). \tag{2.8}$$

and

$$\phi_2(x_1, 1, \mathbf{z}_1, \mathbf{z}_2) = \phi_1(x_1, \mathbf{z}_1)\phi_1(1, \mathbf{z}_2). \tag{2.9}$$

In order to show (2.8) we use (2.2) for $b = 2$ and $\mathbf{n}_2 = \mathbf{0}$ and use the fact that $\phi_2(1, 1, \mathbf{z}_1, \mathbf{0}) = \phi_1(1, \mathbf{z}_1)\phi_1(1, \mathbf{0}) = \phi_1(1, \mathbf{z}_1)$.

The proof of (2.9) is similar. We again use (2.2) for $b = 2$ but sum over all $k_2$. We also have to use the relation $\phi_2(1, 1, \mathbf{z}_1, \mathbf{z}_2) = \phi_1(1, \mathbf{z}_1)\phi_1(1, \mathbf{z}_2)$ in order to get correct normalization.

The equations (2.8) and (2.9) can be also used to derive a simpler expression for $B_{s_1,\dots,s_d;s_1,\dots,s_d}$, namely

$$B_{s_1,\dots,s_d;s_1,\dots,s_d} = \mu_{s_1,\dots,s_d} + \frac{\begin{vmatrix} \tilde\kappa_{uu} & \tilde\kappa_{uv_1} & \cdots & \tilde\kappa_{uv_d} \\ \tilde\kappa_{v_1u} & \tilde\kappa_{v_1v_1} & \cdots & \tilde\kappa_{v_1v_d} \\ \vdots & \vdots & & \vdots \\ \tilde\kappa_{v_du} & \tilde\kappa_{v_dv_1} & \cdots & \tilde\kappa_{v_dv_d} \end{vmatrix}}{\begin{vmatrix} \tilde\kappa_{v_1v_1} & \cdots & \tilde\kappa_{v_1v_d} \\ \vdots & & \vdots \\ \tilde\kappa_{v_dv_1} & \cdots & \tilde\kappa_{v_dv_d} \end{vmatrix}}, \tag{2.10}$$

where $\tilde\kappa_{yz}$ $(y, z \in \{u, v_1, \dots, v_d\})$ is defined by

$$\tilde\kappa_{y,z} := \left.\frac{\partial^2(\log\phi_1(e^u, \rho_1e^{v_1}, \dots, \rho_de^{v_d})}{\partial y \partial z}\right|_{u=v_1=\cdots=v_d=0}$$

and $\rho_i = \rho_i(s_1, \dots, s_d)$ $(1 \le i \le d)$ are defined in (2.6). Of course, we have

$$\mathbf{Var}X_m(n_1, \dots, n_d) = m\,B_{n_1/m,\dots,n_d/m,n_1/m,\dots,n_d/m} + O(1).$$

If $t_i = s_i$ then by (2.6) we have $\rho_{i2} = 0$. Thus we can apply (2.9) and get

$$\begin{aligned}
\kappa_{u_1v_i} &= \tilde\kappa_{uv_i},\\
\kappa_{v_iu_2} &= \tilde\kappa_{v_iu},\\
\kappa_{v_iv_j} &= \tilde\kappa_{v_iv_j}.
\end{aligned}$$

Furthermore, we obtain by using (2.8)

$$
\begin{aligned}
\kappa_{u_1u_2} &= \frac{\phi_2 \frac{\partial^2\phi_2}{\partial x_1 \partial x_2} - \frac{\partial \phi_2}{\partial x_1}\frac{\partial \phi_2}{\partial x_2}}{\phi_2^2} \\
&= \frac{\frac{\partial^2}{\partial x_1 \partial x_2}\phi_1(x_1x_2, \rho_{11}, \dots, \rho_{d1})}{\phi_1(x_1x_2, \rho_{11}, \dots, \rho_{d1})} \\
&- \frac{\frac{\partial}{\partial x_1}\phi_1(x_1x_2, \rho_{11}, \dots, \rho_{d1})\frac{\partial}{\partial x_2}\phi_1(x_1x_2, \rho_{11}, \dots, \rho_{d1})}{\phi_1(x_1x_2, \rho_{11}, \dots, \rho_{d1})^2} \\
&= \frac{\frac{\partial \phi_1}{\partial x}(1, \rho_{11}, \dots, \rho_{d1})}{\phi_1(1, \rho_{11}, \dots, \rho_{d1})} \\
&+ \frac{\phi_1(1, \rho_{11}, \dots, \rho_{d1})\frac{\partial^2 \phi_2}{\partial x^2}(1, \rho_{11}, \dots, \rho_{d1}) - \left(\frac{\partial \phi_1}{\partial x}(1, \rho_{11}, \dots, \rho_{d1})\right)^2}{\phi_1(1, \rho_{11}, \dots, \rho_{d1})^2} \\
&= \mu_{s_1, \dots, s_d} + \tilde{\kappa}_{uu}.
\end{aligned}
$$

This proves (2.10).

*Remark* 3. The assumption that the functions $\phi_b$ are entire is not a necessity; the actual requirement is that any saddle point considered during the evaluation of a Cauchy integral throughout the proof be closer to the origin than any singularity of the integrand. Moreover it is not necessary that the generating functions $\phi_b$ are power series in $x_i$. We only have to require that they are analytic in $x_i$ in a neighbourhood of 1 which provides that we can also treat real valued valuations $f(k_1, \dots, k_d)$.

## 3. Join and balanced urns Models

As discussed in the Introduction, some important cases appear when studying join sizes or balanced urns. We specify now the results of Theorem 2.1 for several cases. We first make explicit the covariance formula in the case $d = 2$, then consider the case of models with factorization (see Remark 2 of Section 2), and finally give explicit results for equijoins and semijoins and for balanced urns.

### 3.1. Formulas for the case $d = 2$.

We work with the functions $\phi_1(x, y, z)$ and $\phi_2(x_1, x_2, y_1, z_1, y_2, z_2)$ : In $\phi_1$, $x$ marks the valuation and $y$ and $z$ mark the balls of first and second type; in $\phi_2$, $x_1$ and $x_2$ mark the valuation for the first and second batch, $y_1$ and $y_2$ mark the number of balls of the first type at each time, and $z_1$ and $z_2$ mark the number of balls of the second type for the first and second batch. The correspondance is as follows :

$$z_{11} \to y_1; z_{21} \to z_1; z_{12} \to y_2; z_{22} \to z_2.$$

The saddle points $\rho_{ij}$ that appear in the expression of the covariance are defined by the equations on the derivatives of $\phi_1(x, y, z)|_{x=1}$. The points $\rho_{11} = y$ and $\rho_{21} = z$ are defined by the system[2]

$$
\begin{cases}
\rho_{11}\, \frac{\partial \phi_1}{\partial y}(1, \rho_{11}, \rho_{21}) = s_1\, \phi_1(1, \rho_{11}, \rho_{21}); \\
\rho_{21}\, \frac{\partial \phi_1}{\partial z}(1, \rho_{11}, \rho_{21}) = s_2\, \phi_1(1, \rho_{11}, \rho_{21}).
\end{cases}
$$

[2] We also use the notations $\rho_1$ and $\rho_2$ when there is no ambiguity.

In the same way, the points $\rho_{12}$ and $\rho_{22}$ are the solutions of the system (again for $x = 1$)

$$\begin{cases} \rho_{12}\, \frac{\partial \phi_1}{\partial y}(1, \rho_{12}, \rho_{22}) = (t_1 - s_1)\, \phi_1(1, \rho_{12}, \rho_{22}); \\ \rho_{22}\, \frac{\partial \phi_1}{\partial z}(1, \rho_{12}, \rho_{22}) = (t_2 - s_2)\, \phi_1(1, \rho_{12}, \rho_{22}). \end{cases}$$

The covariance function is

$$B_{s_1,s_2;t_1,t_2} = \frac{\begin{vmatrix} \kappa_{u_1u_2} & \kappa_{u_1v_1} & \kappa_{u_1v_2} \\ \kappa_{v_1u_2} & \kappa_{v_1v_1} & \kappa_{v_1v_2} \\ \kappa_{v_2u_2} & \kappa_{v_2v_1} & \kappa_{v_2v_2} \end{vmatrix}}{\begin{vmatrix} \kappa_{v_1v_1} & \kappa_{v_1v_2} \\ \kappa_{v_2v_1} & \kappa_{v_2v_2} \end{vmatrix}}.$$

The $\kappa_{\cdot,\cdot}$ are obtained from Remark 1 as derivatives of the function $\phi_2(x_1, x_2, y_1, z_1, y_2, z_2)$, taken at the point $(1, 1, \rho_{11}, \rho_{21}, \rho_{12}, \rho_{22})$.

3.1.1. *Expectation and variance.* To get the asymptotic value $\mathbf{E}X_m(n_1, n_2) = m\mu_{n_1/m,n_2/m} + O(1)$, we have to compute

$$\mu_{s_1,s_2} = \frac{\frac{\partial \phi_1}{\partial x}(1, \rho_1, \rho_2)}{\phi_1(1, \rho_1, \rho_2)},$$

where the saddle points $\rho_1$ and $\rho_2$ are defined by the system (we have that $s_i = n_i/m$ in all this section)

$$\begin{cases} \rho_1\, \frac{\partial \phi_1}{\partial y}(1, \rho_1, \rho_2) = s_1\, \phi_1(1, \rho_1, \rho_2); \\ \rho_2\, \frac{\partial \phi_1}{\partial z}(1, \rho_1, \rho_2) = s_2\, \phi_1(1, \rho_1, \rho_2). \end{cases}$$

For the variance we have :

$$\mathbf{Var}X_m(n_1, n_2) = m\, \tilde{B}_{n_1/m,n_2/m} + O(1),$$

with

$$\begin{aligned} \tilde{B}_{s_1,s_2} \quad &:= \quad B_{s_1,s_2;s_1,s_2} = \mu_{s_1,s_2} + \frac{\begin{vmatrix} \tilde\kappa_{u_1u_2} & \tilde\kappa_{u_1v_1} & \tilde\kappa_{u_1v_2} \\ \tilde\kappa_{v_1u_2} & \tilde\kappa_{v_1v_1} & \tilde\kappa_{v_1v_2} \\ \tilde\kappa_{v_2u_2} & \tilde\kappa_{v_2v_1} & \tilde\kappa_{v_2v_2} \end{vmatrix}}{\begin{vmatrix} \tilde\kappa_{v_1v_1} & \tilde\kappa_{v_1v_2} \\ \tilde\kappa_{v_2v_1} & \tilde\kappa_{v_2v_2} \end{vmatrix}} \\ &= \quad \mu_{s_1,s_2} + \tilde\kappa_{u,u} - \frac{\tilde\kappa_{uv_2}^2 \sigma_1^2 + \tilde\kappa_{uv_1}^2 \sigma_2^2 - 2\tilde\kappa_{v_1v_2}^2 \tilde\kappa_{uv_1}^2 \tilde\kappa_{uv_2}^2}{\sigma_1^2\sigma_2^2 - \tilde\kappa_{v_1v_2}^2}, \end{aligned}$$

where we used the notations

$$\sigma_1^2 = \rho_{11}^2 \frac{\frac{\partial^2 \phi_1}{\partial y^2}}{\phi_1}(1, \rho_{11}, \rho_{21}) + s_1 - s_1^2; \qquad \sigma_2^2 = \rho_{21}^2 \frac{\frac{\partial^2 \phi_1}{\partial z^2}}{\phi_1}(1, \rho_{11}, \rho_{21}) + s_2 - s_2^2.$$

3.1.2. *Models with factorization.* For join or balanced urns models without deletion, there is a further simplification in the computation of the variance :

$$\phi_1(1, y, z) = \lambda_1(y)\ \lambda_2(z). \tag{3.1}$$

The saddle points $\rho_{11}$ and $\rho_{21}$ that appear in the expression of the expectation or variance are now defined as solutions of two separate equations :

$$\begin{cases} \rho_{11}\ \frac{d\lambda_1}{dy}(\rho_{11}) = s_1\ \lambda_1(\rho_{11}); \\ \rho_{21}\ \frac{d\lambda_2}{dz}(\rho_{21}) = s_2\ \lambda_2(\rho_{21}). \end{cases}$$

In the computations for the variance, the coefficients $\tilde{\kappa}_{.,.}$ can also be simplified, by taking into account the relations $\frac{\partial\phi_1}{\partial y} = \lambda_1^{'}(y)\lambda_2(z)$ and $\frac{\partial\phi_1}{\partial z} = \lambda_1(y)\lambda_2^{'}(z)$. We get $\tilde{\kappa}_{v_1 v_2} = 0$. This gives the asymptotic variance as $\mathbf{Var}X_m(n_1, n_2) = m\tilde{B}_{s_1,s_2} + O(1)$, with

$$\tilde{B}_{s_1,s_2} = \mu_{s_1,s_2} + \tilde{\kappa}_{uv} - \left(\frac{\tilde{\kappa}_{uv_1}}{\sigma_1}\right)^2 - \left(\frac{\tilde{\kappa}_{uv_2}}{\sigma_2}\right)^2.$$

For the computation of the covariance, we need the saddle points $\rho_{12}$ and $\rho_{22}$ :

$$\begin{cases} \rho_{12}\ \frac{\partial\lambda_1}{\partial y}(\rho_{12}) = (t_1 - s_1)\ \lambda_1(\rho_{12}); \\ \rho_{22}\ \frac{\partial\lambda_2}{\partial z}(\rho_{22}) = (t_2 - s_2)\ \lambda_2(\rho_{22}). \end{cases}$$

Now the derivatives of the function $\phi_2$ that appear in the expressions of $\kappa_{u_1 u_2}$, $\kappa_{v_1 u_2}$ and $\kappa_{v_2 u_2}$ are evaluated at the point $(1, 1, \rho_{11}, \rho_{21}, \rho_{12}, \rho_{22})$. The other terms $\kappa_{i,j}$ do not involve a derivative w.r.t. $x_2$ and we can get them from the variance computations. The equalities $\kappa_{v_1 v_2} = \kappa_{v_2 v_1} = 0$, which come from the factorization property, still hold, which allow for some simplifications :

$$B_{s_1,s_2;t_1,t_2} = \kappa_{u_1 u_2} - \frac{\kappa_{v_1 u_2} . \kappa_{u_1 v_1}}{\sigma_1^2} - \frac{\kappa_{v_2 u_2} . \kappa_{u_1 v_2}}{\sigma_2^2}.$$

We also have simpler expressions for $\sigma_1^2$ and $\sigma_2^2$ :

$$\sigma_1^2 = \rho_{11}^2 \frac{\lambda_1^{''}}{\lambda_1}(\rho_{11}) + s_1 - s_1^2; \qquad \sigma_2^2 = \rho_{21}^2 \frac{\lambda_2^{''}}{\lambda_2}(\rho_{21}) + s_2 - s_2^2.$$

We now present applications of our results to join models (equijoin and semijoin) and to balanced urns.

3.2. **Equijoin.** In this part, $\phi_1(x, y, z) = \sum_{k,l} a_k b_l x^{kl} y^k z^l$, and we use the relations

$$\sum_k k a_k y^k = y\lambda_1^{'}(y); \qquad \sum_k k^2 a_k y^k = y^2 \lambda_1^{''}(y) + y\lambda_1^{'}(y).$$

Of course, similar relations hold for $\lambda_2$. This gives $\frac{\partial\phi_1}{\partial x}(1, y, z) = y\lambda_1^{'}(y)\ z\lambda_2^{'}(z)$; we obtain the asymptotic expectation as

$$\mathbf{E}X_m(n_1, n_2) = \frac{n_1 n_2}{m} + O(1) = m\, s_1\, s_2 + O(1).$$

The expression for the variance simplifies into $\mathbf{Var}X_m(n_1, n_2) = m\tilde{B}_{s_1,s_2} + O(1)$, with

$$\tilde{B}_{s_1,s_2} = \sigma_1^2\, \sigma_2^2.$$

We can get more precise results for infinite or bounded urns. For infinite urns, $\lambda_i(t) = e^t$, the saddle point is $s_i$ and $\sigma_i^2 = s_i$; for urns of bounded size $\delta_i$, $\lambda_i(t) = (1 + t)^{\delta_i}$, $\rho_i = s_i/(\delta_i - s_i)$ and $\sigma_i^2 = s_i(1 - s_i/\delta_i)$ :

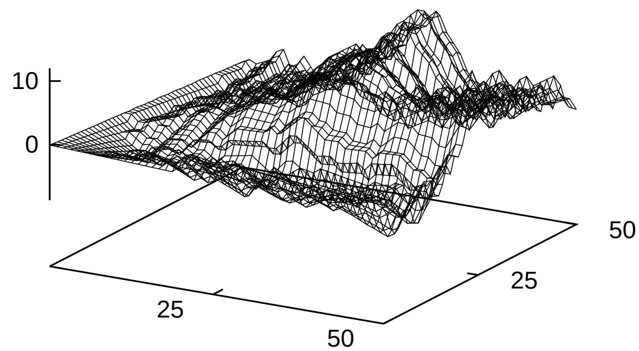


FIGURE 1. Centered process for the equijoin size, infinite urns, $m = 20$ and $n_1, n_2 \leq 50$.

- Urns are unbounded for both types of balls. Then $\tilde{B}_{s_1,s_2} = s_1\, s_2$.
- Urns are unbounded for one type of balls and bounded for the other type. For example, $\lambda_1(t) = e^t$ and $\lambda_2(t) = (1+t)^{\delta_2}$; then $\tilde{B}_{s_1,s_2} = s_1 s_2(1 - s_2/\delta_2)$.
- Urns are bounded for both types of balls. Then $\tilde{B}_{s_1,s_2} = s_1 s_2(1 - s_1/\delta_1)(1 - s_2/\delta_2)$.

The covariance is equal to $mB_{s_1,s_2;t_1,t_2} + O(1)$, with $B_{s_1,s_2;t_1,t_2} = \sigma_1^2\, \sigma_2^2$, hence

$$\mathbf{Cov}(X_m(n_1,n_2), X_m(\tilde{n}_1,\tilde{n}_2)) = \mathbf{Var} X_m(n_1,n_2) + O(1).$$

Now the limiting process can be precisely characterized : If $s_1$ is fixed, the covariance of the process $G(s_1,.)$ between two times $t_1$ and $t_2 \geq t_1$ is $\sigma_1^2 t_1$ if the urns (for the second type of balls) are infinite, and $\sigma_1^2 t_1(1 - t_1/d_2)$ if the urns are bounded. This means that, up to a multiplicative constant, $G$ is a Brownian Motion (for infinite urns) or a Brownian Bridge (for bounded urns). The same result holds if $s_2$ is fixed. Hence $G(s_1, s_2)$ is a Brownian sheet, or a variation thereof [18]. More precisely, we get a Brownian sheet if the urns are infinite for both types of balls, a Kiefer-Müller process when the urns are infinite for one type of balls, and bounded for the other type, and a tucked Brownian sheet for urns bounded on both types of balls.

3.3. **Semijoin.** We now turn to the semijoin. We have that

$$\phi_1(x,y,z) = \lambda_1(y) + \lambda_1(xy)(\lambda_2(z) - 1).$$

The asymptotic expectation is $\mathbf{E}X_m(n_1,n_2) = m\mu_{s_1,s_2} + O(1)$, with

$$\mu_{s_1,s_2} = s_1\left(1 - \frac{1}{\lambda_2(\rho_2)}\right).$$

The variance is $\mathbf{Var}X_m(n_1,n_2) = m\tilde{B}_{s_1,s_2} + O(1)$ with

$$\tilde{B}_{s_1,s_2} = \frac{(s_1^2 + \sigma_1^2)(\lambda_2(\rho_2) - 1)}{\lambda_2^2(\rho_2)} - \frac{s_1^2\, s_2^2}{\sigma_2^2\, \lambda_2^2(\rho_2)}.$$

Again we can get results for bounded or unbounded urns : We just plug in the values for $\sigma_1$ and $\sigma_2$ which were computed in Section 3.2.

- Urns are unbounded for both types of balls. Then

$$\tilde{B}_{s_1,s_2} = s_1 e^{-s_2}\left(1 + s_1 - e^{-s_2}(1 + s_1 + s_1 s_2)\right).$$

- Urns are unbounded for the first type of balls and bounded for the second type. Then

$$\tilde{B}_{s_1,s_2} = s_1 \left(\frac{\delta_2 - s_2}{\delta_2}\right)^{2\delta_2 - 1} \left[(1 + s_1)\left(1 - \frac{s_2}{\delta_2}\right) - s_1 s_2\right].$$

- Urns are bounded for the first type of balls and unbounded for the second type.

$$\tilde{B}_{s_1,s_2} = s_1 e^{-s_2}\left[1 + s_1 - \frac{s_1}{\delta_1} - e^{-s_2}\left(1 + s_1 - \frac{s_1}{\delta_1} + s_1 s_2\right)\right].$$

- Urns are bounded for both types of balls.

$$\tilde{B}_{s_1,s_2} = s_1 \left(\frac{\delta_2 - s_2}{\delta_2}\right)^{2\delta_2 - 1} \left[\left(1 + s_1 - \frac{s_1}{\delta_1}\right)\left(1 - \frac{s_2}{\delta_2}\right) - s_1 s_2\right].$$

The generating function for the two-dimensional distributions is

$$\Phi_2(x_1, x_2, y_1, y_2, z_1, z_2) = \phi_2(x_1, x_2, y_1, y_2, z_1, z_2)^m$$

where the function $\phi_2$ describing the possible states of one urn is equal to

$\lambda_1(y_1)\lambda_2(y_2) + \lambda_1(x_2 y_1)\lambda_1(x_2 y_2)(\lambda_2(z_2) - 1) + \lambda_1(x_1 x_2 y_1)\lambda_1(x_2 y_2)(\lambda_2(z_1) - 1)\lambda_2(z_2).$

This gives $\mathbf{Cov} X_m = m B_{s_1,s_2;t_1,t_2} + O(1)$, with

$$\begin{aligned} B_{s_1,s_2;t_1,t_2} &= \frac{\lambda_2(\rho_{21}) - 1}{\lambda_2(\rho_{21})^2 \lambda_2(\rho_{22})}\left[(s_1^2 + \sigma_1^2) + s_1(s_1 - t_1)(\lambda_2(\rho_{21})\lambda_2(\rho_{22}) - 1)\right] \\ &\quad - \frac{s_1 t_1 s_2^2}{\sigma_2^2 \lambda_2(\rho_{21})^2 \lambda_2(\rho_{22})}. \end{aligned}$$

We can also write $B_{s_1,s_2;t_1,t_2}$ as

$$\frac{1}{\lambda_2(\rho_{22})}\left(\tilde{B}_{s_1,s_2} - \frac{s_1(t_1 - s_1)}{\lambda_2(\rho_{21})}\left[\frac{s_2^2}{\sigma_2^2} + (\lambda_2(\rho_{21}) - 1)(\lambda_2(\rho_{21})\lambda_2(\rho_{22}) - 1)\right]\right).$$

For example, infinite urns on both types of balls give

$$B_{s_1,s_2;t_1,t_2} = s_1\left(-(t_1 - s_1)(1 - e^{-s_2}) + e^{-t_2}(1 + t_1 - e^{-s_2}[1 + t_1 + s_2 t_1])\right)$$

3.4. **Urns of balance $q$.** The valuation of the urn is equal to 1 if the difference between the number of balls of the first type and the number of balls of the second type is $q$, and to 0 otherwise. We shall use Hadamard products to express our results; we recall that the Hadamard product of the two functions $f(t) = \sum_k f_k t^k$ and $g(t) = \sum_k g_k t^k$ is

$$(f \odot g)(t) = \sum_k f_k g_k t^k.$$

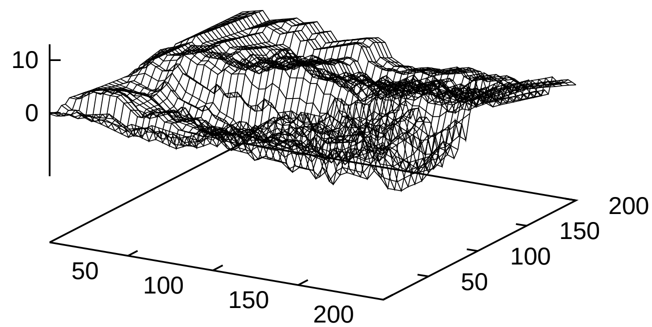


FIGURE 2. Centered process for the semijoin size, unbounded urns, $m = 80$ and $n_1, n_2 \leq 200$.

We define a shifted version of the Hadamard product of the functions $\lambda_1$ and $\lambda_2$ (defined by the equation (3.1)) as

$$g_q(t) := \sum_l a_{l+q} b_l t^l.$$

Of course $g_0(t) = \lambda_1 \odot \lambda_2(t)$.

We have here $\phi_1(x, y, z) = \lambda_1(y)\lambda_2(z) + (x-1)y^q g_q(yz)$, which we can also write as

$$\phi_1(x, y, z) = \lambda_1(y)\lambda_2(z) + (x-1)\psi_q(y, z) \qquad \text{with} \qquad \psi_q(y, z) := [u^q]\lambda_1(uy)\lambda_2\left(\frac{z}{u}\right).$$

This comes from the fact that the generating function marking balls of the first and second kind by $y$ and $z$ and the balance by $u$ is simply $\lambda_1(uy)\lambda_2\left(\frac{z}{u}\right)$. We also have that $\psi_q(y, z) = y^q g_q(yz)$.

In the same vein, the generating function for allocations in two batches can be written as

$$\begin{aligned}
\phi_2(x_1, x_2, y_1, y_2, z_1, z_2) \quad = \quad & (x_1 - 1)(x_2 - 1)\, \psi_q(y_1, z_1)\, \psi_0(y_2, z_2) \\
& + (x_1 - 1)\, \psi_q(y_1, z_1)\, \lambda_1(y_2)\, \lambda_2(z_2) \\
& + (x_2 - 1)\, \Pi_q(y_1, y_2, z_1, z_2) \\
& + \lambda_1(y_1)\, \lambda_1(y_2)\, \lambda_2(z_1)\, \lambda_2(z_2),
\end{aligned}$$

with

$$\Pi_q(y_1, y_2, z_1, z_2) := [u^q]\, \lambda_1(uy_1)\, \lambda_1(uy_2)\, \lambda_2\left(\frac{z_1}{u}\right)\, \lambda_2\left(\frac{z_2}{u}\right).$$

The asymptotic expectation is

$$\mathbf{E}X_m(n_1, n_2) = m\mu_{s_1,s_2}(q) + O(1) \qquad \text{with} \qquad \mu_{s_1,s_2}(q) := \frac{\rho_1^q\, g_q(\rho_1\rho_2)}{\lambda_1(\rho_1)\lambda_2(\rho_2)}.$$

We now define

$$\tau := \rho_1\rho_2 \frac{g_q^{'}(\rho_1\rho_2)}{g_q(\rho_1\rho_2)}.$$

Then the asymptotic variance is $\mathbf{Var} X_m(n_1, n_2) = m\tilde{B}_{s_1,s_2}(q) + O(1)$, with

$$\tilde{B}_{s_1,s_2}(q) = \mu_{s_1,s_2}(q)\left(1 - \mu_{s_1,s_2}(q)\left[1 + \frac{(\tau + q - s_1)^2}{\sigma_1^2} + \frac{(\tau - s_2)^2}{\sigma_2^2}\right]\right).$$

To compute the covariance, we introduce

$$\alpha := \frac{\rho_{11}}{\Pi_q}\frac{\partial \Pi_q}{\partial y_1}; \qquad \beta := \frac{\rho_{21}}{\Pi_q}\frac{\partial \Pi_q}{\partial z_1}.$$

We get

$$\tilde{B}_{s_1,s_2}(q) = \\ \mu_{s_1,s_2}(q)\left(\frac{g_0(y_2 z_2)}{\phi_1(y_2, z_2)} - \frac{\Pi_q}{\phi_2}\left[1 + \frac{(\alpha - s_1)(q + \tau - s_1)}{\sigma_1^2} + \frac{(\beta - s_2)(\tau - s_2)}{\sigma_2^2}\right]\right).$$

*For infinite urns,* the generating functions $\phi_1$ and $\phi_2$ can be expressed in terms of Bessel functions. We have that $\lambda_1(t) = \lambda_2(t) = e^t$ (as in Section 3.2), which gives $\sigma_1^2 = s_1 = \rho_1$ and $\sigma_2^2 = s_2 = \rho_2$; moreover $g_q(t) = t^{-q/2} I_q(2\sqrt{t})$ and $g_q^{'}(t) = t^{-(q+1)/2} I_{q+1}(2\sqrt{t}) = g_{q+1}(t)$. We get

$$\mu_{s_1,s_2}(q) = \left(\frac{s_1}{s_2}\right)^{q/2} I_q(2\sqrt{s_1 s_2})e^{-s_1 - s_2}.$$

Now $\tau = \sqrt{s_1 s_2}\, I_{q+1}(2\sqrt{s_1 s_2})/I_q(2\sqrt{s_1 s_2})$, and we get the variance by

$$\tilde{B}_{s_1,s_2}(q) = \mu_{s_1,s_2}(q)\left(1 - \mu_{s_1,s_2}(q)\left[1 + \frac{1}{s_1}(\tau - s_1 + q)^2 + \frac{1}{s_2}(\tau - s_2)^2\right]\right).$$

We also have that

$$\Pi_q(y_1, y_2, z_1, z_2) = \psi_q(y_1 + y_2, z_1 + z_2) = \left(\frac{y_1 + y_2}{z_1 + z_2}\right)^{q/2} I_q(2\sqrt{(y_1 + y_2)\,(z_1 + z_2)}).$$

Define

$$\eta := \sqrt{t_1 t_2}\, I_{q+1}(2\sqrt{t_1 t_2})/I_q(2\sqrt{t_1 t_2});$$

then $\alpha = s_1(q + \eta)/t_1$ and $\beta = s_2\eta/t_2$. We finally get

$$\begin{aligned} B_{s_1,s_2;t_1,t_2} \quad = \quad & \left(\frac{s_1}{s_2}\right)^{q/2} e^{-t_1 - t_2} I_q(2\sqrt{s_1 s_2}). \\ & \left(I_0(2\sqrt{(t_1 - s_1)(t_2 - s_2)}) - \xi\,\left(\frac{t_1}{t_2}\right)^{q/2} e^{-s_1 - s_2} I_q(2\sqrt{t_1 t_2})\right), \end{aligned}$$

with

$$\xi = 1 + \frac{(q + \eta - t_1)(q + \tau - t_1)}{t_1} + \frac{(\eta - t_2)(\tau - t_2)}{t_2}.$$

## 4. Models with Deletions

In some instances, e.g. when modelizing dynamic databases to study the evolution of projection or join sizes, we need to allow both the deletion of items (balls), and the existence of queries that do not modify the current state of the system (no ball is added or deleted). We shall assume that *the balls inserted are all of the same type, and that a deletion specifies the urn in which a ball must be deleted, but not the precise ball among those present in the urn (which are undistinguishable).* In this section, we first present the ideas that allow us to write generating functions for models with deletions, then show on an example how these functions can be written down and used to characterize the limiting Gaussian process. The study of general update models (including those with queries) will be presented in a further paper.

### 4.1. What happens in a single urn?

We introduce a new type of balls to mark deletions : We shall consider white balls, the ones originally thrown into the urn, and black balls, corresponding to deletions. Of course, we cannot delete balls that are not present in the urn. The difference between the number of white balls and the number of black balls is the number of balls remaining in the urn, and cannot become negative. This leads us to consider a scheme of allocation with two types of balls such that, at each time, the *balance* of the urn, defined as the number of white balls minus the number of black balls, is positive or null.

Such a situation is related to the framework presented in [10]. There we proved that, starting from a general combinatorial structure for which we have the enumerating generating function, and assuming that the basic items can take two colors, we can easily obtain the bivariate generating function marking the size and the color balance, by taking the Hadamard product of the initial enumerating function and of the function associated with the sequence of balances. Requiring that the sequence of balances is always positive simply means that this sequence is the prefix of a Dyck path, for which the enumerating function is well-known. Allowing for queries that do not add or delete balls, we simply take prefixes of Motzkin paths as allowed sequences.

To sum up, what happens in a single urn is described by taking the Hadamard product of the function for an urn with a single type of balls and of the function for the sequence of balances; a white ball is an insertion and a black ball is a deletion; the current balance is the number of balls in the urn at the specified time. Hence we obtain the generating function for the urn as a Hadamard product.

To make clear our ideas, we present now a simple model. Our assumptions are as follows :

- The urns are infinite : there is no upper limit on the number of white balls that an urn can receive.
- The balls that are in the same urn are undistinguishable, when performing either an insertion or a deletion.
- We first choose the urn, then the operation to be done in this urn; the only possible operations are insertion or deletion of a ball.
- The urns are chosen with uniform probability $1/m$.
- Assuming that the urn that has been chosen is not empty, the probabilities of insertion and deletion in this urn are equal. If the urn is empty, then we perform an insertion.

We modelize this situation with two types of balls : White balls correspond to insertions, and are thrown according to the usual rules (there is no upper limit on the number of white balls in an urn); black balls correspond to deletions, and are thrown in such a way that the balance of an urn, defined as the number of white balls minus the number of black balls, is always positive or null.

In the multivariate generating function associated to an urn, we use the variables $x$ to mark the fact that the urn is empty, $z$ to mark the balance of the urn, and $t$ to mark the total number of balls (black and white) that this urn has received : $t$ marks the number of insertions/deletions relative to the urn. The global generating function relative to the sequence of $m$ urns is obtained by taking the $m$-th power of the function for one urn, where the variables $x$, $z$ and $t$ mark respectively the number of empty urns, the number of remaining balls (balls inserted and not deleted) in the sequence of urns, and the total number of operations, i.e. the time.

4.2. **Generating functions.** The function describing the allocation of balls into one urn is[3]

$$\lambda(t,z) = g(t) \odot_t P(t,z),$$

where $g(t)$ is the function describing the allocation of (white and black) balls into the urn (usually $g(t) = e^t$), and $P(t,z) = \sum_{n,q} p_{n,q} t^n z^q$ is the bivariate function enumerating the allowed sequences of allocations of black and white balls into the urn. The number of such sequences with $n$ operations (the final time is equal to $n$) and an excess of $q$ insertions over deletions is $p_{n,q}$ : At the end of the sequence, the urn contains $q$ balls that have not been deleted (this is closely related to the well-known *ballot numbers*). Now $P(t,z)$ is simply the generating function for prefixes of Dyck paths, with $t$ marking the length and $z$ the final height : An up step corresponds to an insertion, a down step to a deletion, we cannot go under the zero axis, and the final height is positive (or null for Dyck paths). Let $d(t) := (1-\sqrt{1-4t^2})/2t^2$ be the function enumerating Dyck paths; then $P(t,z) = d(t)/(1-tzd(t))$. We also mention that the function enumerating prefixes of Dyck paths that end at height $q$ is $p_q(t) := \sum_n p_{n,q} t^n = t^q d(t)^{q+1}$.

The function describing the behaviour of one urn, with $x$ marking the fact that the urn is empty, $t$ marking the number of operations relative to the urn, and $z$ marking the number of remaining balls, is

$$\phi_1(x,t,z) = \lambda(t,z) + (x-1)\lambda(t,0) = g(t) \odot_t \left( \frac{d(t)}{1-tzd(t)} + (x-1)d(t) \right).$$

We consider now what happens at two successive times. Let $\pi_{n_1,n_2,q_1,q_2}$ be the number of sequences of balances of length $n_1+n_2$, such that after $n_1$ steps, the balance is $q_1$, and that the final balance is $q_2$, and define the generating function of these numbers :

$$\pi(t_1,t_2,z_1,z_2) := \sum_{n_1,n_2,q_1,q_2} \pi_{n_1,n_2,q_1,q_2}\, t_1^{n_1} t_2^{n_2} z_1^{q_1} z_2^{q_2}$$

Our next step is to compute this function $\pi$. At least as long as we are working with unbounded urns, it does not depend on the function $g(z)$ enumerating allocations in one urn, and it is simply the generating function for prefixes of Dyck path, enumerated according to their total length $n_1+n_2$ and final height $q_2$, and to

[3]In the case of multivariate functions, we index the Hadamard product by the relevant variable.

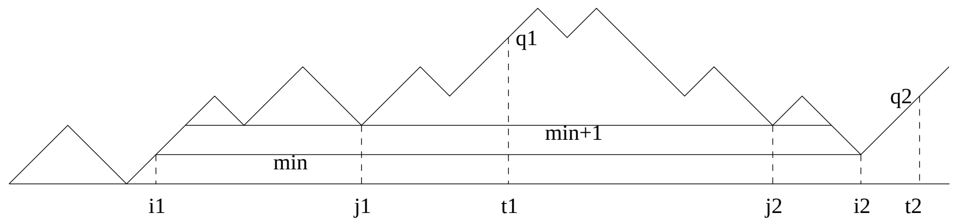


FIGURE 3. Decomposition of a Dyck prefix

some intermediate length $n_1$ and corresponding height $q_1$. We decompose the paths according to their minimal height $min$ between the times $t_1$ and $t_2$ : Let $i_1$ be the time of last passage at $min$ before $t_1$, and let $i_2$ be the time of first passage after $t_1$. Obviously $min \leq q_1, q_2$ and $i_1 \leq t_1 \leq i_2 \leq t_2$.

- The part between 0 and $i_1$ is the prefix of a Dyck path, whose generating function is $d(t_1)/(1 - t_1\, d(t_1))$. Taking into account the heights at times $t_1$ and $t_2$ gives
$$\frac{d(t_1)}{1 - t_1 z_1 z_2 d(t_1)}.$$
- In the central part of the path, the minimal height $min$ can be equal to $q_1$ : then $i_1 = t_1 = i_2$. Otherwise, the path begins by an up step, then stays at height at least $min + 1$ in the interval $[i_1 + 1, i_2 - 1]$. We shall consider the times $j_1$ and $j_2$ of last passage to $min + 1$ before $t_1$, and of first passage to $min + 1$ after $t_1$. The path between $i_1$ and $j_1$ is enumerated by $z_1 t_1 d(t_1)$, and the path between $j_2$ and $i_2$ is enumerated by $t_2 d(t_2)$. Hence the multivariate generating function describing the central part of the path (including the case $q_1 = min$) is
$$\frac{1}{1 - z_1\, t_1\, t_2\, d(t_1)\, d(t_2)}.$$
- Finally, the part between the times $i_2$ and $t_2$ is again a Dyck path, and we mark the final height at time $t_2$, which gives
$$\frac{d(t_2)}{1 - t_2\, z_2\, d(t_2)}.$$

Concatenating the three parts of the path gives :
$$\pi(t_1, t_2, z_1, z_2) = \frac{d(t_1)\, d(t_2)}{(1 - t_1 z_1 z_2 d(t_1))\,(1 - t_2 z_2 d(t_2))\,(1 - t_1 t_2 z_1 d(t_1) d(t_2))}.$$

Now let $\kappa(t_1, t_2, z_1, z_2) := \sum_{n_1, n_2, q_1, q_2} k_{n_1, n_2, q_1, q_2} t_1^{n_1} t_2^{n_2} z_1^{q_1} z_2^{q_2}$ be the function enumerating allocations of black and white balls in two batches, such that, after throwing $n_1$ balls, the balance is $q_1$, and after throwing again $n_2$ balls in the second batch, the balance becomes $q_2$. As for the one-dimensional case, we have that
$$\kappa(t_1, t_2, z_1, z_2) = g(t_1) \odot_{t_1} (g(t_2) \odot_{t_2} \pi(t_1, t_2, z_1, z_2))$$

The function marking the emptyness of the urn at the end of the first or second batches by the variables $x_1$ and $x_2$ is

$$\begin{aligned}\phi_2(x_1,x_2,t_1,t_2,z_1,z_2) &= (x_1-1)(x_2-1)\kappa(t_1,t_2,0,0)+(x_1-1)\kappa(t_1,t_2,0,z_2)\\ &\quad +(x_2-1)\kappa(t_1,t_2,z_1,0)+\kappa(t_1,t_2,z_1,z_2).\end{aligned}$$

We have expressions for the $\kappa(t_1,t_2,.,.)$ as Hadamard products of the entire functions $g(t_1)=e^{t_1}$ and $g(t_2)=e^{t_2}$, and of algebraic functions $\pi(t_1,t_2,.,.)$. Hence the function $\phi_2(x_1,x_2,t_1,t_2,z_1,z_2)$ is an entire function in $t_1$ and $t_2$.

It is not difficult, but cumbersome, to check that we can write down all the desired multivariate generating functions, and that they satisfy the assumptions of Theorem 2.1; hence the associated process converges towards a Gaussian field $G(s,t)$. Note that the first *time* $s=n/m$ corresponds to the total number $n$ of operations (insertions and deletions) and $t=q/m$ to the difference. Hence, $\overline{n}=(n+q)/2$ is the number of insertions and $\overline{q}=(n-q)/2$ is the number of deletions. This means that we can define a modified discrete process $\overline{X}_m$ by

$$\overline{X}_m(\overline{n},\overline{q}) := X_m\left(\overline{n}+\overline{q},\overline{n}-\overline{q}\right)$$

which counts the number of empty urns with $\overline{n}$ insertions and $\overline{q}$ deletions and another Gaussian process $\overline{G}(s,t)$ $(0\le t\le s)$ by

$$\overline{G}(\overline{s},\overline{t}) = G\left(\overline{s}+\overline{t},\overline{s}-\overline{t}\right)$$

so that

$$\overline{X}_m(\overline{n},\overline{q}) \approx \mathbf{E}\overline{X}_m(\overline{n},\overline{q})+\sqrt{m}\cdot\overline{G}(\overline{n}/m,\overline{q}/m).$$

4.3. **Number of empty urns.** In this part, we consider the number of empty urns but for simplicity we just take the total number of operations into account (which is a functional of the bidimensional process we studied above) and show that we can effectively compute the parameters of the limiting process. We get the functions $\phi_1$ and $\phi_2$ by putting $z=z_1=z_2=1$ in the corresponding functions computed in Section 4.2 :

$$\begin{aligned}\varphi_1(x,t) &= (x-1)\;g(t)\odot d(t)+g(t)\odot\frac{d(t)}{1-td(t)};\\ \varphi_2(x_1,x_2,t_1,t_2) &= (x_1-1)(x_2-1)\kappa(t_1,t_2,0,0)+(x_1-1)\kappa(t_1,t_2,0,1)\\ &\quad +(x_2-1)\kappa(t_1,t_2,1,0)+\kappa(t_1,t_2,1,1).\end{aligned}$$

To compute the moments and the limiting distribution, we may use Theorem 2.1 for $d=1$ or the results of [6]. Our first goal is the computation of the saddle point that appears in the expression of the asymptotic expectation and variance. Let us define $f(t):=\lambda(t,0)=g(t)\odot d(t)$ and $g_1(t):=\lambda(t,1)=g(t)\odot d(t)/(1-td(t))$; so that $\phi_1(x,t)=g_1(t)+(x-1)f(t)$. The saddle point $\rho$ is defined as the unique real positive solution of the equation $tg_1^{'}(t)/g_1(t)=s$, where we put $s:=n/m$. The asymptotic expectation is $\mathbf{E}X_m(n)=m\mu_{n/m}+O(1)$, with

$$\mu_s = f(\rho)/g_1(\rho).$$

Now $f(t)=I_1(2t)/t$, with $I_1$ a Bessel function [10]. Further $g_1(t)=e^{2t}(1-K(t))$, where the function $K$ is defined as the solution of the equation $tK^{'}(t)=e^{-2t}I_1(2t)/t$ that becomes null for $t=0$; hence

$$g_1(t)=e^{2t}\left(1-\int_0^t e^{-2u}I_1(2u)\frac{du}{u}\right).$$

Now $g_1^{'}(t) = 2g_1(t) - f(t)$ and the equation defining the saddle point $\rho$ translates to

$$2\rho - \frac{\rho f(\rho)}{g_1(\rho)} = s.$$

This equation can be solved numerically with Maple for any given $s$; it can also be used to simplify the expression of the asymptotic expectation : We get $\mu_s = 2 - s/\rho$, whose numerical value is simple to obtain. For example, $s = 1$ gives $\rho = .6793222511...$ and $\mathbf{E}X_m(n) \sim .527944582...m$.

The asymptotic variance is computed from [6, Corollary to Theorem 2.1] : $\mathbf{Var}X_m(n) = m\tilde{B}_s + O(1)$, with

$$\tilde{B}_s = \tau_0 \left(1 - \tau_0 \left[1 + \frac{s\eta_1^2}{s\eta_2 - (s-1)\eta_1^2} \left(1 - \frac{\rho\tau_1}{s\tau_0}\right)^2\right]\right), \tag{4.1}$$

where the parameters $\tau_0$, $\tau_1$, $\eta_1$ and $\eta_2$ are

$$\tau_0 = \frac{f(\rho)}{g_1(\rho)}; \qquad \tau_1 = \frac{f^{'}(\rho)}{g_1(\rho)}; \qquad \eta_1 = \frac{g_1^{'}(\rho)}{g_1(\rho)}; \qquad \eta_2 = \frac{g_1^{''}(\rho)}{g_1(\rho)}. \tag{4.2}$$

Equations (4.1) and (4.2) are quite general, and we can simplify them by plugging in the exact values of $f$ and $g_1$. We get expressions of the parameters $\tau_i$ and $\eta_i$ using only $s$, $\rho$ and the quotient $\Delta := I_0(2\rho)/I_1(2\rho)$ :

$$\tau_0 = 2 - \frac{s}{\rho}; \qquad \tau_1 = 2(\Delta - \frac{1}{\rho})(2 - \frac{s}{\rho}); \qquad \eta_1 = \frac{s}{\rho}; \qquad \eta_2 = \frac{2}{\rho}(s + 2 - \frac{s}{\rho}) - 2\Delta(2 - \frac{s}{\rho}).$$

This gives us an expression of the asymptotic variance as

$$\tilde{B}_s = \frac{2}{\rho}(2\rho - s)\Delta + \frac{1}{\rho^2}(s\rho - 2\rho^2 - 4\rho + 3s) + \frac{s^2}{\rho^2\left(2\rho(2\rho - s)\Delta - 2s\rho + s^2 - 4\rho + s\right)}.$$

Numerical computation for $s = 1$ gives an asymptotic variance equal to $0.17394268...m$. As another example, we choose $s = 2$; then $\rho = 1.2154678...$, the expected value is $0.3545302...m$, and the variance is $0.1953331...m$.

## 5. Proof of Theorem 2.1

5.1. **Existence of Limiting Gaussian Field with a.s. Continuous Sample Paths.** In order to prove Theorem 2.1 we first have to show that there exists a random field with a.s. continuous sample paths and f.d.d.'s which are characterized by the limiting f.d.d.'s of $Y_m(t_1, \dots, t_d)$. We start with the following property (not including continuity).

**Lemma 5.1.** *There exists a Gaussian field $G(t_1, \dots, t_d)$ with covariance function $B_{s_1,\dots,s_d;t_1,\dots,t_d}$, given by (2.5) such that the finite dimensional distribution of*

$$Y_m(t_1, \dots, t_d) := \frac{X_m(\lfloor mt_1 \rfloor, \dots, \lfloor mt_d \rfloor) - \mathbf{E}X_m(\lfloor mt_1 \rfloor, \dots, \lfloor mt_2 \rfloor)}{\sqrt{m}}$$

*converge weakly to the corresponding finite dimensional distributions of $G(t_1, \dots, t_d)$.*

**Proof.** The limiting distribution of $Y_m$ is characterized by the normalized limit of

$$\mathbf{P}\left\{X_m(n_{1j},\ldots,n_{dj})=k_j, j=1,\ldots,b\right\}$$
$$=\frac{\left[\prod_{j=1}^b x_j^{k_j} z_{1j}^{n_{1j}}\cdots z_{dj}^{n_{dj}}\right]\Phi_1(x_1^{k_1},\ldots,x_b^{k_b};\mathbf{z_1}^{\mathbf{n_1}},\ldots,\mathbf{z_j}^{\mathbf{n_b}})}{[\mathbf{z_1}^{\mathbf{n_1}}\cdots\mathbf{z_b}^{\mathbf{n_b}}]\,\Phi_1(1,\ldots,1;\mathbf{z_1}^{\mathbf{n_1}},\ldots,\mathbf{z_j}^{\mathbf{n_b}})}.$$

where $\mathbf{n_j}=(n_{1j},\ldots,n_{dj})$. Applying the results of Bender and Richmond [1] yields that the above limiting distribution is a Gaussian distribution with asymptotic mean

$$\mathbf{E}X_m(n_1,\ldots,n_d)=m\mu_{n_1/m,\ldots,n_d/m}+O\left(1\right),$$

and covariance

$$\mathbf{Cov}(X_m(n_1,\ldots,n_d),X_m(\tilde n_1,\ldots,\tilde n_d))=mB_{n_1/m,\ldots,n_d/m;\tilde n_1/m,\ldots,\tilde n_d/m}.+O\left(1\right).$$

Especially, for the variance we have

$$\mathbf{Var}X_m(n_1,\ldots,n_d)=mB_{n_1/m,\ldots,n_d/m;n_1/m,\ldots,n_d/m}+O\left(1\right).$$

Here $\mu(s_1,\ldots,s_d)$ and $B_{s_1,\ldots,s_d;s_1,\ldots,s_d}$ are given by

$$\mu_{s_1,\ldots,s_d}=\left.\frac{\partial(\log\lambda_{s_1,\ldots,s_d}(e^u))}{\partial u}\right|_{u=0}\tag{5.1}$$

and by

$$B_{s_1,\ldots,s_d;s_1,\ldots,s_d}=\left.\frac{\partial^2(\log\lambda_{s_1,\ldots,s_d}(e^u))}{\partial^2 u}\right|_{u=0},\tag{5.2}$$

where $\lambda_{s_1,\ldots,s_d}(x)$ denotes

$$\lambda_{s_1,\ldots,s_d}(x)=\frac{\phi_1(x,\rho_1,\ldots,\rho_d)}{\rho_1^{s_1}\cdots\rho_d^{s_d}}$$

and $\rho_i=\rho_i(x,s_1,\ldots,s_d)$ $(1\le i\le d)$ are the saddle points defined by the equations in $z_i$

$$z_i\frac{\partial\phi_1(x,z_1,\ldots,z_d)}{\partial z_i}\quad=\quad s_i\phi_1(x,z_1,\ldots,z_d),\qquad i=1,\ldots,d.\tag{5.3}$$

Furthermore, for $s_i,t_i\ge 0$ we have

$$B_{s_1,\ldots,s_d;t_1,\ldots,t_d}=B_{\min(s_1,t_1),\ldots,\min(s_d,t_d),\max(s_1,t_1),\ldots,\max(s_d,t_d))}$$

and for $s_i<t_i$

$$B_{s_1,\ldots,s_d;t_1,\ldots,t_d}=\left.\frac{\partial^2(\log\lambda_{s_1,\ldots,s_d,t_1,\ldots,t_d}(e^{u_1},e^{u_2}))}{\partial u_1\partial u_2}\right|_{u_1=0,u_2=0}\tag{5.4}$$

with

$$\lambda_{s_1,\ldots,s_d,t_1,\ldots,t_d}(x_1,x_2)=\frac{\phi_2(x_1,x_2,\rho_{11},\ldots,\rho_{d1},\rho_{12},\ldots,\rho_{d2})}{\rho_{11}^{s_1}\cdots\rho_{d1}^{s_d}\rho_{12}^{t_1-s_1}\cdots\rho_{d2}^{t_d-s_d}},$$

where $\rho_{ij}=\rho_{ij}(x_1,x_2,s_1,\ldots,s_d,t_1,\ldots,t_d)$ $(i=1,\ldots,d,\ j=1,2))$ are the saddle points which are defined by the equations in $z_{ij}$

$$z_{i1}\frac{\partial\phi_2(x_1,x_2,\mathbf{z_1},\mathbf{z_2})}{\partial z_{i1}}\quad=\quad s_i\phi_2(x_1,x_2,\mathbf{z_1},\mathbf{z_2}),\qquad i=1,\ldots,d,\tag{5.5}$$

$$z_{i2}\frac{\partial\phi_2(x_1,x_2,\mathbf{z_1},\mathbf{z_2})}{\partial z_{i2}}\quad=\quad(t_i-s_i)\phi_2(x_1,x_2,\mathbf{z_1},\mathbf{z_2})\qquad i=1,\ldots,d.\tag{5.6}$$

Now let $G(t_1,\dots,t_d)$ be the Gaussian field with covariance function $B_{s_1,\dots,s_d;t_1,\dots,t_d}$ (compare with [17], by construction it is clear that the corresponding covariance matrices are positive semi-definite.) The above construction also ensures that all (normalized) finite dimensional distributions of $X_m(n_1,\dots,n_d)$ converge weakly to the corresponding finite dimensional distributions of $G(t_1,\dots,t_d)$.

It remains to show that $\mu(s_1,\dots,s_d)$ and $B_{s_1,\dots,s_d;t_1,\dots,t_d}$ can be represented by the *nice* formulae (2.7), (2.10), and (2.5). This is just an (advanced) exercise in implicit differentiation. (In order to simplify notation we use the abbreviations $\mathbf{s}=(s_1,\dots,s_d)$ and $\mathbf{t}=(t_1,\dots,t_d)$.)

First, let us consider

$$\begin{aligned}\frac{\partial(\log\lambda_{\mathbf{s}}(e^u))}{\partial u} &= \frac{\frac{\partial}{\partial u}\phi_1(e^u,\rho_1(e^u,\mathbf{s}),\dots,\rho_1(e^u,\mathbf{s}))}{\phi_1(\cdots)}\\ &\quad - s_1\frac{\frac{\partial}{\partial u}\rho_1(e^u,\mathbf{s})}{\rho_1(e^u,\mathbf{s})}-\cdots-s_d\frac{\frac{\partial}{\partial u}\rho_d(e^u,\mathbf{s})}{\rho_d(e^u,\mathbf{s})}\\ &= \frac{e^u\frac{\partial\phi_1}{\partial x}(e^u,\rho_1(e^u,\mathbf{s}),\dots,\rho_d(e^u,\mathbf{s}))}{\phi_1(\cdots)}\\ &\quad + e^u\left(\frac{\frac{\partial\phi_1}{\partial z_1}(\cdots)\frac{\partial\rho_1}{\partial x}(\cdots)}{\phi_1(\cdots)}-s_1\frac{\frac{\partial\rho_1}{\partial x}(\cdots)}{\rho_1(\cdots)}\right)+\cdots\\ &\quad + e^u\left(\frac{\frac{\partial\phi_1}{\partial z_d}(\cdots)\frac{\partial\rho_d}{\partial x}(\cdots)}{\phi_1(\cdots)}-s_d\frac{\frac{\partial\rho_d}{\partial x}(\cdots)}{\rho_d(\cdots)}\right)\\ &= \frac{e^u\frac{\partial\phi_1}{\partial x}(e^u,\rho_1(e^u,\mathbf{s}),\dots,\rho_d(e^u,\mathbf{s}))}{\phi_1(\cdots)},\end{aligned}$$

which proves that (2.7) is correct.

Similarly we can treat the second derivatives:

$$\begin{aligned}\frac{\partial^2(\log\lambda_{\mathbf{s}}(e^u))}{\partial u^2} &= \frac{e^u\frac{\partial\phi_1}{\partial x}}{\phi_1}+\frac{e^{2u}\frac{\partial^2\phi_1}{\partial x^2}}{\phi_1}+\sum_{i=1}^d\frac{e^{2u}\frac{\partial^2\phi_1}{\partial x\partial z_i}\frac{\partial\rho_i}{\partial x}}{\phi_1}\\ &\quad -\frac{e^{2u}\left(\frac{\partial\phi_1}{\partial x}\right)^2}{\phi_1^2}-\sum_{i=1}^d\frac{e^{2u}\frac{\partial\phi_1}{\partial x}\frac{\partial\phi_1}{\partial z_i}\frac{\partial\rho_i}{\partial x}}{\phi_1^2}.\end{aligned}$$

Setting $u=0$ and using the definition of $\tilde\kappa_{yz}$ we also get

$$\left.\frac{\partial^2(\log\lambda_{\mathbf{s}}(e^u))}{\partial u^2}\right|_{u=0}=\mu_{\mathbf{s}}+\tilde\kappa_{u,u}+\sum_{i=1}^d\tilde\kappa_{uz_i}\frac{\frac{\partial\rho_i}{\partial x}}{\rho_i}. \tag{5.7}$$

The derivatives $\frac{\partial}{\partial x}\rho_i$ can be calculated by differentiating (5.3) with respect to $x$:

$$\frac{\partial\rho_i}{\partial x}\frac{\partial\phi_1}{\partial z_i}+\rho_i\frac{\partial^2\phi_1}{\partial x\partial z_i}+\rho_i\sum_{j=1}^d\frac{\partial^2\phi_1}{\partial z_i\partial z_j}\frac{\partial\rho_j}{\partial x}=s_i\frac{\partial\phi_1}{\partial x}+s_i\sum_{j=1}^d\frac{\partial\phi_1}{\partial z_j}\frac{\partial\rho_j}{\partial x}$$

or after setting $x = 1$, substitution $s_i = \rho_i \frac{\partial \phi_1}{\partial z_i} / \phi_1$, and using again the definition of $\tilde{\kappa}_{yz}$

$$\sum_{j=1}^{d} \tilde{\kappa}_{z_i z_j} \frac{\frac{\partial \rho_j}{\partial x}}{\rho_j} = -\tilde{\kappa}_{u z_i}. \tag{5.8}$$

By combining (5.7) and (5.8) and using Cramer's rule we directly obtain (2.10).

The verification of (2.5) is similar but a little bit more involved. For convenience we extend the definition of $\kappa_{yz}$ to $y, z \in \{u_1, u_2, v_1, \ldots, v_d, w_1, \ldots, w_d\}$ by

$$\kappa_{y,z} := \left. \frac{\partial^2 (\log \phi_2(e^{u_1}, e^{u_2}, \rho_{11} e^{v_1}, \ldots, \rho_{d1} e^{v_d}, \rho_{12} e^{w_1}, \ldots, \rho_{d2} e^{w_d})}{\partial y \partial z} \right|_{u_1 = u_2 = v_1 = \cdots = v_d = w_1 = \cdots w_d = 0}.$$

From

$$\frac{\partial (\log \lambda_{\mathbf{s};\mathbf{t}}(e^{u_1}, e^{u_2}))}{\partial u_1} = \frac{e^{u_1} \frac{\partial \phi_2}{\partial x_1}(e^{u_1}, e^{u_2}, \rho_{11}(e^{u_1}, e^{u_2}, \mathbf{s}, \mathbf{t}), \ldots, \rho_{d1}(\cdots), \rho_{12}(\cdots), \ldots, \rho_{d2}(\cdots))}{\phi_2(\cdots)},$$

we get (after differentiating with respect to $u_2$, setting $u_1 = u_2 = 0$ and using the definition of $\kappa_{yz}$)

$$\left. \frac{\partial^2 (\log \lambda_{\mathbf{s};\mathbf{t}}(e^{u_1}, e^{u_2}))}{\partial u_1 \partial u_2} \right|_{u_1 = u_2 = 0} = \kappa_{u_1 u_2} + \sum_{j=1}^{d} \kappa_{u_1 v_j} \frac{\frac{\partial \rho_{j1}}{\partial x_2}}{\rho_{j1}} + \sum_{k=1}^{d} \kappa_{u_1 w_2} \frac{\frac{\partial \rho_{k2}}{\partial x_2}}{\rho_{k2}}.$$

Now by differentiating (5.5) with respect to $x_2$ we get (similary as above)

$$\sum_{j=1}^{d} \kappa_{v_i v_j} \frac{\frac{\partial \rho_{j1}}{\partial x_2}}{\rho_{j1}} + \sum_{k=1}^{d} \kappa_{v_i w_k} \frac{\frac{\partial \rho_{k2}}{\partial x_2}}{\rho_{k2}} = -\kappa_{u_2 v_i}$$

and

$$\sum_{j=1}^{d} \kappa_{w_i v_j} \frac{\frac{\partial \rho_{j1}}{\partial x_2}}{\rho_{j1}} + \sum_{k=1}^{d} \kappa_{w_i w_k} \frac{\frac{\partial \rho_{k2}}{\partial x_2}}{\rho_{k2}} = -\kappa_{u_2 w_i}.$$

Now observe that

$$\kappa_{u_1 w_j} = 0 \quad (1 \le j \le d), \tag{5.9}$$

$$\kappa_{v_i w_j} = \kappa_{w_j v_i} = 0 \quad (1 \le i, j \le d). \tag{5.10}$$

This follows from definition and (2.9): $\phi_2(x_1, 1, \mathbf{z}_1, \mathbf{z}_2) = \phi_1(x_1, \mathbf{z}_1) \phi_1(1, \mathbf{z}_2)$. Hence, we have

$$\left. \frac{\partial^2 (\log \lambda_{\mathbf{s};\mathbf{t}}(e^{u_1}, e^{u_2}))}{\partial u_1 \partial u_2} \right|_{u_1 = u_2 = 0} = \kappa_{u_1 u_2} + \sum_{j=1}^{d} \kappa_{u_1 v_j} \frac{\frac{\partial \rho_{j1}}{\partial x_2}}{\rho_{j1}}$$

and

$$\sum_{j=1}^{d} \kappa_{v_i v_j} \frac{\frac{\partial \rho_{j1}}{\partial x_2}}{\rho_{j1}} = -\kappa_{u_2 v_i},$$

Hence, another application of Cramer's rule finally proves (2.5). The relation

$$\sum_{k=1}^{d} \kappa_{w_i w_k} \frac{\frac{\partial \rho_{k2}}{\partial x_2}}{\rho_{k2}} = -\kappa_{u_2 w_i}.$$

might be used to calculate $\frac{\partial \rho_{k2}}{\partial x_2}$ but is not needed for our purposes.

Note also that

$$\rho_{i1}(1, 1, \mathbf{s}, \mathbf{t}) = \rho_i(\mathbf{s})$$

and

$$\rho_{i2}(1, 1, \mathbf{s}, \mathbf{t}) = \rho_i(t_1 - s_1, \dots, t_d - s_d).$$

Thus, the defintion of $\kappa_{yz}$ fits to the calculations above. □

The next step is to show that $G(\mathbf{t})$ has continuous sample paths. This can be done by the Kolmogorov-Čentsov theorem, see [14, Ch. 2, Theorem 2.8 and Problem 2.9]):

**Theorem 5.1.** *Let $\mathbf{t} = (t_1, \dots, t_d)$ and $\mathbf{h} = (h_1, \dots, h_d)$. Then every real-valued random field $X(\mathbf{t})$ for which there exist three constants $\alpha$, $\beta$, $C > 0$ such that*

$$\mathbf{E}[|X(\mathbf{t} + \mathbf{h}) - X(\mathbf{t})|^\alpha] \le C\|\mathbf{h}\|^{d+\beta},$$

*for every $\mathbf{t} \in \mathbb{R}_+^d$ and $\mathbf{h} \in \mathbb{R}_+^d$, has a modification with a.s. continuous sample paths. The same holds on the space $C\left([0,T]^d\right)$ with $\mathbf{t}, \mathbf{t} + \mathbf{h} \in [0,T]^d$.*

The fact that $G(\mathbf{t})$ satisfies this criterion follows immediately by

**Lemma 5.2.** *We have for $\mathbf{s} = (s_1, \dots, s_d)$ and $\mathbf{t} = (t_1, \dots, t_d)$*

$$\mathbf{E}\left(G(\mathbf{t}) - G(\mathbf{s})\right)^{2d+2} = O\left(\|\mathbf{t} - \mathbf{s}\|^{d+1}\right),$$

*uniformly, if $\mathbf{s}$ and $\mathbf{t}$ are contained in an arbitrary bounded but fixed box.*

**Proof.** Since

$$\mathbf{E}\left(G(\mathbf{t}) - G(\mathbf{s})\right)^{2d+2} = \frac{(2d+2)!}{2^{d+1}(d+1)!}\left(B_{\mathbf{s};\mathbf{s}} - 2B_{\mathbf{s};\mathbf{t}} + B_{\mathbf{t};\mathbf{t}}\right)^{d+1}$$

it suffices to prove

$$B_{\mathbf{s};\mathbf{t}} = B_{\mathbf{s};\mathbf{s}} + O\left(\|\mathbf{t} - \mathbf{s}\|\right). \tag{5.11}$$

Due to symmetry it suffices to consider the case $s_1 \le t_1, \dots, s_d \le t_d$..

First we observe that by definition (5.3) the saddle points $\rho_i(\mathbf{t} - \mathbf{s})$ satisfy

$$\rho_i(\mathbf{t} - \mathbf{s}) = (t_i - s_i)\frac{\phi_1(1, \rho_1, \dots, \rho_d)}{\frac{\partial \phi_1}{\partial z_i}(1, \rho_1, \dots, \rho_d)}.$$

Thus, $\rho_i = O\left(t_i - s_i\right)$ and consequently

$$\rho_i(\mathbf{t} - \mathbf{s}) = (t_i - s_i)\left(c_i + O\left(\|\mathbf{t} - \mathbf{s}\|\right)\right), \tag{5.12}$$

where $c_i = 1/\frac{\partial \phi_1}{\partial z_i}(1, \mathbf{0}) \ne 0$.

We also use the property that $\phi_2$ (considered as a power series) can be represented as

$$\phi_2(x_1, x_2, \mathbf{z}_1, \mathbf{z}_2) = \phi_1(x_1 x_2, \mathbf{z}_1) + \sum_{j=1}^{d} z_{j2}^{l_j} R_j(x_1, x_2, \mathbf{z}_1, \mathbf{z}_2),$$

where $l_j \ge 1$ are integers and $R_j$ are proper power series with $R_j(x_1, x_2, \mathbf{z}_1, \mathbf{0}) \ne 0$ (compare with (2.8)). Note that $\phi_2(1, 1, \mathbf{z}_1, \mathbf{z}_2) = \phi_1(1, \mathbf{z}_1)\phi_1(1, \mathbf{z}_2)$ and thus, by assumption, we have $l_j = 1$. We will use this relation for *small* $\mathbf{z}_2$ and use the shorthand notation

$$\phi_2(x_1, x_2, \mathbf{z}_1, \mathbf{z}_2) = \phi_1(x_1 x_2, \mathbf{z}_1) + O\left(\mathbf{z}_2\right). \tag{5.13}$$

For convenience we set $\varepsilon = \|\mathbf{t} - \mathbf{s}\|$.

The next step is to show that

$$\kappa_{u_1u_2} = \mu_{\mathbf{s}} + \tilde{\kappa}_{uu} + O(\varepsilon) \tag{5.14}$$

$$\kappa_{u_1v_i} = \tilde{\kappa}_{uv_i} + O(\varepsilon) \quad (1 \le i \le d) \tag{5.15}$$

$$\kappa_{v_iu_2} = \tilde{\kappa}_{v_iu} + O(\varepsilon) \quad (1 \le i \le d) \tag{5.16}$$

$$\kappa_{v_iv_j} = \tilde{\kappa}_{v_iv_j} + O(\varepsilon) \quad (1 \le i, j \le d) \tag{5.17}$$

For the proof of (5.14) we use (for $x_1 = x_2 = 1$)

$$\frac{\partial\phi_2(1,1,\mathbf{z}_1,\mathbf{z}_2)}{\partial x_1} = \frac{\partial\phi_1(1,\mathbf{z}_1)}{\partial x} + O(\mathbf{z}_2),$$

$$\frac{\partial\phi_2(1,1,\mathbf{z}_1,\mathbf{z}_2)}{\partial x_2} = \frac{\partial\phi_1(1,\mathbf{z}_1)}{\partial x} + O(\mathbf{z}_2),$$

and

$$\frac{\partial^2\phi_2(1,1,\mathbf{z}_1,\mathbf{z}_2)}{\partial x_1\partial x_2} = \frac{\partial\phi_1(1,\mathbf{z}_1)}{\partial x} + \frac{\partial^2\phi_1(1,\mathbf{z}_1)}{\partial x^2} + O(\mathbf{z}_2).$$

There relations (together with (5.12)) directly imply (5.14). In the same way we can treat the other cases (5.15)–(5.17).

Thus, combining (5.14)–(5.17) it follows that

$$\begin{vmatrix} \kappa_{u_1u_2} & \kappa_{u_1v_1} & \cdots & \kappa_{u_1v_d} \\ \kappa_{v_1u_2} & \kappa_{v_1v_1} & \cdots & \kappa_{v_1v_d} \\ \vdots & \vdots & & \vdots \\ \kappa_{v_du_2} & \kappa_{v_dv_1} & \cdots & \kappa_{v_dv_d} \end{vmatrix} = \begin{vmatrix} \mu_{\mathbf{s}} + \tilde{\kappa}_{uu} & \tilde{\kappa}_{uv_1} & \cdots & \tilde{\kappa}_{uv_d} \\ \tilde{\kappa}_{v_1u} & \tilde{\kappa}_{v_1v_1} & \cdots & \tilde{\kappa}_{v_1v_d} \\ \vdots & \vdots & & \vdots \\ \tilde{\kappa}_{v_du} & \tilde{\kappa}_{v_dv_1} & \cdots & \tilde{\kappa}_{v_dv_d} \end{vmatrix} + O(\varepsilon)$$

$$= \mu_{\mathbf{s}} \begin{vmatrix} \tilde{\kappa}_{v_1v_1} & \cdots & \tilde{\kappa}_{v_1v_d} \\ \vdots & & \vdots \\ \tilde{\kappa}_{v_dv_1} & \cdots & \tilde{\kappa}_{v_dv_d} \end{vmatrix} + \begin{vmatrix} \tilde{\kappa}_{uu} & \tilde{\kappa}_{uv_1} & \cdots & \tilde{\kappa}_{uv_d} \\ \tilde{\kappa}_{v_1u} & \tilde{\kappa}_{v_1v_1} & \cdots & \tilde{\kappa}_{v_1v_d} \\ \vdots & \vdots & & \vdots \\ \tilde{\kappa}_{v_du} & \tilde{\kappa}_{v_dv_1} & \cdots & \tilde{\kappa}_{v_dv_d} \end{vmatrix} + O(\varepsilon)$$

and that

$$\begin{vmatrix} \kappa_{v_1v_1} & \cdots & \kappa_{v_1v_d} \\ \vdots & & \vdots \\ \kappa_{v_dv_1} & \cdots & \kappa_{v_dv_d} \end{vmatrix} = \begin{vmatrix} \tilde{\kappa}_{v_1v_1} & \cdots & \tilde{\kappa}_{v_1v_d} \\ \vdots & & \vdots \\ \tilde{\kappa}_{v_dv_1} & \cdots & \tilde{\kappa}_{v_dv_d} \end{vmatrix} + O(\varepsilon).$$

Of course, this proves that

$$B_{\mathbf{s};\mathbf{t}} = B_{\mathbf{s};\mathbf{s}} + O(\varepsilon),$$

which completes the proof of the lemma. □

5.2. **Tightness.** We will need the following Lemma bounding certain central moments of the process.

**Lemma 5.3.** *For all integers $\Delta > 0$ there exists a constant $C > 0$ such that for $m \to \infty$*

$$\mathbf{E}(X_m(\mathbf{n}) - \mathbf{E}X_m(\mathbf{n}))^{2\Delta} \le C\|\mathbf{n}\|^{\Delta}, \tag{5.18}$$

*uniformly for $\|\mathbf{n}\| = O(m)$.*

**Proof.** Set $\mathbf{z} = (z_1, \ldots, z_d)$ and

$$c_{\boldsymbol{n},\alpha} := [\mathbf{z}^{\boldsymbol{n}}] \left. \frac{\partial^\alpha}{\partial x^\alpha} \Phi_1(x, \mathbf{z}) \right|_{x=1}$$

where $\mathbf{z}^{\boldsymbol{n}}$ denotes $z_1^{n_1} z_2^{n_2} \cdots z_d^{n_d}$. Furthermore let

$$A_i := \mathbf{E} \prod_{j=0}^{i-1} (X_m(\mathbf{n}) - j) = \frac{c_{\boldsymbol{n},i}}{c_{\boldsymbol{n},0}}.$$

Then the moment occurring in (5.18) can now be expressed by

$$\mathbf{E}(X_m(\mathbf{n}) - \mathbf{E}X_m(\mathbf{n}))^{2\Delta} = \sum_{l=0}^{2\Delta} \binom{2\Delta}{l} (-1)^l A_1^{2\Delta-l} \sum_{k=1}^{l} S_{lk} A_k, \tag{5.19}$$

where $S_{nk}$ denotes the Stirling numbers of the second kind and the empty sum occurring in the above summation for $l = 0$ is supposed to be equal to 1.

Hence we have to compute $c_{\boldsymbol{n},\alpha}$. If we set

$$d_j(\mathbf{z}) = \frac{1}{\mathbf{g}(\mathbf{z})} \left. \frac{\partial^j}{\partial x^j} \phi_1(x, \mathbf{z}) \right|_{x=1},$$

where we use the notation $\mathbf{g}(\mathbf{z}) := g_1(z_1) g_2(z_2) \cdots g_d(z_d)$, then by Faà di Bruno's formula (see e.g. Comtet [4]) we have

$$c_{\boldsymbol{n},\alpha} = \sum_{\sum_j jk_j = \alpha} \frac{\alpha!}{k_1! \cdots k_\alpha!} m(m-1) \cdots (m - k_1 - \cdots - k_\alpha + 1) [\mathbf{z}^{\boldsymbol{n}}] \mathbf{g}(\mathbf{z})^m \prod_{j=1}^{\alpha} \left( \frac{d_j(\mathbf{z})}{j!} \right)^{k_j}.$$

Thus we have to calculate the coefficient

$$[\mathbf{z}^{\boldsymbol{n}}]\, \mathbf{g}(\mathbf{z})^m \prod_{j=1}^{\alpha} \left( \frac{d_j(\mathbf{z})}{j!} \right)^{k_j}. \tag{5.20}$$

For this task we use the ideas developed in detail for simpler urn models in [6]. First note that $c_{\boldsymbol{n},0} = [\mathbf{z}^{\boldsymbol{n}}] \mathbf{g}(\mathbf{z})^m$ and that by Taylor's theorem we have for real $z$

$$g_l\left(ze^{i\theta}\right) = g_l(z) \exp\left( \sum_{j=1}^{k} \frac{(i\theta)^j}{j!} \kappa_{l,j}(z) + O\left(|\theta^{k+1}|z\right) \right)$$

where

$$\kappa_{l1} := \frac{z g_l'(z)}{g_l(z)} \quad \text{and} \quad \kappa_{l,j+1}(z) := z\kappa_{l,j}'(z), \quad j \geq 1, l = 1, \ldots, d.$$

Since there exist no $r, d$ such that $g_n \neq 0$ if and only if $g_{l,n} \equiv r \mod d$ we have moreover

$$|g_l(ze^{i\theta})| \leq g_l(z) e^{-c\theta^2}$$

for some positive constant $c$. Hence we can apply the saddle point method. If $\mu_l = \kappa_{l1}^{-1}$ for $l = 1, \ldots, d$, then the saddle points of $g_l(z_l)^m z_l^{-n_l}$ for $l = 1, \ldots, d$ are given by

$$\rho_l = \mu_l \left( \frac{n_l}{m} \right) = \frac{g_{l0}}{g_{l1}} \frac{n_l}{m} \left( 1 + O\left( \frac{n_l}{m} \right) \right)$$

Note that $g_{lk} \neq 0$ for $l = 1, \dots, d$ and $k = 0, 1$, since we allow an urn to be empty or to contain only one ball regardless of its type. Now define functions $\bar{\kappa}_{lj}$ by

$$\bar{\kappa}_{lj}\left(\frac{n}{m}\right) = \frac{m}{n}\kappa_{lj}\left(\mu_l\left(\frac{n}{m}\right)\right)$$

which are analytic functions with $\bar{\kappa}_{lj}(0) = 1$. Let $\boldsymbol{\rho} = (\rho_1, \dots, \rho_d)$, as well as $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ Furthermore define $\mathbf{z}^k := (z_1^k, \dots, z_d^k)$. Then applying the saddle point method yields

$$\begin{aligned}
[\mathbf{z}_1^{\boldsymbol{n}}]\mathbf{g}(\mathbf{z})^m &= \frac{\mathbf{g}(\boldsymbol{\rho})^m}{(2\pi)^d \boldsymbol{\rho}^{\boldsymbol{n}}} \underset{B}{\int \cdots \int} \exp\left(-\sum_{l=1}^{d} \frac{\theta_l^2}{2} n_l \bar{\kappa}_{l2}\left(\frac{n_l}{m}\right) + \sum_{l=1}^{d}\sum_{j=3}^{k} \frac{(i\theta_{l1})^j}{j!} n_l \bar{\kappa}_{lj}\left(\frac{n_l}{m}\right)\right. \\
&\quad \left. + O\left(m \sum_{l=1}^{d} |\rho_l \theta_l^{k+1}|\right)\right) d\theta_1 \cdots d\theta_d \\
&= \frac{\mathbf{g}(\boldsymbol{\rho})^m}{(2\pi)^d \boldsymbol{\rho}^{\boldsymbol{n}} \sqrt{\prod_{l=1}^{d} n_l \kappa_{l2}(n_l/m)}} \underset{\tilde{B}}{\int \cdots \int} \exp\left(-\sum_{l=1}^{d} \frac{u_l^2}{2} + \sum_{l=1}^{d}\sum_{j=3}^{k} \frac{(iu_l)^j}{j!} n_l^{1-j/2} \tilde{\kappa}_{lj}\left(\frac{n_l}{m}\right)\right. \\
&\quad \left. + O\left(m \sum_{l=1}^{d} \rho_l \left|\frac{u_l}{\sqrt{n_l}}\right|^{k+1}\right)\right) du_1 \cdots du_d
\end{aligned}$$

where

$$\tilde{\kappa}_{lj}(x) = \bar{\kappa}_{lj}(x)\bar{\kappa}_{l2}(x)^{-j/2},$$

and the integration domain $B$ is given by

$$B = \{(\boldsymbol{\theta}) \,|\, |\theta_l| \leq (m\rho_l)^{-1/2+\varepsilon},\ l = 1, \dots, d\}$$

and $\tilde{B}$ is its transformation due to the substitutions $\theta_l = u_l/\sqrt{n_l \bar{\kappa}_{l2}(n_l/m)}$ and for $l = 1, \dots, d$. Now we could expand this into a series and evaluate the integral. In the general case ($\alpha > 0$) this yields some very complicated expressions involving, for example, Hermite polynomials (cf. [5] for expansions of similar type) which are quite hard to deal with. Fortunately, we need only some structural properties rather than the exact expansion in order to complete the proof.

Observe that, if we expand the integrand, except those terms containing only squares of $u_l$, into a series and set

$$V(\boldsymbol{\rho}, \mathbf{n}, m) = \frac{\mathbf{g}(\boldsymbol{\rho})^m}{(2\pi)^{d/2} \boldsymbol{\rho}^{\boldsymbol{n}} \sqrt{\prod_{l=1}^{d} n_l \kappa_{l2}(n_l/m)}},$$

we obtain

$$
\begin{aligned}
[\mathbf{z}^{\boldsymbol{n}}]\mathbf{g}(\mathbf{z})^m &= \frac{V(\boldsymbol{\rho},\mathbf{n},m)}{(2\pi)^{d/2}} \int\cdots\int_{\tilde{B}} \exp\left(-\sum_{l=1}^{d}\frac{u_l^2}{2}\right) \\
&\times \left(1-\sum_{l=1}^{d}\frac{u_l^3}{3!}n_l^{-1/2}\tilde{\kappa}_{l3}\left(\frac{n_l}{m}\right)+\sum_{l=1}^{d}\frac{u_l^4}{4!}n_l^{-1}\tilde{\kappa}_{l4}\left(\frac{n_l}{m}\right)\right. \\
&\left. + O\left(\sum_{l=1}^{d} u_{l1}^5 n_l^{-3/2}+u_l^6 n_l^{-1}\right)+O\left(m\sum_{l=1}^{d}\rho_l\left|\frac{u_l}{\sqrt{n_l}}\right|^{k+1}\right)\right) du_1\cdots du_d \\
&\sim V(\boldsymbol{\rho},\mathbf{n},m)\left(1+\sum_{l=1}^{d}\tilde{\kappa}_{l4}\left(\frac{n_l}{m}\right)\frac{1}{8n_l}\right)
\end{aligned}
$$

Using more terms this procedure yields a multivariate asymptotic series expansion of the form

$$
[\mathbf{z}^{\boldsymbol{n}}]\mathbf{g}(\mathbf{z})^m \sim V(\boldsymbol{\rho},\mathbf{n},m)\sum_{j_1,\ldots,j_d\geq 0} a_{j_1\cdots j_d}\left(\frac{n_1}{m},\ldots,\frac{n_d}{m}\right) n_1^{-j_1}\cdots n_d^{-j_d}
$$

where $a_{j_1\cdots j_d}(t_1,\ldots,t_{2d})$ are explicitly computable analytic functions.

The next task is analyzing $c_{\boldsymbol{n},\alpha}$ for $\alpha>0$, where we have to cope with the additional factor in (5.20). W.l.o.g. let us assume that the term containing none of the factors $z_1,\ldots,z_d$ vanishes. Then $d_1(\mathbf{z})$ can be represented in the form

$$
d_1(\mathbf{z})=\sum_{l=1}^{d} c_l^{(1)}(\mathbf{z})z_l
$$

with analytic functions $c_l^{(1)}(\mathbf{z})$. Due to the definition of $d_j(\mathbf{z})$ this implies

$$
d_j(\mathbf{z})=\sum_{l=1}^{d} c_l^{(j)}(\mathbf{z})z_l
$$

where $c_l^{(j)}(\mathbf{z})$ are again analytic functions. Hence $c_{\boldsymbol{n},\alpha}$ can be represented as a sum of terms with the shape

$$
m(m-1)\cdots(m-\beta+1)[\mathbf{z}^{\boldsymbol{n}}]\mathbf{g}(\mathbf{z})^m K_\beta(\mathbf{z})
$$

with coefficients independent of $\mathbf{n}$ and $m$. Here $K_\beta(\mathbf{z})$ is an analytic function admitting a representation of the form

$$
K_\beta(\mathbf{z})=\sum_{\substack{\gamma_1,\ldots,\gamma_d\geq 0\\ \sum_j\gamma_j=\beta}} L_{\gamma_1\cdots\gamma_d}(\mathbf{z})\prod_{l=1}^{d} z_l^{\gamma_l}
$$

with an analytic functions $L_{\gamma_1\cdots\gamma_d}(\mathbf{z})$.

For simplicity, assume that the above sum has only one term. Let $L(\mathbf{z}_1,\mathbf{z}_2)$ be the additional factor corresponding to a choice of $\gamma_1,\ldots,\gamma_d$ with $\sum_j\gamma_j=\beta$. Then we have for $\mathbf{z}\in\mathbb{R}^d$

$$L\left(z_1e^{i\theta_1},\dots,z_de^{i\theta_d},z_{12}\right)=L(\mathbf{z})\exp\left(\sum_{j_1+\dots+j_d>0}^{k}\frac{\prod_{l=1}^d(i\theta_l)^{j_l}}{j_l!}\lambda_{\boldsymbol{j_1}}(\mathbf{z})+O\left(\sum_{l=1}^{d}z_l|\theta_l^{k+1}|\right)\right)$$

with

$$\lambda_{\boldsymbol{e}_\mu}(\mathbf{z})=z_\mu\frac{\frac{\partial}{\partial z_\mu}L(\mathbf{z})}{L(\mathbf{z})}\quad\text{for }\mu=1,\dots,d$$
$$\lambda_{\boldsymbol{j_1}+\boldsymbol{e}_\mu}(\mathbf{z})=z_\mu\frac{\partial}{\partial z_\mu}\lambda_{\boldsymbol{j_1}}(\mathbf{z})\quad\text{for }\mu=1,\dots,d$$

where $\mathbf{e}_\mu$ denotes the $\mu$th unit vector in $\mathbb{R}^d$. Thus we can proceed as in the case $\alpha=0$. Set

$$\bar{\lambda}_{\boldsymbol{j_1}}(\mathbf{z})=\frac{\lambda_{\boldsymbol{j_1}}(\mu_1(z_1),\dots,\mu_d(z_d)))}{z_1\cdots z_2}$$

and get

$$\begin{aligned}
&m(m-1)\cdots(m-\beta+1)[\mathbf{z}^{\boldsymbol{n}}]\mathbf{g}(\mathbf{z})^mK_\beta(\mathbf{z})\\
&=\frac{m(m-1)\cdots(m-\beta+1)L(\boldsymbol{\rho})\mathbf{g}(\boldsymbol{\rho})^m}{(2\pi)^d\boldsymbol{\rho}^{\boldsymbol{n}}}\prod_{l=1}^{d}\rho_l^{\gamma_l}\\
&\times\int_B\cdots\int\exp\left(-\sum_{l=1}^{d}\frac{\theta_l^2}{2}n_l\bar{\kappa}_{l2}\left(\frac{n_l}{m}\right)+\sum_{l=1}^{d}\sum_{j=3}^{k}\frac{(i\theta_l)^j}{j!}n_l\bar{\kappa}_{lj}\left(\frac{n_l}{m}\right)\right.\\
&+i\sum_{l=1}^{d}\gamma_l\theta_l+\sum_{j_1+\dots+j_d>0}^{k}\frac{\prod_{l=1}^d(i\theta_l)^{j_l}n_l}{j_l!m^d}\bar{\lambda}_{\boldsymbol{j_1}}\left(\frac{n_1}{m},\dots,\frac{n_d}{m}\right)\\
&\left.+O\left(m\sum_{l=1}^{d}|\rho_l\theta_l^{k+1}|\right)\right)d\theta_1\cdots d\theta_d
\end{aligned}$$

$$\begin{aligned}
&=\frac{m(m-1)\cdots(m-\beta+1)L(\boldsymbol{\rho})V(\boldsymbol{\rho},\mathbf{n},m)}{(2\pi)^{d/2}}\prod_{l=1}^{d}\rho_l^{\gamma_l}\\
&\times\int_{\tilde{B}}\cdots\int\exp\left(-\sum_{l=1}^{d}\frac{u_l^2}{2}+\sum_{l=1}^{d}\sum_{j=3}^{k}\frac{(iu_l)^j}{j!}n_l^{1-j/2}\tilde{\kappa}_{lj}\left(\frac{n_l}{m}\right)+\sum_{l=1}^{d}\gamma_l\frac{u_l}{\sqrt{n_l\bar{\kappa}_{l2}(n_l/m)}}\right.\\
&+\sum_{j_1+\dots+j_d>0}^{k}\frac{\prod_{l=1}^d(iu_l)^{j_l}n_l^{1-j_l/2}}{j_l!m^d}\tilde{\lambda}_{\boldsymbol{j_1}}\left(\frac{n_1}{m},\dots,\frac{n_d}{m}\right)\\
&\left.+O\left(m\sum_{l=1}^{d}\rho_l\left|\frac{u_l}{\sqrt{n_l}}\right|^{k+1}\right)\right)du_1\cdots du_d
\end{aligned}$$

with

$$\tilde{\lambda}_{\boldsymbol{j_1}}(x_1,\dots,x_d)=\bar{\lambda}_{\boldsymbol{j_1}}\bar{\kappa}_{l1}(x_l)^{-j_l/2}.$$

Expanding the exp-term into a series and evaluating the integral yields finally an asymptotic series expansion of the form

$$\begin{aligned}
c_{\boldsymbol{n},\alpha} &\sim \sum_{\beta} m(m-1)\cdots(m-\beta+1)L(\boldsymbol{\rho})V(\boldsymbol{\rho},\mathbf{n},m)\prod_{l=1}^{d}\rho_l^{\gamma_l} \\
&\quad\times \sum_{\substack{j_1,\dots,j_d\geq 0,\\ \delta_1,\dots,\delta_d\geq 0}} a^{(\alpha)}_{j_1\cdots j_d}\left(\frac{n_1}{m},\dots,\frac{n_d}{m}\right) n_1^{-j_1}\cdots n_d^{-j_d} h_1^{-\delta_1}\cdots h_d^{-\delta_d} \\
&\sim \sum_{\beta} \frac{m(m-1)\cdots(m-\beta+1)}{m^{\beta}} L(\boldsymbol{\rho})V(\boldsymbol{\rho},\mathbf{n},m) \\
&\quad\times \sum_{\substack{j_1,\dots,j_d\geq 0,\\ \delta_1,\dots,\delta_d\geq 0}} a^{(\alpha)}_{j_1\cdots j_d}\left(\frac{n_1}{m},\dots,\frac{n_d}{m}\right) \frac{n_1^{\gamma_1}\cdots n_d^{\gamma_d}}{n_1^{j_1}\cdots n_d^{j_d}}
\end{aligned} \tag{5.21}$$

with explicitly computable analytic functions $a^{(\alpha)}_{j_1\cdots j_d}(t_1,\dots,t_d)$ analytic functions.

Inserting this into (5.19) implies that for $m\to\infty$ $\mathbf{E}(X_m(\mathbf{n})-\mathbf{E}X_m(\mathbf{n}))^{2\Delta}$ is asymptotically equal to a rational function in $n_1,\dots,n_d$. If we choose $s_1,\dots,s_d$ fixed and require $n_1=s_1m,\dots,n_d=s_dm$, then by (5.2) we have $\mathbf{E}(X_m(\mathbf{n})-\mathbf{E}X_m(\mathbf{n}))^{2\Delta}=O\left(\|\mathbf{n}\|^{\Delta}\right)$ as desired. Since on the one hand this holds for any choice of $s_1,\dots,s_d$ and on the other hand all terms in (5.21) (and thus in the asymptotic series for $\mathbf{E}(X_m(\mathbf{n})-\mathbf{E}X_m(\mathbf{n}))^{2\Delta}$) have up to constant factors the shape

$$\frac{n_1^{\gamma_1}\cdots n_d^{\gamma_d}}{n_1^{j_1}\cdots n_d^{j_d}}.$$

we must have

$$\gamma_1+\cdots+\gamma_d-j_1-\cdots-j_d\leq\Delta.$$

But as to the fact that $n_1^{\gamma_1'}\cdots n_d^{\gamma_d'}\leq\|\mathbf{n}\|^{\beta}$, if $\gamma_1'+\cdots+\gamma_d'\leq\beta$, the above inequality guarantees the validity of (5.18) for all $\mathbf{n}$ satisfying $\mathbf{n}=O(m)$ and the proof is complete. □

In order to prove tightness, by [17, Ch. XIII, Ex. 1.12] it suffices to show the following Lemma.

**Lemma 5.4.** *Let* $\mathbf{n}=(n_1,\dots,n_d)$ *and* $\mathbf{h}=(h_1,\dots,h_d)$. *Then there exists apositive constant* $C$ *such that*

$$\mathbf{E}\frac{(X_m(\mathbf{n}+\mathbf{h})-X_m(\mathbf{n})-\mathbf{E}(X_m(\mathbf{n}+\mathbf{h})-X_m(\mathbf{n})))^{2d+2}}{m^{d+1}}\leq C\left(\frac{\|\mathbf{h}\|}{m}\right)^{d+1} \tag{5.22}$$

*uniformly for* $\|\mathbf{n}\|=O(m)$ *as* $m\to\infty$.

**Corollary.** *The sequence* $Y_m(t)$ *is tight.*

**Proof.** In order to treat the difference $Z_m(\mathbf{n},\mathbf{h})=X_m(\mathbf{n}+\mathbf{h})-X_m(\mathbf{n})$ we distinguish two cases. If $\|\mathbf{n}\|=O(\|\mathbf{h}\|)$, then set $X_m^c(\mathbf{n}):=X_m(\mathbf{n})-\mathbf{E}X_m(\mathbf{n})$ and use the crude estimate

$$\mathbf{E}Z_m(\mathbf{n},\mathbf{h})^{2d+2}\leq\sum_{k=0}^{d+1}\binom{2d+2}{2k}\mathbf{E}X_m^c(\mathbf{n}+\mathbf{h})^{2k}\mathbf{E}X_m^c(\mathbf{n})^{2d+2-2k} \tag{5.23}$$

in conjunction with Lemma 5.3.

If $\|\mathbf{n}\| = O(\|\mathbf{h}\|)$ does not hold, we may without loss of generality assume that $\|\mathbf{h}\|/\|\mathbf{n}\| \to 0$. We use the generating function that enumerates the change of the valuation between the first and the second batch, i.e., $\Phi_2(1/x, x, z_{11}, \dots, z_{d1}, z_{12}, \dots, z_{d2})$. Set $\mathbf{z}_1 = (z_{11}, \dots, z_{d1})$, $\mathbf{z}_2 = (z_{12}, \dots, z_{d2})$, and

$$c_{\boldsymbol{n},\boldsymbol{h},\alpha} := \left[\mathbf{z}_1^{\boldsymbol{n}} \mathbf{z}_2^{\boldsymbol{h}}\right] \frac{\partial^\alpha}{\partial x^\alpha} \Phi_2\left(\frac{1}{x}, x, \mathbf{z}_1, \mathbf{z}_2\right)\bigg|_{x=1}.$$

Furthermore let

$$A_i := \mathbf{E} \prod_{j=0}^{i-1} (Z_m(\mathbf{n}, \mathbf{h}) - j) = \frac{c_{\boldsymbol{n},\boldsymbol{h},i}}{c_{\boldsymbol{n},\boldsymbol{h},0}}.$$

Then the moment occurring in (5.22) can again be expressed by

$$\mathbf{E}(Z_m(\mathbf{n}, \mathbf{h}) - \mathbf{E} Z_m(\mathbf{n}, \mathbf{h}))^{2d+2} = \sum_{l=0}^{2d+2} \binom{2d+2}{l} (-1)^l A_1^{2d+2-l} \sum_{k=1}^{l} S_{lk} A_k. \tag{5.24}$$

Hence we have to compute $c_{\boldsymbol{n},\boldsymbol{h},\alpha}$. If we set

$$d_j(\mathbf{z}_1, \mathbf{z}_2) = \frac{1}{\mathbf{g}(\mathbf{z}_1)\mathbf{g}(\mathbf{z}_2)} \frac{\partial^j}{\partial x^j} \phi_2\left(\frac{1}{x}, x, \mathbf{z}_1, \mathbf{z}_2\right)\bigg|_{x=1},$$

then Faà di Bruno's formula yields

$$\begin{aligned} c_{\boldsymbol{n},\boldsymbol{h},\alpha} = &\sum_{\sum_j jk_j = \alpha} \frac{\alpha!}{k_1! \cdots k_\alpha!} m(m-1)\cdots(m-k_1-\cdots-k_\alpha+1) \\ &\times [\mathbf{z}_1^{\boldsymbol{n}} \mathbf{z}_2^{\boldsymbol{h}}] \mathbf{g}(\mathbf{z}_1)^m \mathbf{g}(\mathbf{z}_2)^m \prod_{j=1}^{\alpha} \left(\frac{d_j(\mathbf{z}_1, \mathbf{z}_2)}{j!}\right)^{k_j}. \end{aligned}$$

Thus we have to calculate the coefficient

$$\left[\mathbf{z}_1^{\boldsymbol{n}} \mathbf{z}_2^{\boldsymbol{h}}\right] \mathbf{g}(\mathbf{z}_1)^m \mathbf{g}(\mathbf{z}_2)^m \prod_{j=1}^{\alpha} \left(\frac{d_j(\mathbf{z}_1, \mathbf{z}_2)}{j!}\right)^{k_j}. \tag{5.25}$$

The calculation of $c_{\boldsymbol{n},\boldsymbol{h},0} = [\mathbf{z}_1^{\boldsymbol{n}} \mathbf{z}_2^{\boldsymbol{h}}] \mathbf{g}(\mathbf{z}_1)^m \mathbf{g}(\mathbf{z}_2)^m$ is easy since it factorizes. The saddle points of $g_l(z_{l1})^m z_{l1}^{-n_{l1}}$ and $g_l(z_{l2})^m z_{l2}^{-n_{l2}}$ for $l = 1, \dots, d$ are given by

$$\rho_{lj} = \mu_l\left(\frac{n_{lj}}{m}\right) = \frac{g_{l0}}{g_{l1}} \frac{n_{lj}}{m}\left(1 + O\left(\frac{n_{lj}}{m}\right)\right) \qquad \text{for } j = 1, 2.$$

Setting $\boldsymbol{\rho_1} = (\rho_{11}, \dots, \rho_{d1})$, $\boldsymbol{\rho_2} = (\rho_{12}, \dots, \rho_{d2})$ as well as $\mathbf{z}^k := (z_1^k, \dots, z_d^k)$ and

$$V(\boldsymbol{\rho_1}, \boldsymbol{\rho_2}, \mathbf{n}, \mathbf{h}, m) = \frac{\mathbf{g}(\boldsymbol{\rho_1})^m \mathbf{g}(\boldsymbol{\rho_2})^m}{(2\pi)^d \boldsymbol{\rho_1}^{\boldsymbol{n}} \boldsymbol{\rho_2}^{\boldsymbol{h}} \sqrt{\prod_{l=1}^d n_l \kappa_{l2}(n_l/m) h_l \kappa_{l2}(h_l/m)}},$$

we get as above

$$
\begin{aligned}
[\mathbf{z}_1^{\boldsymbol{n}}\mathbf{z}_2^{\boldsymbol{h}}]\mathbf{g}(\mathbf{z}_1)^m\mathbf{g}(\mathbf{z}_2)^m &= \frac{V(\boldsymbol{\rho_1},\boldsymbol{\rho_2},\mathbf{n},\mathbf{h},m)}{(2\pi)^d}\int\cdots\int_{\tilde{B}} \exp\left(-\sum_{l=1}^d \frac{u_{l1}^2}{2} - \sum_{l=1}^d \frac{u_{l2}^2}{2}\right) \\
&\times \left(1 - \sum_{l=1}^d \frac{u_{l1}^3}{3!} n_l^{-1/2}\tilde{\kappa}_{l3}\left(\frac{n_l}{m}\right) - \sum_{l=1}^d \frac{u_{l2}^3}{3!} h_l^{-1/2}\tilde{\kappa}_{l3}\left(\frac{h_l}{m}\right)\right. \\
&+ \sum_{l=1}^d \frac{u_{l1}^4}{4!} n_l^{-1}\tilde{\kappa}_{l4}\left(\frac{n_l}{m}\right) + \sum_{l=1}^d \frac{u_{l2}^4}{4!} h_l^{-1}\tilde{\kappa}_{l4}\left(\frac{h_l}{m}\right) \\
&+ O\left(\sum_{l=1}^d u_{l1}^5 n_l^{-3/2} + u_{l1}^6 n_l^{-1}\right) + O\left(\sum_{l=1}^d u_{l2}^5 h_l^{-3/2} + u_{l2}^6 h_l^{-1}\right) \\
&\left. + O\left(m\sum_{l=1}^d \rho_{l1}\left|\frac{u_{l1}}{\sqrt{n_l}}\right|^{k+1}\right) + O\left(m\sum_{l=1}^d \rho_{l2}\left|\frac{u_{l2}}{\sqrt{h_l}}\right|^{k+1}\right)\right) du_{11}\cdots du_{d1}\, du_{12}\cdots du_{d2} \\
&\sim V(\boldsymbol{\rho_1},\boldsymbol{\rho_2},\mathbf{n},\mathbf{h},m)\left(1 + \sum_{l=1}^d \tilde{\kappa}_{l4}\left(\frac{n_l}{m}\right)\frac{1}{8n_l} + \sum_{l=1}^d \tilde{\kappa}_{l4}\left(\frac{h_l}{m}\right)\frac{1}{8h_l}\right)
\end{aligned}
$$

Using more terms we get a multivariate asymptotic series expansion of the form

$$
\begin{aligned}
&[\mathbf{z}_1^{\boldsymbol{n}}\mathbf{z}_2^{\boldsymbol{h}}]\mathbf{g}(\mathbf{z}_1)^m\mathbf{g}(\mathbf{z}_2)^m \\
&\quad\sim V(\boldsymbol{\rho_1},\boldsymbol{\rho_2},\mathbf{n},\mathbf{h},m) \sum_{\substack{j_1,\ldots,j_d\ge 0\\ \delta_1,\ldots,\delta_d\ge 0}} a_{j_1\cdots j_d\delta_1\cdots\delta_d}\left(\frac{n_1}{m},\ldots,\frac{n_d}{m},\frac{h_1}{m},\ldots,\frac{h_d}{m}\right) n_1^{-j_1}\cdots n_d^{-j_d} h_1^{-\delta_1}\cdots h_d^{-\delta_d}
\end{aligned}
$$

where $a_{j_1\cdots j_d,\delta_1\cdots\delta_d}(t_1,\ldots,t_{2d})$ are explicitly computable analytic functions.

Now we turn to $c_{\boldsymbol{n},\boldsymbol{h},\alpha}$ for $\alpha > 0$. Therefore we first analyze the additional factor occurring in (5.25). By (2.8):

$$
\phi_2(x_1,x_2,\mathbf{z}_1,\mathbf{0}) = \phi_1(x_1x_2,\mathbf{z}_1)
$$

we obtain

$$
\frac{\partial}{\partial x_1}\phi_2(x_1,x_2,\mathbf{z}_1,\mathbf{0}) - \frac{\partial}{\partial x_2}\phi_2(x_1,x_2,\mathbf{z}_1,\mathbf{0})\bigg|_{x_1=x_2=1} = (x_2-x_1)\frac{\partial}{\partial x}\phi_1(x,\mathbf{z}_1)\bigg|_{x_1=x_2=1} = 0
$$

and thus

$$
d_1(\mathbf{z}_1,\mathbf{z}_2) = \frac{\partial}{\partial x_2}\phi_2(x_1,x_2,\mathbf{z}_1,\mathbf{z}_2) - \frac{\partial}{\partial x_1}\phi_2(x_1,x_2,\mathbf{z}_1,\mathbf{z}_2)\bigg|_{x_1=x_2=1} = \sum_{l=1}^d c_l^{(1)}(\mathbf{z}_1,\mathbf{z}_2)z_{l2}
$$

with analytic functions $c_l^{(1)}(\mathbf{z}_1,\mathbf{z}_2)$. As in the proof of Lemma 5.3, the definition of $d_j(\mathbf{z}_1,\mathbf{z}_2)$ guarantees that there exist analytic functions $c_l^{(j)}(\mathbf{z}_1,\mathbf{z}_2)$ such that

$$
d_j(\mathbf{z}_1,\mathbf{z}_2) = \sum_{l=1}^d c_l^{(j)}(\mathbf{z}_1,\mathbf{z}_2)z_{l2}
$$

Hence $c_{\boldsymbol{n},\boldsymbol{h},\alpha}$ can be represented as a sum of terms with the shape

$$
m(m-1)\cdots(m-\beta+1)[\mathbf{z}_1^{\boldsymbol{n}}\mathbf{z}_2^{\boldsymbol{h}}]\mathbf{g}(\mathbf{z}_1)^m\mathbf{g}(\mathbf{z}_2)^m K_\beta(\mathbf{z}_1,\mathbf{z}_2)
$$

with coefficients independent of $\mathbf{n}$, $\mathbf{h}$, and $m$ and an analytic function $K_\beta(\mathbf{z}_1, \mathbf{z}_2)$ of the form

$$K_\beta(\mathbf{z}_1, \mathbf{z}_2) = \sum_{\substack{\gamma_1, \ldots, \gamma_d \geq 0 \\ \sum_j \gamma_j = \beta}} L_{\gamma_1 \cdots \gamma_d}(\mathbf{z}_1, \mathbf{z}_2) \prod_{l=1}^{d} z_{l2}^{\gamma_l}$$

where $L_{\gamma_1 \cdots \gamma_d}(\mathbf{z}_1, \mathbf{z}_2)$ is again analytic.

As above we assume that this sum has only one term, denoted by $L(\mathbf{z}_1, \mathbf{z}_2)$ and corresponding to a choice of $\gamma_1, \ldots, \gamma_d$ with $\sum_j \gamma_j = \beta$. Then we have for $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^d$

$$\begin{aligned} L\left(z_{11} e^{i\theta_{11}}, \ldots, z_{d1} e^{i\theta_{d1}}, z_{12} e^{i\theta_{12}}, \ldots, z_{d2} e^{i\theta_{d2}}\right) &= L(\mathbf{z}_1, \mathbf{z}_2) \\ \times \exp\left( \sum_{j_{11} + \cdots + j_{d1} + j_{12} + \cdots + j_{d2} > 0}^{k} \frac{\prod_{l=1}^{d} (i\theta_{l1})^{j_{l1}} (i\theta_{l2})^{j_{l2}}}{j_{l1}! j_{l2}!} \lambda_{\boldsymbol{j_1 j_2}}(\mathbf{z}_1, \mathbf{z}_2) \right. \\ \left. + O\left( \sum_{l=1}^{d} z_{l1} |\theta_{l1}^{k+1}| + \sum_{l=1}^{d} z_{l2} |\theta_{l2}^{k+1}| \right) \right) \end{aligned}$$

with

$$\begin{aligned} \lambda_{\boldsymbol{e}_\mu}(\mathbf{z}_1, \mathbf{z}_2) &= z_{\mu 1} \frac{\frac{\partial}{\partial z_{\mu 1}} L(\mathbf{z}_1, \mathbf{z}_2)}{L(\mathbf{z}_1, \mathbf{z}_2)} \quad \text{for } \mu = 1, \ldots, d \\ \lambda_{\boldsymbol{e}_\mu}(\mathbf{z}_1, \mathbf{z}_2) &= z_{\mu 2} \frac{\frac{\partial}{\partial z_{\mu 2}} L(\mathbf{z}_1, \mathbf{z}_2)}{L(\mathbf{z}_1, \mathbf{z}_2)} \quad \text{for } \mu = d+1, \ldots, 2d \\ \lambda_{(\boldsymbol{j_1}, \boldsymbol{j_2}) + \boldsymbol{e}_\mu}(\mathbf{z}_1, \mathbf{z}_2) &= z_{\mu 1} \frac{\partial}{\partial z_{\mu 1}} \lambda_{\boldsymbol{j_1}, \boldsymbol{j_2}}(\mathbf{z}_1, \mathbf{z}_2) \quad \text{for } \mu = 1, \ldots, d \\ \lambda_{(\boldsymbol{j_1}, \boldsymbol{j_2}) + \boldsymbol{e}_\mu}(\mathbf{z}_1, \mathbf{z}_2) &= z_{\mu 2} \frac{\partial}{\partial z_{\mu 2}} \lambda_{\boldsymbol{j_1}, \boldsymbol{j_2}}(\mathbf{z}_1, \mathbf{z}_2) \quad \text{for } \mu = d+1, \ldots, 2d \end{aligned}$$

where this time $\mathbf{e}_\mu$ denotes the $\mu$th unit vector in $\mathbb{R}^{2d}$. Thus we can proceed as in the case $\alpha = 0$. Set

$$\bar{\lambda}_{\boldsymbol{j_1}, \boldsymbol{j_2}}(\mathbf{z}_1, \mathbf{z}_2) = \frac{\lambda_{\boldsymbol{j_1}, \boldsymbol{j_2}}(\mu_1(z_{11}), \ldots, \mu_d(z_{d1}), \mu_1(z_{12}), \ldots, \mu_d(z_{d2}))}{z_{11} \cdots z_{d1} z_{12} \cdots z_{d2}}$$

and get (cf. previous Lemma)

$$= \frac{m(m-1)\cdots(m-\beta+1)L(\boldsymbol{\rho_1},\boldsymbol{\rho_2})V(\boldsymbol{\rho_1},\boldsymbol{\rho_2},\mathbf{n},\mathbf{h},m)}{(2\pi)^d}\prod_{l=1}^d \rho_{l2}^{\gamma_l}$$

$$\times \int_{\tilde{B}}\cdots\int \exp\left(-\sum_{l=1}^d \frac{u_{l1}^2}{2} - \sum_{l=1}^d \frac{u_{l2}^2}{2} + \sum_{l=1}^d\sum_{j=3}^k \frac{(iu_{l1})^j}{j!} n_l^{1-j/2}\tilde{\kappa}_{lj}\left(\frac{n_l}{m}\right)\right.$$

$$+ \sum_{l=1}^d\sum_{j=3}^k \frac{(iu_{l2})^j}{j!} h_l^{1-j/2}\tilde{\kappa}_{lj}\left(\frac{h_l}{m}\right) + \sum_{l=1}^d \gamma_l \frac{u_{l2}}{\sqrt{h_l\bar{\kappa}_{l2}(h_l/m)}}$$

$$+ \sum_{j_{11}+\cdots+j_{d1}+j_{12}+\cdots+j_{d2}>0}^{k} \frac{\prod_{l=1}^d (iu_{l1})^{j_{l1}}(iu_{l2})^{j_{l2}} n_l^{1-j_{l1}/2} h_l^{1-j_{l2}/2}}{j_{l1}!j_{l2}!m^{2d}} \tilde{\lambda}_{\boldsymbol{j_1 j_2}}\left(\frac{n_1}{m},\ldots,\frac{n_d}{m},\frac{h_1}{m},\ldots,\frac{h_d}{m}\right)$$

$$\left. + O\left(m\sum_{l=1}^d \rho_{l1}\left|\frac{u_{l1}}{\sqrt{n_l}}\right|^{k+1}\right) + O\left(m\sum_{l=1}^d \rho_{l2}\left|\frac{u_{l2}}{\sqrt{h_l}}\right|^{k+1}\right)\right) du_{11}\cdots du_{d1}\, du_{12}\cdots du_{d2}$$

with

$$\tilde{\lambda}_{\boldsymbol{j_1 j_2}}(x_1,\ldots,x_{2d}) = \bar{\lambda}_{\boldsymbol{j_1 j_2}}\bar{\kappa}_{l1}(x_l)^{-j_{l1}/2}\bar{\kappa}_{l2}(x_{d+l})^{-j_{l2}/2}.$$

Expanding the exp-term into a series and evaluating the integral yields finally an asymptotic series expansion of the form

$$c_{\boldsymbol{n},\boldsymbol{h},\alpha} \sim \sum_\beta \frac{m(m-1)\cdots(m-\beta+1)}{m^\beta} L(\boldsymbol{\rho_1},\boldsymbol{\rho_2})V(\boldsymbol{\rho_1},\boldsymbol{\rho_2},\mathbf{n},\mathbf{h},m)\prod_{l=1}^d h_l^{\gamma_l}$$

(5.26)

$$\times \sum_{\substack{j_1,\ldots,j_d\geq 0,\\ \delta_1,\ldots,\delta_d\geq 0}} a^{(\alpha)}_{j_1\cdots j_d,\delta_1,\cdots,\delta_d}\left(\frac{n_1}{m},\ldots,\frac{n_d}{m},\frac{h_1}{m},\ldots,\frac{h_d}{m}\right) n_1^{-j_1}\cdots n_d^{-j_d} h_1^{-\delta_1}\cdots h_d^{-\delta_d}$$

with explicitly computable analytic functions $a^{(\alpha)}_{j_1\cdots j_d,\delta_1\cdots\delta_d}(t_1,\ldots,t_{2d})$ analytic functions.

Arguing as in the proof of the previous Lemma, we choose arbitrary constants $s_1,\ldots,s_d$ and $t_1,\ldots,t_d$ and require $n_1 = s_1 m,\ldots,n_d = s_d m$ and $h_1 = t_1 m,\ldots,h_d = t_d m$. Then by Lemma 5.2 we have $\mathbf{E}Z_m(\mathbf{n},\mathbf{h})^{2d+2} = O\left(\|\mathbf{h}\|^{d+1}\right)$. On the other hand, inserting (5.26) into (5.24), shows that $\mathbf{E}Z_m(\mathbf{n},\mathbf{h})^{2d+2}$ is for $m\to\infty$ asymptotically equal to a rational function in $n_1,\ldots,n_d,h_1,\ldots,h_d$ all terms of which have the shape

$$\frac{h_1^{\gamma_1}\cdots h_d^{\gamma_d}}{n_1^{j_1}\cdots n_d^{j_d}h_1^{\delta_1}\cdots h_1^{\delta_d}} \tag{5.27}$$

if we neglect constant factors. Thus Lemma 5.2 implies

$$\gamma_1+\cdots+\gamma_d-\delta_1-\cdots-\delta_d-j_1-\cdots-j_d\leq d+1$$

and (5.27) can be rewritten as

$$h_1^{\gamma_1'} \cdots h_d^{\gamma_d'} \frac{h_1^{j_1} \cdots h_d^{j_d}}{n_1^{j_1} \cdots n_d^{j_d}}$$

with $\gamma_1' + \cdots + \gamma_d' \leq d+1$. Assume without loss of generality that equality holds. If $\prod_{i=1}^d (h_i/n_i)^{j_i} = O(1)$ then

$$h_1^{\gamma_1'} \cdots h_d^{\gamma_d'} \frac{h_1^{j_1} \cdots h_d^{j_d}}{n_1^{j_1} \cdots n_d^{j_d}} \leq \|h\|^{d+1}$$

and we would be finished. If $\prod_{i=1}^d (h_i/n_i)^{j_i}$ is not bounded, then we may assume $\prod_{i=1}^d (h_i/n_i)^{j_i} \to \infty$. In this case set $h_i = t_i m$, for $i = 1, \ldots, d$, with $t_i$ lying in an interval bounded away from zero. On the one hand, this implies the existence of a positive constant $C$ such that $h_1^{\gamma_1'} \cdots h_d^{\gamma_d'} \geq C\|h\|^{d+1}$ and consequently

$$h_1^{\gamma_1'} \cdots h_d^{\gamma_d'} \frac{h_1^{j_1} \cdots h_d^{j_d}}{n_1^{j_1} \cdots n_d^{j_d}} \gg \|h\|^{d+1} \tag{5.28}$$

since we still have $\prod_{i=1}^d (h_i/n_i)^{j_i} \to \infty$. On the other hand, by $\|n\| = O(m)$ we have now $\|n\| = O(\|h\|)$ and thus (5.28) contradicts the conclusion of (5.23) and Lemma 5.3. □

**Thanks** are due to P. Chassaing, A. Mokkadem and A. Rouault for information on multivariate processes, and to J.F. Marckert for plotting the bi-dimensional Gaussian processes.

## References

[1] E. A. Bender and L. B. Richmond, Central and local limit theorems applied to asymptotic enumeration II: multivariate generating functions, *J. Comb. Theory, Ser. A* 34, 255–265, 1983.

[2] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[3] S. Boucheron and D. Gardy, An urn model from learning theory, *Random Struct. Alg.* 10, 43–67, 1997.

[4] L. Comtet, *Advanced combinatorics. The art of finite and infinite expansions*, D. Reidel Publishing Co., Dordrecht, 1974.

[5] M. Drmota, A bivariate asymptotic expansion of coefficients of powers of generating functions, *European Journal of Combinatorics* 15, 1994, 139-152.

[6] M. Drmota, D. Gardy, and B. Gittenberger, A unified presentation of some urn models, *Algorithmica,* Special issue on the average-case analysis of algorithms, Vol. 29 No. 1/2, January/February 2001, pp. 120-147.

[7] P. Flajolet and J. S. Vitter, Average-Case Analysis of Algorithms and Data Structures, in *Handbook of Theoretical Computer Science,* J. van Leeuwen, Ed., vol. A: Algorithms and Complexity. North Holland, 1990, ch. 9, 431–524.

[8] D. Gardy, Normal limiting distributions for projection and semijoin sizes, *SIAM Journal on Discrete Mathematics,* 5(2), 219-248, 1992.

[9] D. Gardy, Join sizes, urn models and normal limiting distributions, *Theoretical Computer Science (A),* 131, 375-414, 1994.

[10] D. Gardy, Enumeration of bicolored objects according to their color balance, FPSAC'99, Eleventh Colloquium on Formal Power Series and Algebraic Combinatorics, C. Martinez, M. Noy and O. Serra, Eds, 213-224, Barcelona (Spain), June 1999.

[11] D. Gardy and G. Louchard, Dynamic analysis of some relational data bases parameters, *Theoretical Computer Science (A), (special issue on mathematical analysis of algorithms),* 144(1-2), 125-159, 1995.

[12] I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*, Wiley, New York, 1983.

[13] N.L. Johnson and S. Kotz, *Urn models and their application,* Wiley & Sons, 1977.

[14] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus,* Springer, New York, 1988.
[15] V. F. Kolchin, B. Sevast'yanov, and V. Chistyakov, *Random Allocations,* Wiley, New York, 1978.
[16] J. Neveu, *Processus aléatoires gaussiens,* Presses de l'Université de Montréal, 1968.
[17] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion,* Springer, 1991.
[18] A.W. van der Vaart and J.A. Wellner, *Weak Convergence and Empirical Processes,* Springer.