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 with 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 specific random allocation models that have been proposed to study the dynamical behaviour of relational databases. In particular, 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 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 in [6] we gave an analysis of the asymptotic process in a restricted dynamic case (where balls are added one at a time and no deletions are allowed). The present paper has its origin in more involved models which are related to *joins,* or where deletions in the database are allowed.

The join operations in a database are basically obtained by making the cartesian product of two tables and applying a restriction on the result. Let us assume that we have two tables $T_1[X, Y]$ and $T_2[X, Z]$, each with two columns : The equijoin of T_1 and T_2 might be defined as the cartesian product $T_1 \times T_2$, restricted to keep only those quadruples (x_1, y, x_2, z) such that $x_1 = x_2$. The semijoin of T_1 with T_2 is (best) defined as the set of couples (x, y) such that there exists some couple (x, t) in T_2. The importance of the equijoin comes from the fact that it allows to build "new" data from data already present in the database; however equijoins are prone to creating large tables, which is not recommanded if one desires the database operations to be executed quickly. Semijoins appear when selecting data to be transmitted from one place to an other, in a database distributed over several places. In both cases, it is important that the database optimization system, which can rewrite a query from the end user in several ways, and must then choose a "best" way, evaluates the sizes of the tables created by a join operation.

Roughly speaking, the modeling of join sizes by urn models is as follows (see [8, 9] for the precise definitions and models). Let us consider a table $T_1(X, Y)$, which will be joined to a table $T_2(X, Z)$ on column X. The values on Y have no influence on the join, as long as they belong to the relevant domain with the right (for the underlying database problem) probability distribution and there are no repetitions. Hence we deal with some number of distinct X values among all the possible values for X (in a database context, there are usually a finite, if large, number of such possible values) and their numbers of occurences. Now let us consider a sequence of urns, labelled by the possible values for X : We associate to each tuple (x, y) a ball that goes into the urn labelled by x. We can do this again for the next table $T_2(X, Z)$, using balls of a different type. The numbers of balls of each type are exactly the sizes (i.e. numbers of rows) of the initial tables T_1 and T_2. Usually, these sizes of tables are parameters of the database, or at least can be known precisely

Date: February 20, 2003.
* 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.

(there is no randomness there). Finally, we represent each tuple of the (equi- or semi-)join by a ball of a third type, according to the rules given below (from the definition of the join operations, we can build tuples for the join by considering the X values separately, i.e. by taking each urn in turn and investigating its contents). The number of balls of the last type is precisely the join size that needs to be evaluated.

Such urn models have turned out to be of interest of their own as combinatorial objects; they can also be applied to completely different fields, e.g. to biological problems etc.[13]

A mathematical formulation 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 jth urn, then we have

$$X_m = \sum_{j=1}^{m} f(K_{1j}, \ldots, K_{dj}), \tag{1.1}$$

where we condition on $\sum_{j=1}^{m} K_{ij} = n_i$, $i = 1, 2, \ldots, d$. This formulation allows us to present a unified treatment of both the join models and of several urn models previously encountered :

- Semijoin and equijoin models in dynamical databases, where we have two types of balls and the valuation is the join size :

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

  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 the (normalized) process $X_m = X_m(n_1, \ldots, n_d)$ with a specified number n_i of balls of each type $i = 1, 2, \ldots, d$ converges weakly towards a Gaussian field (with *time* variables n_i/m), whose covariance function can be explicitly computed.

In fact, our main result (Theorem 2.1) is even more general. It just refers to properties of corresponding generating function defining the process. For example, this result can be also applied to model deletion of balls. The source of our interest in such a model comes again from databases, that are now *dynamic,* i.e. the user can add or delete items.

The plan of the paper is as follows. In Section 2 we show that the above urn model can be encoded in terms of generating functions and we formulate our main result concerning the convergence of X_m towards a Gaussian field. We study several examples in Section 3 (join and balanced urns models); for example the equijoin leads to 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.

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

We assume that there are d types of balls which are thrown into m urns. First let us consider just one urn and let $a_{n_1,n_2,\dots,n_d}$ denote the number of ways n_i balls of type $i=1,2,\dots,d$ can be allocated in one urn. Then the exponential generating function[1] describing the allocation of balls in one urn and marking the valuation of this urn with x is given by

$$\phi_1(x,z_1,\dots,z_d)=\sum_{n_1,\dots,n_d\ge 0}\frac{a_{n_1,n_2,\dots,n_d}}{n_1!n_2!\cdots n_d!}x^{f(n_1,\dots,n_d)}z_1^{n_1}\cdots z_d^{n_d}.$$

In the *standard model* one has $a_{n_1,n_2,\dots,n_d}=1$ and hence the function

$$\phi_1(1,z_1,\dots,z_d)=e^{z_1}e^{z_2}\cdots e^{z_d}$$

splits into a product of exponential functions. Another example – which is frequently used in this paper – is

$$a_{n_1,n_2,\dots,n_d}=\prod_{i=1}^{d}\delta_i(\delta_i-1)\cdots(\delta_i-n_i+1),$$

which means that every urn has exactly δ_i possible places for balls of type $i=1,2,\dots,d$. Here we get

$$\phi_1(1,z_1,\dots,z_d)=(1+z_1)^{\delta_1}(1+z_2)^{\delta_2}\cdots(1+z_d)^{\delta_d}.$$

Note that in general there are no factorizations like that.

If we denote (as above) $X_m(n_1,n_2,\dots,n_d)$ the (additive) value of these m urns, where n_i balls of type i, $1\le i\le d$, have been thrown, then by additivity we have

$$\mathbf{E}\left(x^{X_m(n_1,n_2,\dots,n_d)}\right)=\frac{[z_1^{n_1}\cdots z_d^{n_d}]\phi_1(x,z_1,\dots,z_d)^m}{[z_1^{n_1}\cdots z_d^{n_d}]\phi_1(1,z_1,\dots,z_d)^m}.$$

In a similar way we can also consider the joint distribution of the valuations of $X_m(\mathbf{n}_1), X_m(\mathbf{n}_1+\mathbf{n}_2),\dots,X_m(\mathbf{n}_1+\mathbf{n}_2+\dots+\mathbf{n}_b)$ for some $b\ge 1$, where $\mathbf{n}_j=(n_{1j},\dots,n_{dj})$, $j=1,\dots,b$. Let $a_{\mathbf{n}_1,\mathbf{n}_2,\dots,\mathbf{n}_b}$ denote the number of ways to allocate first n_{i1} balls of type $i=1,2,\dots,d$, then n_{i2} balls of type $i=1,2,\dots,d$ etc. and set

$$\begin{aligned}&\phi_b(x_1,x_2,\dots,x_b;\mathbf{z}_1,\dots,\mathbf{z}_b)\\ &=\sum_{\substack{n_{ij}\ge 0\\ 1\le i\le d,1\le j\le b}}\prod_{j=1}^{b}\left(x_j^{f(n_{11}+\dots+n_{1j},n_{21}+\dots+n_{2j},\dots,n_{d1}+\dots+n_{dj})}\prod_{i=1}^{d}\frac{a_{\mathbf{n}_1,\dots,\mathbf{n}_b}}{n_{1j}!\cdots n_{1j}!}z_{ij}^{n_{ij}}\right)\end{aligned}\tag{2.1}$$

with $\mathbf{z_j}=(z_{1j},\dots,z_{dj})$. Then we have

$$\begin{aligned}&\mathbf{E}\left(x_1^{X_m(\mathbf{n}_1)}x_2^{X_m(\mathbf{n}_1+\mathbf{n}_2)}\cdots x_d^{X_m(\mathbf{n}_1+\mathbf{n}_2+\dots+\mathbf{n}_b)}\right)\\ &\qquad=\frac{[\mathbf{z}_1^{\mathbf{n}_1}\cdots\mathbf{z}_d^{\mathbf{n}_b}]\phi_b(x_1,x_2,\dots,x_b;\mathbf{z}_1,\dots,\mathbf{z}_b)^m}{[\mathbf{z}_1^{\mathbf{n}_1}\cdots\mathbf{z}_d^{\mathbf{n}_b}]\phi_b(1,1,\dots,1;\mathbf{z}_1,\dots,\mathbf{z}_b)^m}.\end{aligned}\tag{2.2}$$

For example, for the standard model we get (for $x_1=\cdots x_b=1$)

$$\phi_b(1,\dots,1;\mathbf{z_1},\dots,\mathbf{z_b})=\prod_{j=1}^{b}\prod_{i=1}^{d}e^{z_{ij}}=\prod_{j=1}^{b}\phi_1(1,\mathbf{z_j}).$$

For the second mentioned model we have a nice representation, too, (for $x_1=\cdots x_b=1$)

$$\phi_b(1,\dots,1;\mathbf{z}_1,\dots,\mathbf{z}_b)=\prod_{i=1}^{d}(1+z_{i1}+z_{i2}+\dots+z_{ib})^{\delta_i},$$

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

but we do not have a factorization of the form $\phi_b = \phi_1 \cdots \phi_1$.

2.2. **Main Result.** The nature of $\Phi_1(x, z_1, \dots, z_d)$ (i.e., an m-th power) allows a straightforward application of proper limit theorems (e.g. Bender and Richmond [1]), which directly show that $(X_m - \mathbf{E}\, X_m)/\sqrt{\mathbf{Var} X_m}$ has a Gaussian limiting distribution where expected value $\mathbf{E}X_m(n_1, \dots, n_d)$ and variance $\mathbf{Var}X_m(n_1, \dots, n_d)$ are both of oder m (if n_i and m are proportional). 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 which need not be related to urn models.

Theorem 2.1. *Let $X_m = X_m(n_1, \dots, n_d)$ ($m \ge 1$, $n_i \ge 0$ integers) be a sequence of discrete stochastic processes, such that for every $b \ge 1$ there exist functions*

$$\phi_b(x_1, x_2, \dots , x_b; \mathbf{z}_1, \dots , \mathbf{z}_b)$$

which are analytic for $\mathbf{z_j} = (z_{1j}, \dots , z_{dj})$ around $\mathbf{0}$ and $2d + 2$ times continuously differentiable with respect to $(x_1, \dots , x_d)$ such that

$$\begin{aligned}\mathbf{E}&\left(x_1^{X_m(\mathbf{n}_1)} x_2^{X_m(\mathbf{n}_1+\mathbf{n}_2)} \cdots x_d^{X_m(\mathbf{n}_1+\mathbf{n}_2+\cdots+\mathbf{n}_b)}\right)\\ &= \frac{[\mathbf{z}_1^{\mathbf{n}_1} \cdots \mathbf{z}_d^{\mathbf{n}_b}]\phi_b(x_1, x_2, \dots , x_b; \mathbf{z}_1, \dots , \mathbf{z}_b)^m}{[\mathbf{z}_1^{\mathbf{n}_1} \cdots \mathbf{z}_d^{\mathbf{n}_b}]\phi_b(1, 1, \dots , 1; \mathbf{z}_1, \dots , \mathbf{z}_b)^m}\end{aligned} \tag{2.3}$$

and

$$\phi_2(x_1, x_2; \mathbf{z}_1, \mathbf{0}) = \phi_1(x_1 x_2, \mathbf{z_1}). \tag{2.4}$$

as well as

$$[\mathbf{z}_1^{\mathbf{n}_1} \cdots \mathbf{z}_d^{\mathbf{n}_b}]\phi_b(1, 1, \dots , 1; \mathbf{z}_1, \dots , \mathbf{z}_b)^m > 0 \tag{2.5}$$

for all $\mathbf{n}_i \ge \mathbf{0}$.

Then there exists a centered and continuous Gaussian field $G(\mathbf{t})$, $\mathbf{t} = (t_1, \dots t_d) \in \mathbf{T}^\circ$, (where $\mathbf{0} \in \mathbf{T} \subseteq \mathbb{R}^d$ is a proper connected set, see below) such that the following functional limit theorem holds:

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

If $B_{\mathbf{s},\mathbf{t}}$ denotes the covariance function of $G(\mathbf{t})$ then

$$\mathbf{Cov}(X_m(n_1, \dots , n_d), X_m(\tilde{n}_1, \dots , \tilde{n}_d)) = m\, B_{n_1/m,\dots,n_d/m;\tilde{n}_1/m,\dots,\tilde{n}_d/m} + O\left(1\right)$$

uniformly for $m, n_i, \tilde{n}_i \to \infty$ such that n_i/m resp. $\tilde{n}_i/m$ are contained in a fixed compact set contained in $\mathbf{T}^\circ$. Furthermore, there exists a continuous function $\mu_\mathbf{t}$, $\mathbf{t} \in \mathbf{T}^\circ$ such that

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

uniformly for $m, n_i \to \infty$ such that n_i/m are contained in a fixed compact set contained in $\mathbf{T}^\circ$.

Remark 1. We want to mention that the univariate case ($d = b = 1$) for the standard model (i.e. $\phi_1(1, z) = e^z$) has appeared quite early in the literature, e.g. by Quine and Robinson [17]. They proved a (univariate) central limit theorem for $X_m(n)$ under very general moment conditions (which are much weaker than our analyticity conditions). Their method is based on the observation that $X_m(n)$ may be considered as the sum $\sum_{j=1}^m f(U_j(n/m))$ conditioned on $\sum_{j=1}^m U_j(n/m) = n$, where $U_j(t)$ denote independent Poisson random variables with parameter t. (With help of this interpretation it is also quite easy to interprete mean and variance of $X_m(n)$ in terms of moments of $U_j(t)$ resp. of $f(U_j(t))$, compare with [17]).

In order to describe the Gaussian field $G(\mathbf{t})$ in Theorem 2.1 we just have to provide the covariance funcion $B_{\mathbf{s},\mathbf{t}}$ and the set $\mathbf{T}$. The formulas for $B_{\mathbf{s},\mathbf{t}}$ (and $\mu_{\mathbf{t}}$) we present here depend on *saddle point equations* (2.6) and (2.7) (resp. on (2.10)) and do not explicitly refer to the distribution of K_{ij} as it has been done in [17].[2]

Let $\Phi_2(x_1, x_2; \mathbf{z}_1, \mathbf{z}_2)$ be given. For $\mathbf{s} = (s_1, \dots, s_d)$ and $\mathbf{t} = (t_1, \dots, t_d)$ with $0 \le s_i < t_i$, $i = 1, 2, \dots, d$, let $\boldsymbol{\rho}_1 = \boldsymbol{\rho}_1(\mathbf{s}, \mathbf{t}) = (\rho_{11}, \dots, \rho_{d1})$ and $\boldsymbol{\rho}_2 = \boldsymbol{\rho}_2(\mathbf{s}, \mathbf{t}) = (\rho_{12}, \dots, \rho_{d2})$ be defined by

$$\rho_{i1} \frac{\partial \phi_2(1, 1, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2)}{\partial z_{i1}} = s_i \phi_2(1, 1, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2), \qquad i = 1, \dots, d, \tag{2.6}$$

and by

$$\rho_{i2} \frac{\partial \phi_2(1, 1, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2)}{\partial z_{i2}} = (t_i - s_i) \phi_2(1, 1, \boldsymbol{\rho}_1, \boldsymbol{\rho}_2), \qquad i = 1, \dots, d. \tag{2.7}$$

We will denote by $\mathbf{T}$ the set of all $\mathbf{t}$ such that $\boldsymbol{\rho}_1(\mathbf{s}, \mathbf{t})$ and $\boldsymbol{\rho}_2(\mathbf{s}, \mathbf{t})$ exists for all $\mathbf{s}$ with $\mathbf{0} \le \mathbf{s} < \mathbf{t}$.

Now set

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

where $a, b \in \{u_1, u_2, v_1, \dots, v_d, w_1, \dots, w_d\}$), and we obtain

$$B_{\mathbf{s},\mathbf{t}} = \frac{\begin{vmatrix} \kappa_{u_1 u_2} & \kappa_{u_1 v_1} & \cdots & \kappa_{u_1 v_d} & \kappa_{u_1 w_1} & \cdots & \kappa_{u_1 w_d} \\ \kappa_{v_1 u_2} & \kappa_{v_1 v_1} & \cdots & \kappa_{v_1 v_d} & \kappa_{v_1 w_1} & \cdots & \kappa_{v_1 w_d} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \kappa_{v_d u_2} & \kappa_{v_d v_1} & \cdots & \kappa_{v_d v_d} & \kappa_{v_d w_1} & \cdots & \kappa_{v_d w_d} \\ \kappa_{w_1 u_2} & \kappa_{w_1 v_1} & \cdots & \kappa_{w_1 v_d} & \kappa_{w_1 w_1} & \cdots & \kappa_{w_1 w_d} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \kappa_{w_d u_2} & \kappa_{w_d v_1} & \cdots & \kappa_{w_d v_d} & \kappa_{w_d w_1} & \cdots & \kappa_{w_d w_d} \end{vmatrix}}{\begin{vmatrix} \kappa_{v_1 v_1} & \cdots & \kappa_{v_1 v_d} & \kappa_{v_1 w_1} & \cdots & \kappa_{v_1 w_d} \\ \vdots & & \vdots & \vdots & & \vdots \\ \kappa_{v_d v_1} & \cdots & \kappa_{v_d v_d} & \kappa_{v_d w_1} & \cdots & \kappa_{v_d w_d} \\ \kappa_{w_1 v_1} & \cdots & \kappa_{w_1 v_d} & \kappa_{w_1 w_1} & \cdots & \kappa_{w_1 w_d} \\ \vdots & & \vdots & \vdots & & \vdots \\ \kappa_{w_d v_1} & \cdots & \kappa_{w_d v_d} & \kappa_{w_d w_1} & \cdots & \kappa_{w_d w_d} \end{vmatrix}}. \tag{2.8}$$

For general $\mathbf{s}, \mathbf{t} \in \mathbf{T}$ we set $B_{\mathbf{s},\mathbf{t}} = B_{\min(\mathbf{s},\mathbf{t}),\max(\mathbf{s},\mathbf{t})}$.

Furthermore, we have

$$\mu_{\mathbf{s}} = \frac{\frac{\partial}{\partial x} \phi_1(1, \boldsymbol{\rho}(\mathbf{s}))}{\phi_1(1, \boldsymbol{\rho}(\mathbf{s}))}, \tag{2.9}$$

in which $\boldsymbol{\rho}(\mathbf{s}) = (\rho_1, \dots, \rho_d)$ denotes the solution of the equation

$$\rho_i \frac{\partial \phi_1(1, \boldsymbol{\rho})}{\partial z_i} = s_i \phi_1(1, \boldsymbol{\rho}), \qquad i = 1, \dots, d. \tag{2.10}$$

Remark 2. Note that the covariance function $B_{\mathbf{s},\mathbf{t}}$ is just defined if $s_i \ne t_i$ for all $i = 1, 2, \dots, d$. However, we will see in the proof of Theorem 2.1 that it extends continuously to the missing values $s_i = t_i$. Especially we have

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

---

[2]If we consider the urn model $X_m = \sum_{j=1}^m f(K_{1j}, \dots, K_{dj})$ conditioned by $\sum_{j=1}^m K_{ij} = n_i$ then, for given m and n_i we can choose properly scaled K_{ij} such that $\sum_{j=1}^m \mathbf{E}\, K_{ij} = n_i$. This relation is *hidden* in the (univariate) saddle point equation (2.10). Thus, if one is interested in the asymptotics of $\mathbf{E}\, X_m$ and $\mathbf{Var}\, X_m$ this can be worked out in the same vein as in [17]. However, for the covariance we need the joint distribution $(X_m(\mathbf{n}), X_m(\tilde{\mathbf{n}}))$ and two (differently) scaled versions K_{ij} and $\tilde{K}_{ij}$ with $\sum_{j=1}^m \mathbf{E}\, K_{ij} = n_i$ and $\sum_{j=1}^m \mathbf{E}\, \tilde{K}_{ij} = \tilde{n}_i$ which are encoded in (2.6) and (2.7). This would lead to a probabilistic – however, not really elegant – interperation of our formulae.

Furthermore, $B_{\mathbf{s},\mathbf{s}}$ is a little bit easier to calculate than $B_{\mathbf{s},\mathbf{t}}$.

$$B_{\mathbf{s};\mathbf{s}} = \mu_{\mathbf{s}} + \frac{\begin{vmatrix} \tilde\kappa_{uu} & \tilde\kappa_{uv_1} & \cdots & \tilde\kappa_{uv_d} \\ \tilde\kappa_{v_1u} & \tilde\kappa_{v_1v_1} & \cdots & \tilde\kappa_{v_1v_d} \\ \vdots & \vdots & & \vdots \\ \tilde\kappa_{v_du} & \tilde\kappa_{v_dv_1} & \cdots & \tilde\kappa_{v_dv_d} \end{vmatrix}}{\begin{vmatrix} \tilde\kappa_{v_1v_1} & \cdots & \tilde\kappa_{v_1v_d} \\ \vdots & & \vdots \\ \tilde\kappa_{v_dv_1} & \cdots & \tilde\kappa_{v_dv_d} \end{vmatrix}}, \tag{2.11}$$

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

$$\tilde\kappa_{y,z} := \left.\frac{\partial^2(\log\phi_1(e^u, \rho_1 e^{v_1}, \dots, \rho_d e^{v_d}))}{\partial y \partial z}\right|_{u=v_1=\cdots=v_d=0}$$

and $\boldsymbol{\rho}(\mathbf{s}) = (\rho_1, \dots, \rho_d)$ is defined in (2.6).

Remark 3. We also want to mention that the formula for $B_{\mathbf{s},\mathbf{t}}$ is much easier if

$$\phi_b(\mathbf{1}; \mathbf{z}_1, \dots, \mathbf{z}_b) = \prod_{j=1}^{b}\prod_{i=1}^{d} e^{z_{ij}},$$

which we usually refer as the *standard model*. Here we have $\boldsymbol{\rho}_1(\mathbf{s},\mathbf{t}) = \mathbf{s}$ and $\boldsymbol{\rho}_2(\mathbf{s},\mathbf{t}) = \mathbf{t} - \mathbf{s}$ and

$$\kappa_{u_1w_j} = 0 \quad (1 \le j \le d), \tag{2.12}$$

$$\kappa_{v_iw_j} = \kappa_{w_jv_i} = 0 \quad (1 \le i, j \le d). \tag{2.13}$$

3. Join and Balanced Urn 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 these cases. We consider the case $d = 2$ and models with factorization, i.e., where we have

$$\phi_1(1, y, z) = g_1(y)\ g_2(z), \tag{3.1}$$

and give explicit results for equijoins and semijoins, and for balanced urns which generalize those in [3, 8, 9, 10]. (For the sake of shortness we will calculate the covariance function only explicitly for the standard model of infinte urns.)

3.1. **Equijoin.** For the equijoin models we have two types of balls and the valuation $f(k,l) = kl$. Thus $\phi_1(x, y, z) = \sum_{k,l} a_k b_l x^{kl} y^k z^l$ with $g_1(y) = \sum_k a_k y^k$ and $g_2(z) = \sum_l b_l z^l$ and hence $\frac{\partial\phi_1}{\partial x}(1, y, z) = y g_1'(y)\, z g_2'(z)$. Throughout this section set $s_i := n_i/m$ for $i = 1, 2$, where n_i denotes the number of balls of type i. We give results for the four cases, where the urns are either bounded or unbounded w.r.t. balls of type one and two. Denote these models by UU, UB, BU, BB, where the ith letter indicates whether the urns are bounded (by δ_i) or not w.r.t. balls of type i. Inserting the generating functions into Theorem 2.1 we get the results in Table 1.

model	generating function	$\mathbf{E}X_m(n_1, n_2)$	$\mathbf{Var}X_m(n_1, n_2)$
UU	$g_1(z) = g_2(z) = e^z$	$ms_1s_2 + O(1)$	$ms_1s_2 + O(1)$
UB	$g_1(z) = e^z,\ g_2(z) = (1+z)^{\delta_2}$	$ms_1s_2 + O(1)$	$ms_1s_2(1 - s_2/\delta_2) + O(1)$
BU	$g_1(z) = (1+z)^{\delta_1},\ g_2(z) = e^z$	$ms_1s_2 + O(1)$	$ms_1s_2(1 - s_1/\delta_1) + O(1)$
BB	$g_1(z) = (1+z)^{\delta_1},\ g_2(z) = (1+z)^{\delta_2}$	$ms_1s_2 + O(1)$	$ms_1s_2(1 - s_1/\delta_1)(1 - s_2/\delta_2) + O(1)$

Table 1. Expectation and variance for the equijoin models ($s_i = n_i/m$)

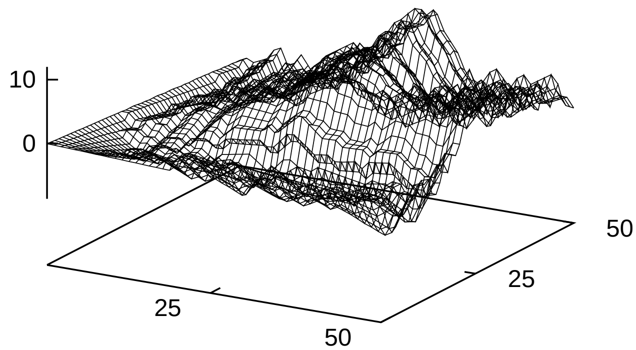


FIGURE 1. Centered process for the equijoin size, infinite urns, $m = 20$ and $n_1, n_2 \leq 50$.

In a similar way we can calculate the covariance function. For example, in the case of infinite urns we have $B_{s_1,s_2;t_1,t_2} = s_1 s_2$ if $s_1 \leq t_1$ and $s_2 \leq t_2$. Hence the limiting process G is precisely a Brownian sheet (cf. [19], see also Figure 1).

3.2. **Semijoin.** We now turn to the semijoin. By $f(k,l) = k\mathbf{1}_{[l>0]}$ we have

$$\phi_1(x,y,z) = g_1(y) + g_1(xy)(g_2(z) - 1)$$

where g_1 and g_2 are chosen as for the equijoin models UU, UB, BU, and BB, respectively. Thus we get the results in Table 2.

model	$\mathbf{E}X_m(n_1,n_2)$	$\mathbf{Var}X_m(n_1,n_2)$
UU	$ms_1\left(1 - \frac{1}{e^{s_2}}\right) + O(1)$	$ms_1 e^{-s_2}\left((1+s_1)(1-e^{-s_2}) - s_1 s_2 e^{-s_1 s_2}\right) + O(1)$
UB	$ms_1\left(1 - \left(1 - \frac{s_2}{\delta_2}\right)^{-\delta_2}\right) + O(1)$	$ms_1\left(1 - \frac{s_2}{\delta_2}\right)^{2\delta_2 - 1}\left[(1+s_1)\left(1 - \frac{s_2}{\delta_2}\right) - s_1 s_2\right] + O(1)$
BU	$ms_1\left(1 - \frac{1}{e^{s_2}}\right) + O(1)$	$ms_1 e^{-s_2}\left[(1 + s_1 - \frac{s_1}{\delta_1})(1-e^{-s_2}) - s_1 s_2 e^{-s_1 s_2}\right] + O(1)$
BB	$ms_1\left(1 - \left(1 - \frac{s_2}{\delta_2}\right)^{-\delta_2}\right) + O(1)$	$ms_1\left(1 - \frac{s_2}{\delta_2}\right)^{2\delta_2 - 1} \times \left[\left(1 + s_1 - \frac{s_1}{\delta_1}\right)\left(1 - \frac{s_2}{\delta_2}\right) - s_1 s_2\right] + O(1)$

TABLE 2. Expectation and variance for the semijoin models ($s_i = n_i/m$)

The generating function for the two-dimensional distributions is $\Phi_2 = \phi_2^m$, where

$$\phi_2 = g_1(y_1 + y_2) + g_1(x_2 y_1 + x_2 y_2)(g_2(z_2) - 1) + g_1(x_1 x_2 y_1 + x_2 y_2)(g_2(z_1 + z_2) - g_2(z_2)).$$

For example, infinite urns on both types of balls give

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

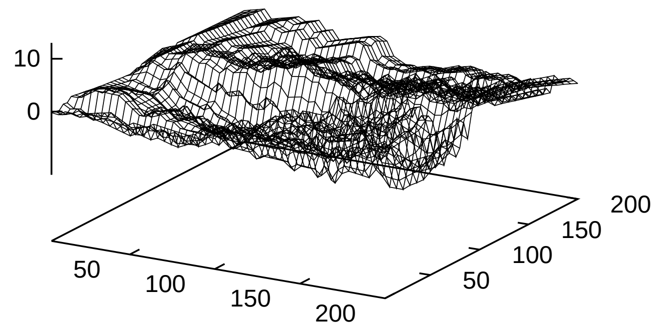


FIGURE 2. Centered process for the semijoin size, unbounded urns, $m = 80$ and $n_1, n_2 \leq 200$.

3.3. **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. 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 $(f \odot g)(t) = \sum_k f_k g_k t^k$. We define a shifted version of the Hadamard product of the functions g_1 and g_2 (defined by the equation (3.1)) as

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

Of course $\lambda_0(t) = g_1 \odot g_2(t)$.

We have here $\phi_1(x, y, z) = g_1(y)g_2(z) + (x-1)y^q \lambda_q(yz)$, which we can also write as

$$\phi_1(x, y, z) = g_1(y)g_2(z) + (x-1)\psi_q(y, z) \qquad \text{with} \qquad \psi_q(y, z) := y^q \lambda_q(yz) = [u^q] g_1(uy) g_2(z/u).$$

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 $g_1(uy)\,(z/u)$. In the same vein, the generating function for allocations in two batches can be written as

$$\begin{aligned} \phi_2(x_1, x_2, y_1, y_2, z_1, z_2) &= (x_1-1)(x_2-1)\,[u^q v^0] g_1(uy_1 + vy_2) g_2\left(\frac{z_1}{u} + \frac{z_2}{v}\right) \\ &\quad + (x_1-1)\,[u^q] g_1(uy_1 + y_2) g_2\left(\frac{z_1}{u} + z_2\right) + (x_2-1)\,\psi_q(y_1+y_2, z_1+z_2) \\ &\quad + g_1(y_1+y_2)\, g_2(z_1+z_2). \end{aligned}$$

The asymptotic expectation is

$$\mathbf{E}X_m(n_1, n_2) = m\mu_{s_1,s_2}(q) + O(1) \qquad \text{with} \qquad \mu_{s_1,s_2}(q) := \frac{\rho_1^q \lambda_q(\rho_1\rho_2)}{g_1(\rho_1) g_2(\rho_2)}$$

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

$$\tilde{B}_{s_1,s_2}(q) = \mu_{s_1,s_2}(q)\left(1 - \mu_{s_1,s_2}(q)\left[1 + \frac{(\tau + q - s_1)^2}{\sigma_1^2} + \frac{(\tau - s_2)^2}{\sigma_2^2}\right]\right) \tag{3.2}$$

with $\tau := \rho_1\rho_2\lambda_q'(\rho_1\rho_2)/\lambda_q(\rho_1\rho_2)$ and $\sigma_1^2 = \rho_1^2 \frac{g_1''}{g_1}(\rho_1) + s_1 - s_1^2$ and $\sigma_2^2 = \rho_2^2 \frac{g_2''}{g_2}(\rho_2) + s_2 - s_2^2$ where ρ_1 and ρ_2 are defined as solutions of $\rho_1 \frac{dg_1}{dy}(\rho_1) = s_1\, g_1(\rho_1)$ and $\rho_2 \frac{dg_2}{dz}(\rho_2) = s_2\, g_2(\rho_2)$. (In the same way we can compute the covariance function.)

For infinite urns the generating functions ϕ_1 and ϕ_2 can be expressed in terms of Bessel functions. We have $\lambda_q(t) = t^{-q/2} I_q(2\sqrt{t})$ and $\lambda_q'(t) = t^{-(q+1)/2} I_{q+1}(2\sqrt{t}) = g_{q+1}(t)$ and obtain

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

Furthermore

$$\psi_q(y,z) = \left(\frac{y}{z}\right)^{q/2} I_q(2\sqrt{(y)\,(z)}).$$

Thus, we also obtain a (simple) representation of the covariance function

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

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

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

4. Models with Deletions

In some instances, e.g. when modeling dynamic databases to study the evolution of projection or join sizes, we need to allow new operations, for example the deletion of items (balls), or the existence of queries that do not modify the current state of the system (no ball is added or deleted). In what follows we explicitly determine the corresponding generating functions with a combinatorial approach. In a similar way general update models (including those with queries) could be studied.

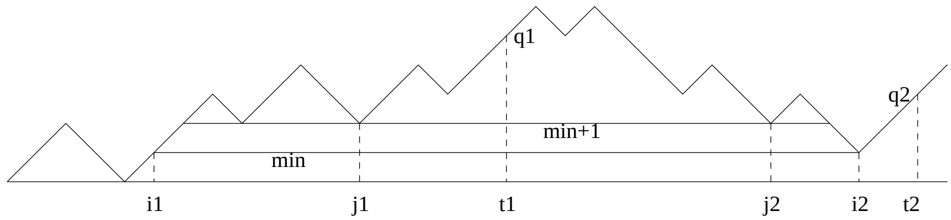
4.1. Allocations and deletions in a single urn.

Our model is based on the following assumptions:

- The urns have infinite capacity and are chosen with uniform probability $1/m$.
- The balls 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.
- 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 model 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 (number of white balls minus number of black balls) is always positive or null. Thus the balance is the actual number of balls in the urn.

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. (If we consider also queries that do not add or delete balls, we would simply take prefixes of Motzkin paths as allowed sequences.)



Put here the figure for Dyck paths....

FIGURE 3. Decomposition of a Dyck prefix

4.2. **Generating functions.** In the 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. 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 current number of balls (balls inserted and not deleted) in the sequence of urns, and the total number of operations, i.e., the time.

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. 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 $P(t, z) = d(t)/(1 - tzd(t))$. The function describing the behaviour of one urn 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 π_{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:

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

At least as long as we are working with unbounded urns, 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 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 $z_1 t_1 d(t_1)$,

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

and the path between j_2 and i_2 is enumerated by $t_2 d(t_2)$. Hence the central part of the path (including the case $q_1 = min$) is enumerated by

$$\frac{1}{1 - z_1\, t_1\, t_2\, d(t_1)\, d(t_2)}.$$

- 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 now clear 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. We now define a modified discrete process $\overline{X}_m$ by

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

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

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

such 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; 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 ϕ_1 and ϕ_2 by putting $z = z_1 = z_2 = 1$ in the corresponding functions computed in Section 4.2:

$$\begin{aligned}\phi_1(x, t) &= (x - 1)\, g(t) \odot d(t) + g(t) \odot \frac{d(t)}{1 - td(t)};\\ \phi_2(x_1, x_2, t_1, t_2) &= (x_1 - 1)(x_2 - 1)\kappa(t_1, t_2, 0, 0) + (x_1 - 1)\kappa(t_1, t_2, 0, 1)\\ &\quad +(x_2 - 1)\kappa(t_1, t_2, 1, 0) + \kappa(t_1, t_2, 1, 1).\end{aligned}$$

Set $f(t) := \lambda(t, 0) = g(t) \odot d(t)$ 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 asymptotic expectation is $\mathbf{E}X_m(n) = m\mu_{n/m} + O(1)$, with $\mu_s = f(\rho)/g_1(\rho)$, where ρ is defined as the unique real positive solution of the equation $tg_1^{'}(t)/g_1(t) = s$, with $s = n/m$. For $g(t) = e^t$ we have $f(t) = I_1(2t)/t$, 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

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

and the equation defining the saddle point becomes $2\rho - \rho f(\rho)/g_1(\rho) = s$. We also have $\mu_s = 2 - s/\rho$. For example, $s = 1$ gives $\rho = 0.6793222511...$ and $\mu_1 = 0.527944582....$ For $s = 2$ we get $\rho = 1.2154678...$ and $\mu_2 = 0.3545302....$

Next we obtain

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

with $\Delta := I_0(2\rho)/I_1(2\rho)$. Numerically we have $\tilde{B}_1 = 0.17394268...$ and $\tilde{B}_2 = 0.1953331....$

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 $Y_m(t_1, \dots, t_d)$. The following two lemmata will be proved together.

Lemma 5.1. *There exists a Gaussian field $G(\mathbf{t})$ with covariance function $B_{\mathbf{s};\mathbf{t}}$, given by (2.8) such that almost all sample paths are continuous.*

Lemma 5.2. *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 X_m is characterized by

$$\mathbf{E}\left(x^{X_m(\mathbf{n})}\right) = \frac{[\mathbf{z}^{\mathbf{n}}]\phi_1(x, \mathbf{z})^m}{[\mathbf{z}^{\mathbf{n}}]\phi_1(1, \mathbf{z})^m}.$$

Thus by standard saddle point techniques (compare with [1] or [5]) it follows that

$$\mathbf{E}\left(x^{X_m(\mathbf{n})}\right) = \left(\frac{\lambda_{\frac{1}{m}\mathbf{n}}(x)}{\lambda_{\frac{1}{m}\mathbf{n}}(1)}\right)^m \left(1 + O\left(\frac{1}{m}\right)\right),$$

where $\lambda_{\mathbf{s}}(x) = \lambda_{s_1,\dots,s_d}(x)$ denotes

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

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

$$z_i\,\frac{\partial\phi_1(x, z_1, \dots, z_d)}{\partial z_i} \;=\; s_i\phi_1(x, z_1, \dots, z_d), \qquad i = 1, \dots, d. \tag{5.1}$$

Consequently, by applying the results of Bender and Richmond [1] one directly obtains that the limiting distribution of X_m is Gaussian (if m and n_i are proportional) with asymptotic mean $\mathbf{E}\,X_m(\mathbf{n}) = m\mu_{\frac{1}{m}\mathbf{n}} + O(1)$ and $\mathbf{Var}\,X_m(\mathbf{n}) = m\sigma^2_{\frac{1}{m}\mathbf{n}} + O(1)$, where

$$\mu_{\mathbf{s}} = \left.\frac{\partial(\log\lambda_{\mathbf{s}}(e^u))}{\partial u}\right|_{u=0} \tag{5.2}$$

and by

$$\sigma^2_{\mathbf{s}} = \left.\frac{\partial^2(\log\lambda_{\mathbf{s}}(e^u))}{\partial^2 u}\right|_{u=0}. \tag{5.3}$$

By an (advanced) exercise in implicit differentiation it follows that $\mu_{\mathbf{s}}$ and $\sigma_{\mathbf{s}}^2 = \tilde{B}_{\mathbf{s}}$ are exactly given by (2.9) and by (2.11).

By another use of saddle point techniques it directly follows that the joint distribution of $(X_m(\mathbf{n}_1), \dots, X_m(\mathbf{n}_1 + \cdots + \mathbf{n}_b))$ is also Gaussian for any fixed $b \ge 2$ (if m and n_{ij} are proportional). This shows that there is a Gaussian field behind. Of course, a Gaussian field is characterized by a covariance function $B_{\mathbf{s};\mathbf{t}}$ which can be determined just by considering the bivariate distribution $(X_m(\mathbf{n}_1), X_m(\mathbf{n}_1 + \mathbf{n}_2))$.

By applying the above procecure it follows that

$$\mathbf{Cov}(X_m(\mathbf{n}_1, X_m(\mathbf{n}_1 + \mathbf{n}_2)) = mB_{\frac{1}{m}\mathbf{n}_1; \frac{1}{m}(\mathbf{n}_1 + \mathbf{n}_2)} + O(1),$$

where

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

with

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

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

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

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

Another exercise in implicit differentiation shows that $B_{\mathbf{s};\mathbf{t}}$ is exactly given by (2.8).

Now let $G(\mathbf{t})$ be the Gaussian field with covariance function $B_{\mathbf{s};\mathbf{t}}$ (compare with [18], 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 $X_m(\mathbf{n})$ converge weakly to the corresponding finite dimensional distributions of $G(\mathbf{t})$ (with $\mathbf{t} = \frac{1}{m}\mathbf{n}$).

In a final step we have to show that $G(\mathbf{t})$ has a modification with a.s. continuous sample paths. For this purpose we will prove that

$$B_{\mathbf{s};\mathbf{t}} = B_{\mathbf{s};\mathbf{s}} + O(\|\mathbf{t}-\mathbf{s}\|). \tag{5.7}$$

Namely, if (5.7) holds then

$$\mathbf{Var}\,(G(\mathbf{t}) - G(\mathbf{s})) = B_{\mathbf{s};\mathbf{s}} - 2B_{\mathbf{s};\mathbf{t}} + B_{\mathbf{t};\mathbf{t}} = O(\|\mathbf{t}-\mathbf{s}\|), \tag{5.8}$$

which implies that $G(\mathbf{t})$ has a modification with a.s. continuous sample paths (see [14, Ch. 2, Theorem 2.8 and Problem 2.9]).

First we observe that by definition (2.7) the saddle point $\boldsymbol{\rho}_2(\mathbf{s},\mathbf{t})$ satisfy

$$\boldsymbol{\rho}_2(\mathbf{s},\mathbf{t}) = O(\|\mathbf{t}-\mathbf{s}\|).$$

We also use the property that ϕ_2 (considered as a power series) can be represented as

$$\phi_2(x_1,x_2,\mathbf{z}_1,\mathbf{z}_2) = \phi_1(x_1x_2,\mathbf{z}_1) + \sum_{j=1}^d z_{j2}R_j(x_1,x_2,\mathbf{z}_1,\mathbf{z}_2),$$

where $R_j \ne 0$ are proper power series in $\mathbf{z}_1,\mathbf{z}_2$ with $R_j(x_1,x_2,\mathbf{z}_1,\mathbf{0}) \ne 0$ (compare with (2.5)). We will use this relation for *small* $\mathbf{z}_2$ and use the shorthand notation

$$\phi_2(x_1,x_2,\mathbf{z}_1,\mathbf{z}_2) = \phi_1(x_1x_2,\mathbf{z}_1) + O(\mathbf{z}_2). \tag{5.9}$$

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

The next step is to show that

$$\kappa_{u_1u_2} = \mu_{\mathbf{s}} + \tilde{\kappa}_{uu} + O(\varepsilon) \tag{5.10}$$

$$\kappa_{u_1v_i} = \tilde{\kappa}_{uv_i} + O(\varepsilon) \quad (1 \le i \le d) \tag{5.11}$$

$$\kappa_{v_iu_2} = \tilde{\kappa}_{v_iu} + O(\varepsilon) \quad (1 \le i \le d) \tag{5.12}$$

$$\kappa_{v_iv_j} = \tilde{\kappa}_{v_iv_j} + O(\varepsilon) \quad (1 \le i, j \le d) \tag{5.13}$$

For the proof of (5.10) 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 directly imply (5.10). In the same way we can treat the other cases (5.11)–(5.13).

Thus, combining (5.10)–(5.13) it follows that the left upper parts of the determinants in (2.8) are given by

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

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

and by

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

Next observe that

$$\kappa_{u_1w_j} = \rho_{j2}\frac{\phi_2\frac{\partial^2\phi_2}{\partial x_1\partial z_{j2}} - \frac{\partial\phi_2}{\partial x_1}\frac{\partial\phi_2}{\partial z_{j2}}}{\phi_2^2},$$

$$\kappa_{v_iw_j} = \rho_{i1}\rho_{j2}\frac{\phi_2\frac{\partial^2\phi_2}{\partial z_{i1}\partial z_{j2}} - \frac{\partial\phi_2}{\partial z_{i1}}\frac{\partial\phi_2}{\partial z_{j2}}}{\phi_2^2},$$

$$\kappa_{w_iw_j} = \rho_{i2}\frac{\frac{\partial\phi_2}{\partial z_{i2}}}{\phi_2}\cdot\delta_{ij} + \rho_{i2}\rho_{j2}\frac{\phi_2\frac{\partial^2\phi_2}{\partial z_{i2}\partial z_{j2}} - \frac{\partial\phi_2}{\partial z_{i2}}\frac{\partial\phi_2}{\partial z_{j2}}}{\phi_2^2},$$

where we have to evaluate at $(x_1, x_2, \mathbf{z}_1, \mathbf{z}_2) = (1, 1, \boldsymbol{\rho}_1(\mathbf{s},\mathbf{t}), \boldsymbol{\rho}_2(\mathbf{s},\mathbf{t}))$. Thus we can extract a common factor of the last d colums and last d rows (of the determinants of (2.8)) of the form $\sqrt{\rho_{j2}}$ and obtain for the right upper part of the determinant (in the formula (2.8))

$$\frac{1}{\sqrt{\rho_{12}\cdots\rho_{d2}}}\begin{vmatrix} \kappa_{u_1w_1} & \cdots & \kappa_{u_1w_d} \\ \kappa_{v_1w_1} & \cdots & \kappa_{v_1w_d} \\ \vdots & & \vdots \\ \kappa_{v_dw_1} & \cdots & \kappa_{v_dw_d} \end{vmatrix} = \begin{vmatrix} O(\sqrt{\varepsilon}) & \cdots & O(\sqrt{\varepsilon}) \\ O(\sqrt{\varepsilon}) & \cdots & O(\sqrt{\varepsilon}) \\ \vdots & & \vdots \\ O(\sqrt{\varepsilon}) & \cdots & O(\sqrt{\varepsilon}) \end{vmatrix}.$$

(A similar relation holds for the left lower part of the determinants.) For the right lower part we get

$$\frac{1}{\rho_{12}\cdots\rho_{d2}}\begin{vmatrix}\kappa_{w_1w_1} & \cdots & \kappa_{w_1w_d}\\ \vdots & & \vdots\\ \kappa_{w_dw_1} & \cdots & \kappa_{w_dw_d}\end{vmatrix} = \begin{vmatrix} c_1+O(\varepsilon) & O(\varepsilon) & \cdots & O(\varepsilon)\\ O(\varepsilon) & c_2+O(\varepsilon) & \cdots & O(\varepsilon)\\ \vdots & & \ddots & \vdots\\ O(\varepsilon) & \cdots & O(\varepsilon) & c_d+O(\varepsilon)\end{vmatrix}$$

for certain non-zero numbers $c_1,\dots,c_d$.

By expanding both *big* determinants in (2.8) and comparing them with (2.11) this immediately proves that $B_{\mathbf{s};\mathbf{t}} = B_{\mathbf{s};\mathbf{s}} + O(\varepsilon)$. This completes the proof of Lemma 5.1 and 5.2. □

5.2. **Tightness.** In order to prove tightness we will need two lemmas, one bounding certain polynomial moments of the centered process and one bounding polynomial moments of the increments of the process. In the proof of the first one we will for brevity assume that $\phi_1(1,\mathbf{z}) = \mathbf{g}(\mathbf{z}) := g_1(z_1)g_2(z_2)\cdots g_d(z_d)$ (g_i is the generating function counting the allocations of type i balls into a single urn). The general case is similar but requires mixed cumulants of the functions which is computationally and notationally much more involved. In order to indicate the general case, the proof of the second lemma is given in full generality, but the long and tedious computations are only sketched.

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

$$\mathbf{E}(X_m(\mathbf{n}) - \mathbf{E}X_m(\mathbf{n}))^{2\Delta} \le C\|\mathbf{n}\|^{\Delta}, \tag{5.14}$$

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

Proof. Set $\mathbf{z} = (z_1,\dots,z_d)$ 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

$$A_i := \mathbf{E}\prod_{j=0}^{i-1}(X_m(\mathbf{n}) - j) = \frac{c_{\boldsymbol{n},i}}{c_{\boldsymbol{n},0}}. \tag{5.15}$$

Then the moment occurring in (5.14) can now be expressed by

$$\mathbf{E}(X_m(\mathbf{n}) - \mathbf{E}X_m(\mathbf{n}))^{2\Delta} = \sum_{l=0}^{2\Delta}\binom{2\Delta}{l}(-1)^l A_1^{2\Delta-l}\sum_{k=1}^{l} S_{lk}A_k, \tag{5.16}$$

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

$$d_j(\mathbf{z}) = \frac{1}{\mathbf{g}(\mathbf{z})}\left.\frac{\partial^j}{\partial x^j}\phi_1(x,\mathbf{z})\right|_{x=1}, \tag{5.17}$$

then by Faà di Bruno's formula (see e.g. Comtet [4]) we have

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

where $(m)_k := m!/(m-k)!$. Thus we have to calculate the coefficient

$$[\mathbf{z}^{\boldsymbol{n}}]\,\mathbf{g}(\mathbf{z})^m\prod_{j=1}^{\alpha}\left(\frac{d_j(\mathbf{z})}{j!}\right)^{k_j}. \tag{5.18}$$

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 $\kappa_{l1} := (zg_l'(z))/g_l(z)$ and $\kappa_{l,j+1}(z) := z\kappa_{l,j}'(z)$, for $j \geq 1, l = 1, \ldots, d$. 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 $l = 1, \ldots, d$ 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

$$\bar{\kappa}_{lj}\left(\frac{n}{m}\right) = \frac{m}{n}\kappa_{lj}\left(\mu_l\left(\frac{n}{m}\right)\right) \tag{5.19}$$

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

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

where $\tilde{\kappa}_{lj}(x) = \bar{\kappa}_{lj}(x)\bar{\kappa}_{l2}(x)^{-j/2}$, and the integration domain $\tilde{B}$ is given by transforming $B = \{\boldsymbol{\theta}|\,|\theta_l| \leq (m\rho_l)^{-1/2+\varepsilon},\ l = 1, \ldots, d\}$ according to the substitutions $\theta_l = u_l/\sqrt{n_l\bar{\kappa}_{l2}(n_l/m)}$ for $l = 1, \ldots, d$. 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, set

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

and evaluate the integral we obtain $[\mathbf{z}^{\boldsymbol{n}}]\mathbf{g}(\mathbf{z})^m \sim V(\boldsymbol{\rho}, \mathbf{n}, m)\left(1 + \sum_{l=1}^d\tilde{\kappa}_{l4}(n_l/m)\right)/8n_l$. Using more terms this procedure yields a multivariate asymptotic series expansion of the form

$$[\mathbf{z}^{\boldsymbol{n}}]\mathbf{g}(\mathbf{z})^m \sim V(\boldsymbol{\rho}, \mathbf{n}, m)\sum_{j_1,\ldots,j_d\geq 0}a_{j_1\cdots j_d}\left(\frac{n_1}{m}, \ldots, \frac{n_d}{m}\right)n_1^{-j_1}\cdots n_d^{-j_d} \tag{5.20}$$

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

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

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

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

$$L\left(z_1 e^{i\theta_1},\dots,z_d e^{i\theta_d}\right) = L(\mathbf{z}) \exp\left(\sum_{j_1+\cdots+j_d>0}^{k} \prod_{l=1}^d \frac{(i\theta_l)^{j_l}}{j_l!} \lambda_{\boldsymbol{j}}(\mathbf{z}) + O\left(\sum_{l=1}^d z_l |\theta_l^{k+1}|\right)\right)$$

with

$$\lambda_{\boldsymbol{e}_\mu}(\mathbf{z}) = z_\mu \frac{\frac{\partial}{\partial z_\mu} L(\mathbf{z})}{L(\mathbf{z})}, \quad \lambda_{\boldsymbol{j}+\boldsymbol{e}_\mu}(\mathbf{z}) = z_\mu \frac{\partial}{\partial z_\mu} \lambda_{\boldsymbol{j}}(\mathbf{z}) \quad \text{for } \mu = 1,\dots,d \tag{5.21}$$

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

$$\tilde{\lambda}_{\boldsymbol{j}}(\mathbf{z}) = \frac{\lambda_{\boldsymbol{j}}(\mu_1(z_1),\dots,\mu_d(z_d)))}{z_1\cdots z_2} \prod_{l=1}^d \bar{\kappa}_{l2}(x_l)^{-j_l/2}$$

and get

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

with $\tilde{\lambda}_{\boldsymbol{j}}(x_1,\dots,x_d) = \bar{\lambda}_{\boldsymbol{j}} \prod_{l=1}^d \bar{\kappa}_{l2}(x_l)^{-j_l/2}$. Expanding the exp-term into a series and evaluating the integral yields finally an asymptotic series expansion of the form

$$c_{\boldsymbol{n},\alpha} \sim \sum_\beta \frac{(m)_\beta}{m^\beta} L(\boldsymbol{\rho}) V(\boldsymbol{\rho},\mathbf{n},m) \sum_{j_1,\dots,j_d\ge 0,} a_{\boldsymbol{j}}^{(\alpha)}\left(\frac{1}{m}\cdot\mathbf{n}\right) \frac{\mathbf{n}^{\boldsymbol{\gamma}}}{\mathbf{n}^{\boldsymbol{j}}} \tag{5.22}$$

with $\boldsymbol{\gamma} = (\gamma_1,\dots,\gamma_d)$ and explicitly computable analytic functions $a_{j_1\cdots j_d}^{(\alpha)}(t_1,\dots,t_d)$.

Inserting this into (5.16) 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 $n_1 = s_1 m,\dots,n_d = s_d m$, then by 5.8 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 (5.22) (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 $n_1^{\gamma_1}\cdots n_d^{\gamma_d}/n_1^{j_1}\cdots n_d^{j_d}$ we must have $\gamma_1+\cdots+\gamma_d - j_1 - \cdots - j_d \le \Delta$. But as to the fact that $n_1^{\gamma_1'}\cdots n_d^{\gamma_d'} \le \|\mathbf{n}\|^\beta$, if $\gamma_1'+\cdots+\gamma_d' \le \beta$, the above inequality guarantees the validity of (5.14) for all $\mathbf{n}$ satisfying $\mathbf{n} = O(m)$ and the proof is complete. □

In order to prove tightness, by [18, Ch. XIII, Ex. 1.12] it suffices to show the following Lemma.

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

$$\mathbf{E}\frac{(X_m(\mathbf{n}+\mathbf{h}) - X_m(\mathbf{n}) - \mathbf{E}(X_m(\mathbf{n}+\mathbf{h}) - X_m(\mathbf{n})))^{2d+2}}{m^{d+1}} \le C\left(\frac{\|\mathbf{h}\|}{m}\right)^{d+1} \tag{5.23}$$

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

$$\mathbf{E}Z_m(\mathbf{n}, \mathbf{h})^{2d+2} \le \sum_{k=0}^{d+1} \binom{2d+2}{2k} \mathbf{E}X_m^c(\mathbf{n}+\mathbf{h})^{2k}\mathbf{E}X_m^c(\mathbf{n})^{2d+2-2k} \tag{5.24}$$

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., $\Phi_2(1/x, x, z_{11}, \ldots, z_{d1}, z_{12}, \ldots, z_{d2})$. Set $\mathbf{z}_1 = (z_{11}, \ldots, z_{d1})$, $\mathbf{z}_2 = (z_{12}, \ldots, z_{d2})$, and

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

Proceeding as in (5.15)–(5.17), the moment occurring in (5.23) can again be expressed by

$$\mathbf{E}(Z_m(\mathbf{n}, \mathbf{h}) - \mathbf{E}Z_m(\mathbf{n}, \mathbf{h}))^{2d+2} = \sum_{l=0}^{2d+2} \binom{2d+2}{l} (-1)^l A_1^{2d+2-l} \sum_{k=1}^{l} S_{lk}A_k \tag{5.25}$$

where $A_i = c_{\boldsymbol{n},\boldsymbol{h},i}/c_{\boldsymbol{n},\boldsymbol{h},0}$. If we set

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

and apply as in the proof of Lemma 5.3 Faà di Bruno's formula, we are left with the task of calculating the coefficient

$$\left[\mathbf{z}_1^{\boldsymbol{n}}\mathbf{z}_2^{\boldsymbol{h}}\right] \phi_2(1, 1, \mathbf{z}_1, \mathbf{z}_2)^m \prod_{j=1}^{\alpha} \left(\frac{d_j(\mathbf{z}_1, \mathbf{z}_2)}{j!}\right)^{k_j}. \tag{5.26}$$

The calculation of $c_{\boldsymbol{n},\boldsymbol{h},0} = [\mathbf{z}_1^{\boldsymbol{n}}\mathbf{z}_2^{\boldsymbol{h}}]\mathbf{g}(\mathbf{z}_1 + \mathbf{z}_2)^m$ can be done in the same manner as the derivation of (5.20). Let $\boldsymbol{v} \in \mathbb{N}_0^{2d}$ and let $\kappa_{\boldsymbol{v}}(\mathbf{z}_1, \mathbf{z}_2)$ and $\bar{\kappa}_{\boldsymbol{v}}$ denote the cumulants of $\phi_2(1, 1, \mathbf{z}_1, \mathbf{z}_2)$ defined analogously to (5.21) and (5.19), respectively. Then the saddle points $\boldsymbol{\rho_1} = (\rho_{11}, \ldots, \rho_{d1})$ and $\boldsymbol{\rho_2} = (\rho_{12}, \ldots, \rho_{d2})$ of $\phi_2(1, 1, \mathbf{z}_1, \mathbf{z}_2)^m\mathbf{z}_1^{-\boldsymbol{n}}\mathbf{z}_2^{-\boldsymbol{h}}$ are given by $(\boldsymbol{\rho_1}, \boldsymbol{\rho_2}) = \boldsymbol{\mu}\left(\frac{1}{m}\cdot\mathbf{n}, \frac{1}{m}\cdot\mathbf{h}\right)$ where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{2d})$ is the inverse of $(\kappa_{\mathbf{e}_1}, \ldots, \kappa_{\mathbf{e}_{2d}})$. Define $\bar{\kappa}_{\boldsymbol{v}}$

$$V(\boldsymbol{\rho_1}, \boldsymbol{\rho_2}, \mathbf{n}, \mathbf{h}, m) = \frac{\phi_2(1, 1, \boldsymbol{\rho_1}, \boldsymbol{\rho_2})^m}{(2\pi)^d\boldsymbol{\rho_1}^{\boldsymbol{n}}\boldsymbol{\rho_2}^{\boldsymbol{h}}\sqrt{\prod_{l=1}^d n_l h_l \bar{\kappa}_{\mathbf{e}_l}\left(\frac{1}{m}\cdot\mathbf{n}, \frac{1}{m}\cdot\mathbf{h}\right)\bar{\kappa}_{\mathbf{e}_{d+l}}\left(\frac{1}{m}\cdot\mathbf{n}, \frac{1}{m}\cdot\mathbf{h}\right)}},$$

Then we get as in the proof of the previous lemma

$$[\mathbf{z}_1^{\boldsymbol{n}}\mathbf{z}_2^{\boldsymbol{h}}]\phi_2(1, 1, \boldsymbol{\rho_1}, \boldsymbol{\rho_2})^m V(\boldsymbol{\rho_1}, \boldsymbol{\rho_2}, \mathbf{n}, \mathbf{h}, m) \sum_{\substack{j_1, \ldots, j_d \ge 0 \\ \delta_1, \ldots, \delta_d \ge 0}} a_{j_1\cdots j_d\delta_1\cdots\delta_d}\left(\frac{1}{m}\cdot\mathbf{n}, \frac{1}{m}\cdot\mathbf{h}\right)\mathbf{n}^{-\boldsymbol{j}}\mathbf{h}^{-\boldsymbol{\delta}}$$

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 (5.26). By (2.4) 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

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

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 $(m)_\beta[\mathbf{z}_1^{\boldsymbol{n}}\mathbf{z}_2^{\boldsymbol{h}}]\phi_2(1,1,\boldsymbol{\rho_1},\boldsymbol{\rho_2})^m K_\beta(\mathbf{z}_1,\mathbf{z}_2)$ 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

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

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 $\gamma_1,\dots,\gamma_d$ with $\sum_j \gamma_j = \beta$. Then we have for $\mathbf{z}_1,\mathbf{z}_2 \in \mathbb{R}^d$

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

where $\lambda_{\boldsymbol{j_1 j_2}}$ are the cumulants of $L(\mathbf{z}_1,\mathbf{z}_2)$ defined analogously to (5.21). Thus we can proceed as in the case $\alpha = 0$. Set

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

as well as $\boldsymbol{j}! = (j_1,\dots,j_d)! := \prod_{i=1}^{d} j_i!$ and $M := \{0,1,2,\dots,k\}$. Then we get (cf. proof of the previous lemma)

$$\begin{aligned} (m)_\beta[\mathbf{z}_1^{\boldsymbol{n}}\mathbf{z}_2^{\boldsymbol{h}}]\mathbf{g}(\mathbf{z}_1+\mathbf{z}_2)^m K_\beta(\mathbf{z}_1,\mathbf{z}_2) &= \frac{(m)_\beta L(\boldsymbol{\rho_1},\boldsymbol{\rho_2})V(\boldsymbol{\rho_1},\boldsymbol{\rho_2},\mathbf{n},\mathbf{h},m)}{(2\pi)^d}\prod_{l=1}^{d}\rho_{l2}^{\gamma_l}\\ \times \int_{\tilde{B}}\cdots\int \exp\Bigg(&-\sum_{l=1}^{d}\frac{u_{l1}^2}{2} - \sum_{l=1}^{d}\frac{u_{l2}^2}{2}\\ + \sum_{\substack{\boldsymbol{j_1},\boldsymbol{j_2}\in M^d\\ \|\boldsymbol{j_1}\|_1,\|\boldsymbol{j_2}\|_1\geq 2}} \frac{(i\boldsymbol{u_1})^{\boldsymbol{j_1}}}{\boldsymbol{j_1}!}\frac{(i\boldsymbol{u_2})^{\boldsymbol{j_2}}}{\boldsymbol{j_2}!}\mathbf{n}^{\mathbf{1}-\boldsymbol{j_1}/2}\mathbf{h}^{\mathbf{1}-\boldsymbol{j_2}/2}\tilde{\kappa}_{\boldsymbol{j_1},\boldsymbol{j_2}}&\left(\frac{1}{m}\cdot\mathbf{n},\frac{1}{m}\cdot\mathbf{h}\right) + \sum_{l=1}^{d}\gamma_l\frac{u_{l2}}{\sqrt{h_l\bar{\kappa}_{l2}(h_l/m)}}\\ + \sum_{j_{11}+\cdots+j_{d1}+j_{12}+\cdots+j_{d2}>0}^{k}\prod_{l=1}^{d}\frac{(iu_{l1})^{j_{l1}}(iu_{l2})^{j_{l2}}n_l^{1-j_{l1}/2}h_l^{1-j_{l2}/2}}{j_{l1}!j_{l2}!m^{2d}}&\tilde{\lambda}_{\boldsymbol{j_1 j_2}}\left(\frac{1}{m}\cdot\mathbf{n},\frac{1}{m}\cdot\mathbf{h}\right)\\ +O\left(m\sum_{l=1}^{d}\rho_{l1}\left|\frac{u_{l1}}{\sqrt{n_l}}\right|^{k+1}\right) + O\left(m\sum_{l=1}^{d}\rho_{l2}\left|\frac{u_{l2}}{\sqrt{h_l}}\right|^{k+1}\right)\Bigg)& du_{11}\cdots du_{d1}\,du_{12}\cdots du_{d2} \end{aligned}$$

with $\tilde{\lambda}_{\boldsymbol{j_1 j_2}}(x_1,\dots,x_{2d}) = \bar{\lambda}_{\boldsymbol{j_1 j_2}}\prod_{l=1}^{d}\bar{\kappa}_{l22}(x_l,x_{d+l})^{(-j_{l1}-j_{l2})/2}$. Expanding the exp-term into a series and evaluating the integral yields finally an asymptotic series expansion of the form

$$c_{\boldsymbol{n},\boldsymbol{h},\alpha} \sim \sum_{\beta} \frac{(m)_\beta}{m^\beta} L(\boldsymbol{\rho_1},\boldsymbol{\rho_2}) V(\boldsymbol{\rho_1},\boldsymbol{\rho_2},\mathbf{n},\mathbf{h},m) \prod_{l=1}^{d} h_l^{\gamma_l} \sum_{\substack{j_1,\ldots,j_d\geq 0,\\ \delta_1,\ldots,\delta_d\geq 0}} a_{\boldsymbol{j},\boldsymbol{\delta}}\left(\frac{1}{m}\cdot\mathbf{n},\frac{1}{m}\mathbf{h}\right)\mathbf{n}^{-\boldsymbol{j}}\mathbf{h}^{-\boldsymbol{\delta}} \tag{5.27}$$

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

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 (5.8) we have $\mathbf{E} Z_m(\mathbf{n},\mathbf{h})^{2d+2} = O\left(\|\mathbf{h}\|^{d+1}\right)$. On the other hand, inserting (5.27) into (5.25), 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

$$\frac{h_1^{\gamma_1}\cdots h_d^{\gamma_d}}{n_1^{j_1}\cdots n_d^{j_d} h_1^{\delta_1}\cdots h_1^{\delta_d}} \tag{5.28}$$

if we neglect constant factors. Thus

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

and (5.28) 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

$$h_1^{\gamma_1'}\cdots h_d^{\gamma_d'}\frac{h_1^{j_1}\cdots h_d^{j_d}}{n_1^{j_1}\cdots n_d^{j_d}}\gg \|h\|^{d+1} \tag{5.29}$$

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 (5.29) contradicts the conclusion of (5.24) and Lemma 5.3. □

Acknowledgement. Thanks are due to P. Chassaing, A. Mokkadem and A. Rouault for information on multivariate processes, to J.F. Marckert for plotting the bi-dimensional Gaussian processes, and to an anonymous referee for valuable comments and suggestions.

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] M. P. QUINE AND J. ROBINSON, *Normal approximations to sums of scores based on occupancy numbers*, Ann. Probab. 12, 794–804, 1984.

[18] D. REVUZ AND M. YOR, *Continuous Martingales and Brownian Motion,* Springer, 1991.

[19] A.W. VAN DER VAART AND J.A. WELLNER, *Weak Convergence and Empirical Processes,* Springer.