Math. Proc. Camb. Phil. Soc. (1993), **114**, 347
Printed in Great Britain

On random polynomials over finite fields

BY RICHARD ARRATIA
Department of Mathematics, University of Southern California, Los Angeles, CA 90089-1113, *U.S.A.*

A. D. BARBOUR
Institut für Angewandte Mathematik, Universität Zürich, Rämistrasse 74, *CH*-8001 *Zürich, Switzerland*

AND SIMON TAVARÉ
Department of Mathematics, University of Southern California, Los Angeles, CA 90089-1113, *U.S.A.*

(*Received* 20 *July* 1992)

Abstract

We consider random monic polynomials of degree n over a finite field of q elements, chosen with all q^n possibilities equally likely, factored into monic irreducible factors. More generally, relaxing the restriction that q be a prime power, we consider that multiset construction in which the total number of possibilities of weight n is q^n. We establish various approximations for the joint distribution of factors, by giving upper bounds on the total variation distance to simpler discrete distributions. For example, the counts for particular factors are approximately independent and geometrically distributed, and the counts for all factors of sizes $1, 2, \ldots, b$, where $b = O(n/\log n)$, are approximated by independent negative binomial random variables. As another example, the joint distribution of the large factors is close to the joint distribution of the large cycles in a random permutation. We show how these discrete approximations imply a Brownian motion functional central limit theorem and a Poisson–Dirichlet limit theorem, together with appropriate error estimates. We also give Poisson approximations, with error bounds, for the distribution of the total number of factors.

1*a*. *Introduction*

For integers $q \geqslant 2$ and $n \geqslant 1$ we consider random variables $Y_j \equiv Y_j(n)$, for $j = 1, 2, \ldots, n$, whose joint distribution is given by

$$\mathbb{P}[Y_1 = y_1, \ldots, Y_n = y_n] = q^{-n} \prod_{i=1}^{n} \binom{N_q(i) + y_i - 1}{y_i} \tag{1·1}$$

where $y_1, \ldots, y_n \geqslant 0$ satisfy $\sum_{i=1}^{n} i y_i = n$ and

$$N_q(i) = \frac{1}{i} \sum_{j|i} \mu(i/j)\, q^j, \quad i \geqslant 1. \tag{1·2}$$

In (1·2), μ is the Möbius function, with $\mu(d) = (-1)^k$ if d is the product of k distinct primes, and $\mu(d) = 0$ if d is divisible by a square. By Möbius inversion, (1·2) is equivalent to

$$q^n = \sum_{d|n} dN_q(d), \quad n \geqslant 1. \tag{1·3}$$

From (1·3) one sees immediately that $q^n \geqslant nN_q(n)$, and furthermore these two expressions are asymptotic, with relative error decaying exponentially fast as $n \to \infty$, since their difference is

$$\sum_{d|n,\, d<n} dN_q(d) \leqslant \sum_{d|n,\, d<n} q^d \leqslant \sum_{d \leqslant n/2} q^d < (q/(q-1))\, q^{n/2} \leqslant 2q^{n/2}.$$

We will make repeated use of this simple bound, for our purposes the crucial property of the $N_q(i)$.

When q is a prime power, and F_q is the finite field with q elements, $N_q(i)$ is the number of monic irreducible polynomials of order i over F_q; see Lidl and Niederreiter [**20**, p. 82ff.]. With π a random monic polynomial of degree n over F_q, chosen uniformly at random from the q^n possibilities, and Y_j the number of irreducible factors of degree j in π, we have the joint distribution given by (1·1).

The decomposition of random polynomials into irreducible factors is an example of a multiset construction. For $i \geqslant 1$ there are $N_q(i)$ different types of objects of weight i, with an unlimited supply of each type of object. Among all possible multisets of total weight n, we select one at random, and let Y_j be the number of objects of weight j included; the joint distribution of these counts is given by (1·1). See Flajolet and Soria [**14**] and Arratia and Tavaré [**1**] for probabilistic treatments of multisets in general. With $N_q(i)$ given by (1·2), where q is any positive integer, the total number of possible multisets of weight n is q^n, and (1·3) is valid. This multiset construction for general $q \geqslant 2$, $n \geqslant 1$ can be interpreted in terms of necklaces (Metropolis and Rota [**21**, **22**]).

The purpose of this paper is to investigate simplifying approximations, with error bounds, for the joint distribution in (1·1). Our starting point was the result from Car [**5**], that for large n and $k \ll \log n$, the number of polynomials with exactly k factors is very close to $\{n^{-1} q^n (\log n)^{k-1}/(k-1)!\}$. Effectively, the Poisson distribution with mean $\log n$ serves to approximate the distribution of the total number of factors, minus 1. Our approximations to (1·1) are also expressed in terms of comparison to simpler random objects, such as independent negative binomial random variables, Poisson processes, and random permutations. In particular, the joint distribution of large factor sizes of a random polynomial is similar to the joint distribution of large cycle sizes in a uniform random permutation, about which much is known, see for example Kolchin [**17**].

Using a bijection discovered by Gessel and Reutenauer [**15**] between $\{1, \ldots, q\}^n$ and multisets of necklaces, Diaconis, McGrath, and Pitman [**10**] found the distribution (1·1) for the cycle structure of non-uniformly distributed random permutations of n elements derived from random riffle shuffles. In their setting, the parameter q represents the number of decks into which the original deck is cut before the riffle shuffle, and can be any positive integer, not necessarily a prime power. They give a variety of exact formulae for the distribution of the counts, analogous to classical formulae for the uniform case. Using the method of moments, they obtain the

asymptotic independence and negative binomial distribution of the small counts, and the same Poisson–Dirichlet limit for the big counts as for the cycles of a uniformly distributed random permutation. Hansen[**16**] establishes a Poisson–Dirichlet limit distribution for a general class of combinatorial structures whose generating functions have a logarithmic singularity, and observes that polynomials over finite fields satisfy this condition. Here, we systematically establish approximations for the joint distribution of factors, by giving upper bounds on the total variation distance to simpler discrete distributions. For example, the counts for individual factors are approximately independent and geometrically distributed, and the counts for all factors of sizes $1, 2, \ldots, b$, where $b = O(n/\log n)$, are approximated by independent negative binomial random variables. As another example, the joint distribution of the large factors is close to the joint distribution of the large cycles in a random permutation. We show how these discrete approximations imply, as easy corollaries, a Brownian motion functional central limit theorem and a Poisson–Dirichlet limit theorem, together with appropriate error estimates. We conclude with Poisson approximations, with error bounds, for the distribution of the total number of factors.

1*b*. *Notation*

We use the language of random polynomials throughout this paper. For the general case, q can be any integer at least two, not necessarily a prime power. The multisets considered here, having q^n objects of weight n, may still be described in terms of polynomials decomposed into products of monic irreducible factors. To do this, use the field of rationals. Pick $N_q(i)$ irreducible polynomials of degree i, such as $(x^i + p_k)$, $k = 1, 2, \ldots, N_q(i)$, where p_k is the kth prime. Consider all products of these, and select at random any of the q^n such products of degree n, with all possibilities equally likely.

Y_i is the number of factors of degree i, so $\sum_1^n iY_i \equiv n$.

$\tilde{Y}_i$ are independent negative binomial $(N_q(i), q^{-i})$ random variables, which give the distributional limit of the Y_i, as $n \to \infty$.

$\hat{Y}_i$ are independent Poisson random variables, with the same means as the $\tilde{Y}_i$.

X_j is the number of occurrences of the irreducible factor ϕ_j, under an arbitrary fixed enumeration of the possible factors.

$\delta(j)$ is the degree of ϕ_j, so that $\sum \delta(j) X_j \equiv n$ and $Y_i \equiv \sum X_j 1(\delta(j) = i)$.

$\tilde{X}_j$ are independent geometric $(q^{-\delta(j)})$ random variables, which give the distributional limit of the X_j as $n \to \infty$.

$K_0' \equiv \sum Y_i \equiv \sum X_j$ is the total number of irreducible factors, for a randomly chosen polynomial of degree n.

M_k is the label of the kth factor selected in size-biassed sampling, so $\delta(M_k)$ is the size of that factor. We take $\delta(0) \equiv 0$ and set $M_k = 0$ in case $k > K_0'$. Check that $X_j = \sum_{k \geqslant 1} 1(M_k = j)$.

L_k is the size of the kth largest factor degree in a randomly chosen polynomial of degree n, with value 0 if there are fewer than k factors.

Δ_k is the size of the kth oldest cycle in a random permutation of n objects, with value 0 if $k > K_0$, the total number of cycles. Check that $n = \sum_{k \geqslant 1} \Delta_k$.

C_i is the number of cycles of size i, so $\sum iC_i \equiv n$ and $C_i = \sum_k 1(\Delta_k = i)$.

Λ_k is the size of the kth largest cycle in a random permutation of n objects, with value 0 if there are fewer than k cycles.

2. *Size-biassed sampling*

The cycle structure of random permutations is most easily analysed in terms of a size-biassed sampling scheme, which not only generates the cycle structure but also gives an ordering among the cycles. We describe here two constructions that generate an ordered list of lengths of cycles. The first construction is motivated by a 'record value process' (Rényi [24]). Let $(I_j)_{j\geqslant 1}$ be independent Bernoulli (j^{-1}) random variables, set $T_1 = 1 = \min\{j > 0 : I_j = 1\}$, and define

$$T_i = \min\{j > T_{i-1} : I_j = 1\} \wedge (n+1), \quad i \geqslant 2.$$

Let $K_0 = \max\{i : T_i < n+1\}$. The T_i can be used to generate a uniformly distributed permutation of $\{1, 2, \ldots, n\}$, where K_0 is the number of cycles, $\Delta_1 \equiv n+1-T_{K_0}$ is the length of the cycle containing 1, $\Delta_2 \equiv T_{K_0} - T_{K_0-1}$ is the length of the cycle containing the smallest element not in the first cycle, and so on. The length Δ_i of the ith cycle is $T_{K_0-i+2} - T_{K_0-i+1}$ if $i \leqslant K_0$, with $\Delta_i = 0$ if $i > K_0$.

A second description of the same size biassed sampling scheme has random variables $\tilde{T}_i$ and $\tilde{K}_0$ such that $(T_1, T_2, \ldots, T_{K_0+1})$ has the same distribution as $(\tilde{T}_{\tilde{K}_0+1}, \tilde{T}_{\tilde{K}_0}, \ldots, \tilde{T}_2, \tilde{T}_1)$. Set $\tilde{T}_1 = n+1$, and, given $\tilde{T}_1, \ldots, \tilde{T}_{i-1}$, if $\tilde{T}_{i-1} > 1$ choose $\tilde{T}_i$ uniformly at random from the integers $\{1, 2, \ldots, \tilde{T}_{i-1}-1\}$; otherwise if $\tilde{T}_{i-1} = 1$, define $\tilde{K}_0 = i-2$.

From this second coupling one sees that for all $d_i \geqslant 1$, $1 \leqslant i \leqslant k$ such that $\sum_{i=1}^k d_i \leqslant n$,

$$\mathbb{P}[\Delta_1 = d_1, \ldots, \Delta_k = d_k] = \prod_{i=1}^{k} \left(n - \sum_{j=1}^{i-1} d_j\right)^{-1}. \tag{2·1}$$

Aspects of these two constructions of the cycle structure of a random permutation have been exploited in several places, among them Feller [12], Vershik and Shmidt [26], Diaconis and Pitman [9], Donnelly and Joyce [11], Barbour [3], and Arratia, Barbour and Tavaré [2].

In order to obtain a parallel construction for the factorization of a random polynomial π of degree n, let ϕ_0 denote the unit polynomial, and let the allowable irreducible monic polynomials be listed in some order as $\phi_1, \phi_2, \ldots$. Let X_j denote the number of times ϕ_j appears as a factor in π, and let $Y_d = \sum_{j:\delta(j)=d} X_j$, where $\delta(j)$ is the degree of ϕ_j, so that Y_d denotes the number of factors of degree d in π. Now consider the random sequence of integers $(M_k)_{k\geqslant 1}$ constructed as follows. Choose π uniformly at random, and then select its irreducible factors one at a time, by sampling at random from those not already selected, with probabilities proportional to their degree. If factor ϕ_m is selected at step k, set $M_k = m$. If π is exhausted after k steps, set $M_j = 0$, $j > k$, and set $K_0' = k$. Then it is easily seen that

$$\mathbb{P}[M_1 = m] = n^{-1}\delta(m)\,\mathbb{E}X_m. \tag{2·2}$$

The general joint probability is determined by the formula

$$\mathbb{P}[M_1 = m_1, \ldots, M_k = m_k] = \left\{\prod_{i=1}^{k} \frac{\delta(m_i)}{n - \sum_{j=1}^{i-1}\delta(m_j)}\right\} \mathbb{E}\left\{\prod_{j=1}^{l} (X_{m_j'})_{s_j}\right\}, \tag{2·3}$$

whenever $m_i \geqslant 1$, $1 \leqslant i \leqslant k$, and $\sum_{i=1}^k \delta(m_i) \leqslant n$; $m_1', \ldots, m_l'$ represent the distinct

values taken by the sequence $m_1, \ldots, m_k$, and $s_1, s_2, \ldots, s_l$ their multiplicities; and $(x)_r = x(x-1)\ldots(x-r+1)$. The joint distribution of the degrees of the factors sampled in this way is then obtained by adding (2·3) over the relevant choices of $m_1, \ldots, m_k$:

$$\mathbb{P}[\delta(M_1 = d_1, \ldots, \delta(M_k) = d_k] = \left\{\prod_{i=1}^{k} \frac{d_i}{n - \sum_{j=1}^{i-1} d_j}\right\} \sum_{m_1 : \delta(m_1) = d_1} \cdots \sum_{m_k : \delta(m_k) = d_k} \mathbb{E}\left\{\prod_{j=1}^{l} (X_{m_j'})_{s_j}\right\}. \tag{2·4}$$

In order to exhibit a parallel between (2·4) and (2·1), it is necessary to note the following facts about our random polynomials. First, of the q^n allowable polynomials of degree n, q^{n-r} have a given allowable irreducible polynomial ρ of degree $r \leqslant n$ as a factor, which, expressed in terms of probabilities, says that

$$\mathbb{P}[\rho \text{ divides } \pi] = q^{-r}, \quad r \leqslant n. \tag{2·5}$$

Secondly, as proved after (1·3), the number $N_q(d)$ of irreducible monic polynomials of degree d satisfies

$$q^{-1} N_q(1) = 1; \quad 0 \leqslant 1 - q^{-d}\, d N_q(d) \leqslant 2q^{-d/2}, \quad d \geqslant 2. \tag{2·6}$$

Thus, if obtaining the same factor twice were unusual, as is the case if q is big, the right hand side of (2·4) might be expected to be close to

$$\left\{\prod_{i=1}^{k} N_q(d_i)\right\} q^{-\sum_{i=1}^{k} d_i} \left\{\prod_{i=1}^{k} \frac{d_i}{n - \sum_{j=1}^{i-1} d_j}\right\} \tag{2·7}$$

because of (2·5), and (2·6) then suggests that this is in turn almost

$$\prod_{i=1}^{k} \left(n - \sum_{j=1}^{i-1} d_j\right)^{-1},$$

the right hand side of (2·1). Thus a parallel with the cycle structure of a random permutation seems reasonable, insofar as the above argument can be made precise. It turns out that, even for q small, much can be gained by this approach: see Section 5. However, equation (2·5) suggests an even more direct line of investigation.

3. *Factors of small degree*

Let $J_k = \{j : \delta(j) \leqslant k\}$, and write $X(J_k)$ for the vector $(X_j, j \in J_k)$ with components ordered by increasing j. Let $c = (c_j, j \in J_k)$ be a similar vector of non-negative integers. Then (2·5) implies that

$$\mathbb{P}[X(J_k) \geqslant c] = q^{-\sum_{j \in J_k} c_j \delta(j)} = \prod_{j \in J_k} q^{-c_j \delta(j)}, \tag{3·1}$$

whenever $\sum_{j \in J_k} c_j\, \delta(j) \leqslant n$, and $\mathbb{P}[X(J_k) \geqslant c] = 0$ otherwise. This suggests that, if $k \leqslant n$, the distribution of $X(J_k)$ should be close to that of a vector of independent geometric random variables $\tilde{X}_j \sim \mathrm{Ge}(q^{-\delta(j)})$, where $\mathrm{Ge}(\theta)\{r\} = (1-\theta)\,\theta^r, r \geqslant 0$. This is the substance of the following theorem.

THEOREM 3·1. *For all* $k \geqslant 1$,

$$d_{TV}(\mathscr{L}(X(J_k)), \mathscr{L}(\tilde{X}(J_k))) = O(k\, e^{-(n/2k)\log(4/3)}).$$

If $k \geqslant \log n/\log(\frac{4}{5}q^{\frac{1}{2}})$, *we have the better estimate*

$$d_{TV}(\mathscr{L}(X(J_k)), \mathscr{L}(\tilde{X}(J_k))) = O\left(k^2 \exp\left\{-\frac{n}{2k}\log\frac{n}{2k}+\frac{n}{2k}\right\}\right).$$

Remark 3·2. This shows, for instance, that the distributions of $X(J_k)$ and $\tilde{X}(J_k)$ are asymptotically close as $n \to \infty$, if $k = k(n) = O(n/\log n)$, and that the accuracy of approximation is very high if $k = n^\beta$ for some $\beta < 1$.

Proof. To start with, observe that if $Z = (Z_j, j \in J_k)$ is any non-negative random vector,

$$\mathbb{P}[Z = c] = \mathbb{P}[Z \geqslant c] + \sum_{r=1}^{|J_k|} (-1)^r \sum_{B \in \mathscr{D}_r} \mathbb{P}[Z \geqslant c + \sum_{s \in B} e_s], \tag{3·2}$$

where $\mathscr{D}_r$ is the set of all r-subsets of J_k and e_s denotes the sth coordinate vector. Using (3·2) on $X(J_k)$ and $\tilde{X}(J_k)$ for c satisfying $\sum_{j \in J_k} c_j\,\delta(j) = l \leqslant n$ yields

$$|\mathbb{P}[X(J_k) = c] - \mathbb{P}[\tilde{X}(J_k) = c]| \leqslant \sum_{r=1}^{|J_k|} \sum_{B \in \mathscr{D}_r} 1_{\{l+\sum_{s \in B}\delta(s) > n\}}\, q^{-(l+\sum_{s \in B}\delta(s))}. \tag{3·3}$$

On the other hand, using (2·6),

$$\mathbb{P}[\tilde{X}(J_k) = c] = \prod_{j \in J_k} (1-q^{-\delta(j)})\, q^{-c_j\delta(j)} \asymp q^{-l} \exp\left\{-\sum_{i=1}^{k} q^{-i} N_q(i)\right\} \asymp k^{-1} q^{-l}. \tag{3·4}$$

Hence the relative error in approximating $\mathbb{P}[X(J_k) = c]$ by $\mathbb{P}[\tilde{X}(J_k) = c]$ is of order at most

$$O\left(k \sum_{r=1}^{|J_k|} \sum_{B \in \mathscr{D}_r} 1_{\{\sum_{s \in B}\delta(s) > n-l\}}\, q^{-\sum_{s \in B}\delta(s)}\right), \tag{3·5}$$

when $l = \sum_{j \in J_k} c_j\,\delta(j) \leqslant n$.

The contribution to the sum in (3·5) from those $B \subset J_k$ such that $\sum_{s \in B}\delta(s) = t$ is just

$$q^{-t}\{\#\text{ of polynomials of degree } t \text{ in } H_k\} = \mathbb{P}[\pi^{(t)} \in H_k],$$

where H_k is the set of monic polynomials with distinct irreducible factors all of degree no greater than k, and $\pi^{(t)}$ is a uniform random polynomial of degree t. Applying (2·3) with t for n gives

$$\mathbb{P}[\pi^{(t)} \in H_k] \leqslant \sum_{i \geqslant 1} \sum_{\substack{t_1+\ldots+t_i=t \\ 1 \leqslant t_j \leqslant k,\,\text{all } j}} \prod_{u=1}^{i} \frac{t_u N_q(t_u)}{(n-\sum_{v=1}^{u-1} t_v)} q^{-t_u} \leqslant \mathbb{P}[\rho^{(t)} \in H'_k],$$

where $\rho^{(t)}$ is a uniform random permutation of t objects, and H'_k is the set of permutations with all cycles of length no greater than k, in view of (2·1) and (2·6). But now, from (2·1),

$$\mathbb{P}[\rho^{(t)} \in H'_k] \leqslant k^k/\{t(t-k)\ldots(t-(k-1)[t/k])\} \leqslant 1/[t/k]!,$$

where $[\cdot]$ denotes the integer part. Hence (3·5) is of order at most

$$k \sum_{t > n-l} 1/[t/k]! = O(k^2/[(n-l)/k]!) = O\left(k^2 \exp\left\{-\frac{n}{2k}\log\frac{n}{2k}+\frac{n}{2k}\right\}\right), \tag{3·6}$$

provided that $l \leqslant n/2$.

To complete the proof, we need only estimate

$$\mathbb{P}\left[\sum_{j\in J_k} \delta(j)\tilde{X}_j > n/2\right] = \mathbb{P}\left[\sum_{i=1}^{k} i\tilde{Y}_i > n/2\right],$$

where the $\tilde{Y}_i$ are independent negative binomial NB $(N_q(i), q^{-i})$ random variables. Elementary computations, using (2·6) and the inequality

$$(1-p)^{-n} \leqslant \exp\{np+np^2\} \quad \text{in} \quad 0 \leqslant p \leqslant 3/5,$$

show that, for $1 \leqslant z \leqslant \frac{4}{5}q^{\frac{1}{2}} < \frac{3}{5}q$,

$$\mathbb{E}\{z^{\sum_{i=1}^{k} i\tilde{Y}_i}\} = \prod_{i=1}^{k}\left(\frac{1-q^{-i}}{1-z^i q^{-i}}\right)^{N_q(i)} \leqslant \exp\left\{\sum_{i=1}^{k} N_q(i)\,(-q^{-i}+q^{-i}z^i+q^{-2i}z^{2i})\right\}$$

$$\leqslant \exp\left\{\sum_{i=1}^{k} i^{-1}\{(z^i-1)+q^{-i}z^{2i}\}\right\} \leqslant \tfrac{25}{9}\exp\left\{\sum_{i=1}^{k} i^{-1}(z^i-1)\right\} \leqslant 3\exp\{z^k-1\}.$$

Hence, for such z,

$$\mathbb{P}\left[\sum_{i=1}^{k} i\tilde{Y}_i > n/2\right] \leqslant 3\exp\{-(n/2)\log z + z^k - 1\}. \tag{3·7}$$

If $k \geqslant \log n/\log(\frac{4}{5}q^{\frac{1}{2}})$, we can take $z = (n/2k)^{1/k}$ in (3·7), which yields

$$\mathbb{P}\left[\sum_{i=1}^{k} i\tilde{Y}_i > n/2\right] \leqslant 3\,e^{-1}\exp\left\{-\frac{n}{2k}\log\frac{n}{2k}+\frac{n}{2k}\right\} = O\left(\exp\left\{-\frac{n}{2k}\log\frac{n}{2k}+\frac{n}{2k}\right\}\right). \tag{3·8}$$

To obtain the estimate which is valid for all k, take $z = 1+(3k)^{-1} \leqslant 2q/3$ and use the weaker inequality

$$(1-p)^{-n} \leqslant e^{2np} \quad \text{in} \quad 0 \leqslant p \leqslant 2/3,$$

to give

$$\mathbb{P}\left[\sum_{i=1}^{k} i\tilde{Y}_i > n/2\right] \leqslant k\exp\{-(n/2)\log z + e^{\frac{1}{3}} - 1\} = O\left(k\exp\left\{-\frac{n}{2k}\log\frac{4}{3}\right\}\right). \tag{3.9}$$

The theorem follows from (3·5), (3·6), (3·8) and (3·9). ∎

Recall that $\tilde{Y}_i$ are independent negative binomial NB $(N_q(i), q^{-i})$ random variables.

COROLLARY 3·3. *For all* $k \geqslant 1$,

$$d_{TV}(\mathscr{L}(Y_1,\ldots,Y_k), \mathscr{L}(\tilde{Y}_1,\ldots,\tilde{Y}_k)) = O\left(k\exp\left\{-\frac{n}{2k}\log\frac{4}{3}\right\}\right).$$

Proof. This follows from Theorem 3·1 together with the fact that taking functionals never increases total variation distance, so

$$d_{TV}(\mathscr{L}(Y_1,\ldots,Y_k), \mathscr{L}(\tilde{Y}_1,\ldots,\tilde{Y}_k)) \leqslant d_{TV}(\mathscr{L}(X(J_k)), \mathscr{L}(\tilde{X}(J_k))).$$

In fact, we have equality: see Arratia and Tavaré [1]. ∎

Remark 3·4. This last result demonstrates the main difference between the factor structure in a random polynomial and the cycle structure of a random permutation.

In the latter, the numbers of small cycles come very close to having independent Poisson distributions, with mean d^{-1} for cycles of order d. The corresponding approximation for the numbers of factors of small degree is by independent negative binomial distributions, with NB $(N_q(d), q^{-d})$ for the factors of degree d, whose mean, $q^{-d}N_q(d)\,(1-q^{-d})^{-1}$, is nonetheless not too different from d^{-1}.

4. *Factors of medium degree*

In this section, we use the total variation estimate in Corollary 3·3 to study the factors of medium size, proving that the process B_n defined by

$$B_n(t) = \frac{\sum_{i=1}^{[n^t]} Y_i - t\log n}{\surd(\log n)}, \quad 0 \leqslant t \leqslant 1, \tag{4·1}$$

is close to a standard Brownian motion. The basis of the argument is the 'method of the common probability space'. We shall, without further comment, always assume that our space is rich enough to support all our constructions. We begin with the following elementary moment calculation.

LEMMA 4·1. *For* $1 \leqslant k \leqslant n$,

$$\mathbb{E}\sum_{i=k+1}^{n} Y_i \leqslant \mathbb{E}\sum_{i=k+1}^{n} \tilde{Y}_i \leqslant \frac{\log(n/k)}{(1-q^{-k-1})}.$$

Proof. For the left hand inequality, observe that, from (2·5), $\mathbb{E}X_j \leqslant \mathbb{E}\tilde{X}_j$ for all j, and hence that $\mathbb{E}Y_i \leqslant \mathbb{E}\tilde{Y}_i$ for all i. For the right hand inequality,

$$\mathbb{E}\tilde{Y}_i = \frac{N_q(i)\,q^{-i}}{1-q^{-i}} \leqslant \frac{1}{i(1-q^{-i})}$$

follows from (2·6). ∎

LEMMA 4·2. *For each* $n \geqslant 1$, *there is a coupling of* $\{Y_i,\ 1 \leqslant i \leqslant n\}$ *and* $\{\tilde{Y}_i,\ 1 \leqslant i \leqslant n\}$ *such that, if*

$$R_{n,1} \equiv \frac{\sum_{i=1}^{n} |\tilde{Y}_i - Y_i|}{\surd(\log n)},$$

then

$$\mathbb{E}(R_{n,1} \wedge 1) = O\left(\frac{\log\log n}{\surd(\log n)}\right).$$

Proof. Pick

$$k = k(n) = \left[\frac{n\log\frac{4}{3}}{4\log n}\right];$$

then, from Corollary 3·3, there exists a coupling of $(Y_1,\ldots,Y_k)$ and $(\tilde{Y}_1,\ldots,\tilde{Y}_k)$ so that $\mathbb{P}[(Y_1,\ldots,Y_k) \neq (\tilde{Y}_1,\ldots,\tilde{Y}_k)] = O(n^{-1})$. Extend this to a coupling of $(Y_1,\ldots,Y_n)$ and $(\tilde{Y}_1,\ldots,\tilde{Y}_n)$ in any way at all. Then, using Lemma 4·1,

$$\begin{aligned}\mathbb{E}(R_{n,1} \wedge 1) &\leqslant \mathbb{P}[(Y_1,\ldots,Y_k) \neq (\tilde{Y}_1,\ldots,\tilde{Y}_k)] + \sum_{i=k+1}^{n} (\mathbb{E}Y_i + \mathbb{E}\tilde{Y}_i)/\surd(\log n)\\ &= O\left(\frac{\log\log n}{\surd(\log n)}\right). \quad ∎\end{aligned}$$

LEMMA 4·3. *The coupling of Lemma 4·2 can be extended to include a set of independent random variables* $(\hat{Y}_i, 1 \leqslant i \leqslant n)$ *with* $\hat{Y}_i \sim \mathrm{Po}\,(\mathbb{E}\tilde{Y}_i)$, *in such a way that, if*

$$R_{n,2} = \frac{\sum_{i=1}^{n} |\tilde{Y}_i - \hat{Y}_i|}{\sqrt{(\log n)}},$$

then

$$\mathbb{E}(R_{n,2}) = O\left(\frac{1}{\sqrt{(\log n)}}\right).$$

Proof. Given the $\tilde{Y}_i$, then $\hat{Y}_i$ can be constructed in such a way that, for each i,

$$\mathbb{E}\,|\hat{Y}_i - \tilde{Y}_i| = d_W(\hat{Y}_i, \tilde{Y}_i),$$

the Wasserstein distance between the law of $\hat{Y}_i$ and that of $\tilde{Y}_i$. Now

$$\begin{aligned} d_W(\hat{Y}_i, \tilde{Y}_i) &= d_W\left(\mathrm{NB}\,(N_q(i), q^{-i}), \mathrm{Po}\left(\frac{N_q(i)\,q^{-i}}{1-q^{-i}}\right)\right) \\ &\leqslant N_q(i)\, d_W\left(\mathrm{Ge}\,(q^{-i}), \mathrm{Po}\left(\frac{q^{-i}}{1-q^{-i}}\right)\right) \\ &\leqslant N_q(i)\,\frac{2q^{-2i}}{1-q^{-i}}. \end{aligned}$$

This last inequality follows from the estimate

$$d_W(\mathrm{Ge}\,(p), \mathrm{Po}\,(p/(1-p)) \leqslant d_W(\mathrm{Ge}\,(p), \mathrm{Be}\,(p)) + d_W(\mathrm{Be}\,(p), \mathrm{Po}\,(p/(1-p)),$$

because Be (p) is stochastically smaller than the other two distributions, so that the Wasserstein distance is in each case just the difference of the means. Adding over i gives

$$\mathbb{E}\sum_{i=1}^{n} |\hat{Y}_i - \tilde{Y}_i| \leqslant \sum_{i=1}^{n} N_q(i)\, q^{-i}\,\frac{2q^{-i}}{1-q^{-i}} \leqslant \sum_{i=1}^{n} \frac{2q^{-i}}{i(1-q^{-i})} \leqslant -\frac{2\log\,(1-q^{-1})}{(1-q^{-1})},$$

completing the proof. ∎

Now define

$$u_n(t) = \sum_{i=1}^{[n^t]} \mathbb{E}\tilde{Y}_i + (n^t - [n^t])\,\mathbb{E}\tilde{Y}_{[n^t]+1},$$

and observe that, using (2·6) as in the proof of Lemma 4·1,

$$\sup_{0 \leqslant t \leqslant 1} |u_n(t) - t\log n| \leqslant c < \infty$$

for a fixed constant c not depending on n. The partial sums $\sum_{i=1}^{j} \hat{Y}_i$ can then be thought of as the values taken by a Poisson process at times $u_n(t_j)$, where $n^{t_j} = j$. This is the basis for the approximation theorem which follows.

THEOREM 4·4. *It is possible to construct* B_n *and a standard Brownian motion* B *on the same probability space, in such a way that*

$$\mathbb{E}\left\{\sup_{0 \leqslant t \leqslant 1} |B_n(t) - B(t)| \wedge 1\right\} = O\left(\frac{\log\log n}{\sqrt{(\log n)}}\right). \tag{4·2}$$

Proof. Let Z be a Poisson process constructed to satisfy

$$Z(u_n(t)) = \sum_{i=1}^{n^t} \hat{Y}_i$$

for all t such that n^t is integral. A standard Brownian motion $\tilde{B}$ can then be constructed on the same space in such a way that

$$\sup_{t \geqslant 0} \frac{|Z(t)-t-\tilde{B}(t)|}{2 \vee \log t} = K < \infty,$$

where $\mathbb{E}e^{\lambda K} < \infty$ for some $\lambda > 0$, and so, in particular, $\mathbb{E}K < \infty$. This follows from the theorem of Kómlos, Major and Tusnády [**18**]; see also Kurtz [**19**], Lemma 3·1. With this construction,

$$|Z(u_n(t)) - u_n(t) - \tilde{B}(u_n(t))| \leqslant K(2+\log u_n(1)), \quad 0 \leqslant t \leqslant 1. \tag{4·3}$$

Now, by the triangle inequality,

$$|\sqrt{(\log n)}B_n(t) - \tilde{B}(t \log n)| \leqslant |Z(u_n(t)) - u_n(t) - \tilde{B}(u_n(t))| + \left|\sum_{i=1}^{[n^t]} \hat{Y}_i - Z(u_n(t))\right|$$

$$+ \left|\sum_{i=1}^{[n^t]} (\tilde{Y}_i - \hat{Y}_i)\right| + \left|\sum_{i=1}^{[n^t]} (Y_i - \tilde{Y}_i)\right| + |u_n(t) - t \log n| + |\tilde{B}(u_n(t)) - \tilde{B}(t \log n)|, \tag{4·4}$$

and hence, writing $B(t) = \tilde{B}(t \log n)/\sqrt{(\log n)}$,

$$\sup_{0 \leqslant t \leqslant 1} |B_n(t) - B(t)| \leqslant \frac{K(2+\log(c+\log n))}{\sqrt{(\log n)}} + \frac{\max_{1 \leqslant i \leqslant n} \hat{Y}_i}{\sqrt{(\log n)}}$$

$$+ R_{n,1} + R_{n,2} + \frac{c}{\sqrt{(\log n)}} + \frac{\sup_{0 \leqslant t \leqslant 1} |\tilde{B}(u_n(t)) - \tilde{B}(t \log n)|}{\sqrt{(\log n)}}. \tag{4·5}$$

Now we have already established that $\mathbb{E}K < \infty$ and that

$$\mathbb{E}(R_{n,i} \wedge 1) = O\left(\frac{\log \log n}{\sqrt{(\log n)}}\right), \quad i = 1, 2.$$

In addition, it follows easily from Csörgő and Révész [7], Lemma 1.2.1, that

$$\mathbb{E}\left\{\sup_{\substack{0 \leqslant u, v \leqslant \log n + c \\ |u-v| \leqslant c}} |\tilde{B}(u) - \tilde{B}(v)|\right\} = O(\sqrt{(\log \log n)}),$$

and a calculation based on the crude estimate

$$\mathbb{P}\left[\max_{1 \leqslant i \leqslant n} \hat{Y}_i \geqslant r\right] \leqslant \sum_{i=1}^{n} \mathrm{Po}\,(2/i)\{[r, \infty)\}, \quad r \geqslant 2,$$

is enough to show that $\mathbb{E}(\max_{1 \leqslant i \leqslant n} \hat{Y}_i) \leqslant 5$. Equation (4·2) now follows. ∎

Remark 4·5. Theorem 4·4 highlights another similarity between the factor structure of a random polynomial and the cycle structure of a random permutation.

The weak convergence of B_n to B was proved first in the context of random permutations by DeLaurentis and Pittel [8]. The central limit theorem for the total number of factors appears in Flajolet and Soria [14]; this is implied, together with a rate estimate of order $[(\log\log n)/\surd(\log n)]$, by taking $t = 1$ in Theorem 4·4. This rate can actually be improved to order $[1/\surd(\log n)]$, by combining Theorem 6·8 and the Berry–Esseen theorem.

Remark 4·6. Instead of using a Brownian motion as an approximation to B_n, one could equally well use a centred and normalized Poisson process, in the form $\{P(t\log n)-t\log n\}/\surd(\log n)$, in which case no appeal need be made to the Kómlos, Major and Tusnády theorem. In fact, the main conclusion to be drawn from this section, explaining why the factor and cycle structures are alike, is that

$$d_{TV}(\mathscr{L}((Y_i, l \leqslant i \leqslant k)), \mathscr{L}((\hat{Y}_i, l \leqslant i \leqslant k))) = O(n^{-1}), \tag{4·6}$$

where

$$l = l(n) = \log_q n \quad \text{and} \quad k = k(n) = \left[\frac{n\log\frac{4}{3}}{4\log n}\right],$$

the $\hat{Y}_i$s being independent Poisson variates. This estimate, which covers all factors of medium degree, follows directly from the proofs of Lemmas 4·2 and 4·3, since

$$d_{TV}(\mathscr{L}((\tilde{Y}_i, l \leqslant i \leqslant k)), \mathscr{L}((\hat{Y}_i, l \leqslant i \leqslant k))) \leqslant d_W((\tilde{Y}_i, l \leqslant i \leqslant k), (\hat{Y}_i, l \leqslant i \leqslant k)).$$

Another result in the same spirit is given in Theorem 5·8.

5. *Factors of large degree*

Although the joint distribution of the numbers of factors of small degree is not the same as that of the small cycles in a random permutation, the distinction fades as soon as either q or the sizes of the factors become large. For instance, for large n, the distribution of the number of factors of degree d, NB $(N_q(d), q^{-d}) =$ NB $(d^{-1}q^d(1+O(q^{-d/2})), q^{-d})$, is close to Po (d^{-1}) if q^d is large. The results of this section exploit this similarity.

We start by making precise comparisons between (2·1) and (2·4).

LEMMA 5·1. *As for* (2·3), *suppose that* $m_1, \ldots, m_k$ *are positive integers such that* $\sum_{i=1}^k \delta(m_i) \leqslant n$, *and let* $m_1', \ldots, m_l'$ *denote the distinct values taken,* $s_1, s_2, \ldots, s_l$ *their multiplicities. Then*

$$\mathbb{E}\left\{\prod_{j=1}^{l} (X_{m_j'})_{s_j}\right\} \geqslant \prod_{j=1}^{l} s_j!\, q^{-\sum_{i=1}^k \delta(m_i)},$$

and equality holds if $\sum_{i=1}^k \delta(m_i) = n$.

Proof. The lemma follows from (2·5) because

$$\prod_{j=1}^{l} (X_{m_j'})_{s_j} \geqslant \prod_{j=1}^{l} \{s_j!\, I[X_{m_j'} \geqslant s_j]\} = \left\{\prod_{j=1}^{l} s_j!\right\} I\left[\bigcap_{j=1}^{l} \{X_{m_j'} \geqslant s_j\}\right],$$

with equality if $\sum_{i=1}^k \delta(m_i) = n$. ∎

COROLLARY 5·2. *If $d_i \geqslant 1$ for each i and $\sum_{i=1}^k d_i \leqslant n$, then*

$$\mathbb{P}[\delta(M_1) = d_1, \ldots, \delta(M_k) = d_k] \geqslant \prod_{i=1}^{k} \left(n - \sum_{j=1}^{i-1} d_j\right)^{-1} \left\{1 - 2 \sum_{\substack{i=1 \\ d_i \geqslant 2}}^{k} q^{-d_i/2}\right\}.$$

Proof. Adding over the possible choices of $m_1, \ldots, m_k$ consistent with the degree sequence $d_1, \ldots, d_k$ gives

$$\mathbb{P}[\delta(M_1) = d_1, \ldots, \delta(M_k) = d_k] \geqslant \left\{\prod_{i=1}^{k} \frac{d_i}{n - \sum_{j=1}^{i-1} d_j}\right\} q^{-\sum_{i=1}^k d_i} \prod_{i=1}^{k} N_q(d_i), \tag{5·1}$$

and the corollary follows using (2·6). ∎

LEMMA 5·3. *With the notation of Lemma 5·1,*

(i) $$\sum_{m_1 : \delta(m_1) = d} \cdots \sum_{m_j : \delta(m_j) = d} \left[\prod_{i=1}^{l} (X_{m_i'})_{s_i}\right] = \left(\sum_{m : \delta(m) = d} X_m\right)_j,$$

and

(ii) $$\sum_{m_1 : \delta(m_1) = d_1} \cdots \sum_{m_k : \delta(m_k) = d_k} \mathbb{E}\left\{\prod_{j=1}^{l} (X_{m_j'})_{s_j}\right\} \leqslant \prod_{i=1}^{k} \left(\frac{q^{-d_i} N_q(d_i)}{1 - q^{-d_i}}\right).$$

Proof. The first part consists of two different ways of counting the choices of j objects from a total of $\sum_m X_m$, order being distinguished: in the former, they are enumerated by first accounting for their m-grouping.

For the second, we use the fact that $\mathbb{E}\prod_{j=1}^{l} (Z_j)_{s_j}$ is an increasing function of the joint tail probabilities $\mathbb{P}[Z_1 \geqslant z_1, \ldots, Z_l \geqslant z_l]$, so that, from (2·5),

$$\mathbb{E}\left\{\prod_{j=1}^{l} (X_{m_j'})_{s_j}\right\} \leqslant \mathbb{E}\left\{\prod_{j=1}^{l} (\tilde{X}_{m_j'})_{s_j}\right\}, \tag{5·2}$$

where the $\tilde{X}_j$s are independent geometric $\mathrm{Ge}(q^{-\delta(j)})$ random variables. Hence, if $d_1', \ldots, d_t'$ denote the distinct d_i-values, and $u_1, \ldots, u_t$ their multiplicities, the first part yields

$$\sum_{m_1 : \delta(m_1) = d_1} \cdots \sum_{m_k : \delta(m_k) = d_k} \mathbb{E}\left\{\prod_{j=1}^{l} (X_{m_j'})_{s_j}\right\} \leqslant \mathbb{E}\left\{\prod_{v=1}^{t} \left(\sum_{m : \delta(m) = d_v'} \tilde{X}_m\right)_{u_v}\right\}$$
$$= \prod_{v=1}^{t} \left(\frac{q^{-d_v'} N_q(d_v')}{1 - q^{-d_v'}}\right)^{u_v} = \prod_{i=1}^{k} \left(\frac{q^{-d_i} N_q(d_i)}{1 - q^{-d_i}}\right). \quad ∎$$

COROLLARY 5·4. *If $d_1, \ldots, d_k \geqslant 1$ and $\sum_{i=1}^k d_i \leqslant n$,*

$$\mathbb{P}[\delta(M_1) = d_1, \ldots, \delta(M_k) = d_k] \leqslant \left\{\prod_{i=1}^{k} \left(n - \sum_{j=1}^{i-1} d_j\right)^{-1}\right\} \exp\left(\frac{3}{2} \sum_{i=1}^{k} q^{-d_i}\right).$$

Proof. This follows from (2·3) and Lemma 5·3 (ii), using the inequality $(1-x)^{-1} \leqslant \exp\{3x/2\}$ in $0 \leqslant x \leqslant \frac{1}{2}$. ∎

The comparisons of probabilities in Corollaries 5·2 and 5·4 lead immediately to the following comparison between the factor and cycle processes, which shows that they are close in distribution if q is large.

THEOREM 5·5. *As for* (2·1)–(2·3), *let* $(\Delta_i)_{1\leqslant i\leqslant K_0}$ *denote the sequence of orders of the cycles obtained by size-biassed sampling from a uniform random permutation,* $(M_i)_{1\leqslant i\leqslant K_0'}$ *the corresponding sequence of irreducible factors. Then*

$$d_{TV}(\mathscr{L}(\{\delta(M_i)\}_{1\leqslant i\leqslant K_0'}), \mathscr{L}(\{\Delta_i\}_{1\leqslant i\leqslant K_0})) \leqslant 2\sum_{l\geqslant 2} l^{-1} q^{-l/2}.$$

Proof. Let

$$\chi = \{(k, (d_i)_{1\leqslant i<k}); k\geqslant 2, d_i \geqslant 1 \quad \text{for each} \quad i\}.$$

From Corollary 5·2, we have

$$\begin{aligned} &d_{TV}(\mathscr{L}(\{\delta(M_i)\}_{1\leqslant i\leqslant K_0'}), \mathscr{L}(\{\Delta_i\}_{1\leqslant i\leqslant K_0})) \\ &\quad = \sup_{A\subset\chi}\sum_{x\in A} (\mathbb{P}[\{\Delta_i\}_{1\leqslant i\leqslant K_0} = x] - \mathbb{P}[\{\delta(M_i)\}_{1\leqslant i\leqslant K_0'} = x]) \\ &\quad \leqslant \sup_{A\subset\chi}\sum_{x\in A} (\mathbb{P}[\{\Delta_i\}_{1\leqslant i\leqslant K_0} = x].2 \sum_{\substack{i=1\\ d_i\geqslant 2}}^{k} q^{-d_i/2} \\ &\quad \leqslant 2\mathbb{E}\Big(\sum_{1\leqslant i\leqslant K_0} q^{-\frac{1}{2}\Delta_i} I[\Delta_i \geqslant 2]\Big) \\ &\quad = 2\sum_{j=2}^{n} q^{-j/2}\,\mathbb{E}C_j, \end{aligned}$$

where $C_j = \#\{i : \Delta_i = j\}$. The theorem follows because $\mathbb{E}C_j = j^{-1}$ for all $j\leqslant n$. ∎

COROLLARY 5·6. *For the counts of factors of different degrees and of cycles of different orders, we have*

$$d_{TV}(\mathscr{L}(Y_1,\ldots,Y_n), \mathscr{L}(C_1,\ldots,C_n)) \leqslant q^{-1} + O(q^{-\frac{3}{2}}). \quad ∎$$

Remark 5·7. It follows from Corollary 3·3 with $k = 1$ and Feller [13], chapter 4·4, that

$$d_{TV}(\mathscr{L}(Y_1), \mathscr{L}(C_1)) \geqslant e^{-1} - (1-q^{-1})^q - 1/(n+1)! - O(e^{-n\alpha}) \geqslant cq^{-1} - O(e^{-n\alpha}),$$

where $\alpha = \frac{1}{2}\log\frac{4}{3}$, so that the q-order of approximation in Corollary 5·6 cannot be improved.

If q is not large, it still makes sense to approximate the joint distribution of the large factors. For $1\leqslant r\leqslant n$, take $\chi\equiv\chi(r)$ to consist of elements $(k, (d_i)_{i=1}^k)$ such that now $k\geqslant 1$, $d_i\geqslant 1$ for each i, $\sum_{i=1}^{k-1} d_i \leqslant n-r$ and $n-r < \sum_{i=1}^{k} d_i \leqslant n$. Let $(K_r, (\Delta_i)_{1\leqslant i\leqslant K_r})$ denote the random element of χ obtained by taking the cycle lengths sampled as above, but stopping when fewer than r objects are left to be permuted, and let $(K_r', (\delta(M_i))_{1\leqslant i\leqslant K_r'})$ be the corresponding random element derived from the factor process. Then

$$\mathbb{E}\Big\{\sum_{i=1}^{K_r} q^{-\frac{1}{2}\Delta_i} I[\Delta_i\geqslant 2]\Big\} = \sum_{j\geqslant 2} q^{-j/2}\,\mathbb{E}C_{jr},$$

where $C_{jr} = \#\{i\leqslant K_r : \Delta_i = j\}$. Furthermore, using the record value description to compute the expectations,

$$\mathbb{E}C_{jr} = \frac{1}{n} + \sum_{l=(r+1)\vee(j+1)}^{n} \frac{1}{l(l-1)} = \frac{1}{r}\wedge\frac{1}{j}.$$

In consequence, we have the following result.

THEOREM 5·8. *For the size-biassed cycle lengths and factor sizes,*

$$d_{TV}(\mathscr{L}(\{\delta(M_i)\}_{1\leqslant i\leqslant K_r'}), \mathscr{L}(\{\Delta_i\}_{1\leqslant i\leqslant K_r})) \leqslant r^{-1}c_{q1},$$

for any $1 \leqslant r \leqslant n$, *where* $c_{q1} = 2q^{-1}/(1-q^{-\frac{1}{2}})$. *For the counts of factors of large degree and of cycles of large order,*

$$d_{TV}(\mathscr{L}(\{Y_j\}_{j\geqslant r}), \mathscr{L}(\{C_j\}_{j\geqslant r})) \leqslant r^{-1}c_{q1}.$$

Proof. Use Corollary 5·2 as for Theorem 5·5, and note that $\Delta_i < r$ for all $i > K_r$. ∎

The size-biassed sequence of factor degrees can be viewed in terms of a random splitting of the unit interval. The elements $n^{-1}\delta(M_1)$, $n^{-1}\delta(M_2), \ldots$, are thought of as lengths successively removed from $[0,1]$, corresponding to cutting at the points $1-n^{-1}\delta(M_1)$, $1-n^{-1}(\delta(M_1)+\delta(M_2))$, and so on. An alternative splitting is obtained by cutting at the points $1-n^{-1}L_1$, $1-n^{-1}(L_1+L_2)$, and so on, where $L_1 \geqslant L_2 \geqslant \ldots$ denote the degrees of the factors arranged in descending order. In either formulation, there is a natural limit in distribution as $n\to\infty$, in the former case the GEM distribution with parameter 1, and in the latter the Poisson–Dirichlet distribution with parameter 1. In the remainder of the section, the consequences of Theorem 5·8 are investigated in this framework.

Any factor splitting of the unit interval can be represented as a finite decreasing sequence of rationals $1 > x_1 > x_2 > \ldots > 0$, or, equivalently, as the associated atomic measure $\mu = \sum_{j\geqslant 1}\delta_{x_j}$. Let the set of such measures be denoted by $\mathscr{H}$, and define a metric d_H on $\mathscr{H}$ by

$$d_H(\mu,\nu) = \inf\{t>0 : \mu\{(t,1)\} = \nu\{(t,1)\}\} \leqslant 1. \tag{5·3}$$

The space $(\mathscr{H}, d_H)$ is a natural choice for the distributional approximation of factor splittings by cycle splittings, because of Theorem 5·8, but has drawbacks as far as limiting procedures are concerned: it does not support the GEM or Poisson–Dirichlet distributions; removing the restriction of the cut points to the rationals, so as to include these distributions, would give a non-separable space; and, in any case, because these distributions give zero probability to rational cut points, the d_H distance between the 'limiting' distributions and the factor distributions for finite n would not approach zero as $n\to\infty$. So take $\mathscr{G} \supset \mathscr{H}$ to be the set of measures of the form $\mu = \sum_{j\geqslant 1} w_j\delta_{x_j}$, where the w_j are elements of $\mathbb{N}$, and where $1 > x_1 > x_2 > \ldots > 0$ is any possibly infinite sequence of reals which does not accumulate except, if infinite, at 0. Equivalently, μ can be represented as $\sum_{l\geqslant 1}\delta_{y_l}$, where $1 > y_1 \geqslant y_2 \geqslant \ldots > 0$ does not accumulate except perhaps at 0, so that the w_js are replaced by the repeats in the y-sequence. Then define a metric d_G on $\mathscr{G}$ by

$$d_G(\mu,\nu) = \inf\ \{t>0 : |\mu(f)-\nu(f)| \leqslant t \quad \text{for all} \quad f\in\mathscr{E}_t\} \leqslant 1, \tag{5·4}$$

where

$$\mathscr{E}_t = \{f\in C(0,1) : \sup_x |f(x)| \leqslant 1;\ \sup_{x\neq y}\{|f(x)-f(y)|/|x-y|\} \leqslant 1;\ f(x)=0 \quad \text{for} \quad 0\leqslant x\leqslant t\}. \tag{5·5}$$

Clearly, when restricted to $\mathscr{H}$, $d_G \leqslant d_H$. If $d_G(\mu,\nu) < \epsilon$, then μ and ν must be close on $[2\epsilon, 1)$ at least in the (Lévy–Prohorov) sense that, for all $A \subset [2\epsilon, 1)$, $\nu(A^\epsilon) \geqslant \mu(A)$ and vice versa; if $d_H(\mu,\nu) < \epsilon$, then μ and ν must agree exactly on $(\epsilon, 1)$; neither statement implies anything about μ and ν on $(0,\epsilon)$. Thus d_G is somewhat less sensitive than d_H to small changes in the positions of the point masses: for example, if $\epsilon < \frac{1}{2}$,

$$d_G(\delta_{\frac{1}{2}}, \delta_{\epsilon+\frac{1}{2}}) = \epsilon, \quad \text{but} \quad d_H(\delta_{\frac{1}{2}}, \delta_{\epsilon+\frac{1}{2}}) = \epsilon+\tfrac{1}{2}.$$

An element $\mu = \sum_{l\geqslant 1} \delta_{y_l}$ of $\mathscr{G}$, where $1 > y_1 \geqslant y_2 \geqslant \ldots > 0$, can immediately be identified with an element $\tilde{\mu} = (y_1, y_2, \ldots)$ of $[0,1]^\infty$ if the y-sequence is infinite: if it is finite, fill out $\tilde{\mu}$ with zeros. The metric d_G can then be compared with metrics induced by those on $[0,1]^\infty$, such as that given by the following metric for the product topology:

$$d(\mu^{(1)}, \mu^{(2)}) = \sum_{j\geqslant 1} 2^{-j} |y_j^{(1)} - y_j^{(2)}| \leqslant 1. \tag{5·6}$$

Now if $d_G(\mu^{(1)}, \mu^{(2)}) < \epsilon$ and $y_j^{(1)} \vee y_j^{(2)} \geqslant 2\epsilon$, then consideration of the function $f \in \mathscr{E}_\epsilon$ defined by

$$f(x) = \epsilon \wedge (x + \epsilon - y_j^{(1)} \vee y_j^{(2)})^+$$

shows that $|y_j^{(1)} - y_j^{(2)}| < \epsilon$. On the other hand, if $y_j^{(1)} \vee y_j^{(2)} < 2\epsilon$ then clearly $|y_j^{(1)} - y_j^{(2)}| < 2\epsilon$. Hence $d_G(\mu^{(1)}, \mu^{(2)}) < \epsilon$ puts a uniform bound of 2ϵ on the component differences, and

$$d(\mu^{(1)}, \mu^{(2)}) \leqslant 2 d_G(\mu^{(1)}, \mu^{(2)}). \tag{5·7}$$

There can be no comparable inequality in the other direction, because $d(\mu^{(1)}, \mu^{(2)}) < \epsilon$ sets no limit on $\max_{j\geqslant 1} |y_j^{(1)} - y_j^{(2)}|$, and thus rate estimates expressed in terms of d_G, if obtainable, seem preferable to rates in terms of d, because of the extra control that they imply. However, $d(\mu^{(n)}, \mu) \to 0$ easily implies that $d_G(\mu^{(n)}, \mu) \to 0$, so that d_G and d are topologically equivalent. The space $(\mathscr{G}, d_G)$ is thus separable, and, in view of (5·7), is also complete.

Now let $\Theta_P^{(n)}$ and $\Theta_F^{(n)}$ denote the random elements of $\mathscr{H}$ corresponding to the size-biassed cycle lengths and factor degrees respectively: thus, in the notation of Section 2,

$$\Theta_P^{(n)} = \sum_{j=1}^{K_0-1} \delta_{x_j}; \quad x_j = 1 - n^{-1} \sum_{l=1}^{j} \Delta_l \tag{5·8}$$

and

$$\Theta_F^{(n)} = \sum_{j=1}^{K_0'-1} \delta_{x_j'}; \quad x_j' = 1 - n^{-1} \sum_{l=1}^{j} \delta(M_l). \tag{5·9}$$

Then define the size-ordered cycle and factor processes by

$$\Phi_P^{(n)} = \sum_{j=1}^{K_0-1} \delta_{y_j}; \quad y_j = 1 - n^{-1} \sum_{l=1}^{j} \Delta_{[l]}, \tag{5·10}$$

where $\Delta_{[1]} \geqslant \Delta_{[2]} \geqslant \ldots \geqslant \Delta_{[K_0]}$ are the cycle lengths in descending order, and likewise

$$\Phi_F^{(n)} = \sum_{j=1}^{K_0'-1} \delta_{y_j'}; \quad y_j' = 1 - n^{-1} \sum_{l=1}^{j} L_l, \tag{5·11}$$

where L_l denotes the degree of the factor of lth largest degree. Finally, let Θ and Φ be

random elements of $\mathscr{G}$ corresponding to the GEM and Poisson–Dirichlet processes with parameter 1: these can be constructed by setting

$$\Theta = \sum_{j \geqslant 1} \delta_{v_j}; \quad v_j = \prod_{l=1}^{j} U_l, \tag{5·12}$$

where $(U_l, l \geqslant 1)$ are independent uniform U[0, 1] random variables, and

$$\Phi = \sum_{j \geqslant 1} \delta_{v'_j}; \quad v'_j = 1 - \sum_{l=1}^{j} z_l, \tag{5·13}$$

where z_l denotes the lth largest of the differences $v_{j-1} - v_j$, with $v_0 = 1$ and the other v_js as for Θ. With these definitions, the processes of interest are expressed as elements of $\mathscr{H}$ or $\mathscr{G}$. We now compare their distributions, using the metrics d_H and d_G and suitable couplings.

The first result is merely a re-formulation of Theorem 5·8.

THEOREM 5·9. *The processes $\Theta_P^{(n)}$ and $\Theta_F^{(n)}$ can be constructed on the same probability space in such a way that, for any $1 \leqslant r < n$,*

$$\mathbb{P}[d_H(\Theta_P^{(n)}, \Theta_F^{(n)}) > r/n] \leqslant c_{q1}\, r^{-1}.$$

Hence also

$$\mathbb{E}[d_H(\Theta_P^{(n)}, \Theta_F^{(n)})] = O(n^{-1} \log n). \quad \blacksquare$$

The next result compares the size-ordered processes of cycles and factors.

THEOREM 5·10. *The processes $\Phi_P^{(n)}$ and $\Phi_F^{(n)}$ can be constructed on the same probability space in such a way that, for any $1 \leqslant r < n$,*

$$\mathbb{P}[d_H(\Phi_P^{(n)}, \Phi_F^{(n)}) > r/n] \leqslant c_{q2}\, r^{-\frac{1}{2}},$$

where $c_{q2} = 2\sqrt{c_{q1}}$. Hence also

$$\mathbb{E}[d_H(\Phi_P^{(n)}, \Phi_F^{(n)})] = O(n^{-\frac{1}{2}}).$$

Proof. Construct $\Theta_P^{(n)}$ and $\Theta_F^{(n)}$ as in Theorem 5·9, and derive realizations of $\Phi_P^{(n)}$ and $\Phi_F^{(n)}$ from them by the appropriate re-ordering. Then observe that, for $k \leqslant r$, if $\Theta_P^{(n)}$ is the same as $\Theta_F^{(n)}$ on $[n^{-1}k, 1)$, and if all the cycle lengths Δ_l used to construct $\Theta_P^{(n)}$ on the interval $[n^{-1}r, 1)$ are of length at least k, then $\Phi_P^{(n)} = \Phi_F^{(n)}$ on $[n^{-1}r, 1)$. Hence

$$\mathbb{P}[d_H(\Phi_P^{(n)}, \Phi_F^{(n)}) > r/n] \leqslant \mathbb{P}[d_H(\Theta_P^{(n)}, \Theta_F^{(n)}) > k/n] + \mathbb{P}[\min_{1 \leqslant i \leqslant K_r} \Delta_i < k]. \tag{5·14}$$

The first of these probabilities can be bounded above by $c_{q1}\, k^{-1}$, using Theorem 5·9. For the second, use the Bernoulli construction at the beginning of Section 2 to bound it above by

$$\sum_{j=r+1}^{n} \mathbb{E}(I_j\{1 - I[I_l = 0, j-k \leqslant l \leqslant j-1]\}) = \sum_{j=r+1}^{n} \frac{k}{j(j-1)} = \frac{k}{r}. \tag{5·15}$$

Now take $k = (c_{q1}\, r)^{\frac{1}{2}}$. $\blacksquare$

We now turn to approximation by the limit processes. In view of Theorems 5·9 and 5·10, it is enough to work either with factors or with cycles, and we choose the latter, because the structure is simpler.

THEOREM 5·11. *It is possible to construct Θ and $\Theta_P^{(n)}$ on the same probability space, in such a way that, for $4 \log n \leqslant r < n$,*

$$\mathbb{P}[d_G(\Theta_P^{(n)}, \Theta) > n^{-1}(r+2)] \leqslant r^{-1} + n^{-1}.$$

Hence also

$$\mathbb{E}[d_G(\Theta_P^{(n)}, \Theta)] = O(n^{-1} \log n).$$

Proof. Let $(U_l, l \geqslant 1)$ be independent $\mathrm{U}[0,1]$ random variables, and define $V_j = \prod_{l=1}^{j} U_l$ for $j \geqslant 1$, and Θ as in (5·12). We use the second construction of Section 2 as the basis for an explicit coupling of

$$\tilde{V}_j \overset{\mathscr{D}}{=} n^{-1}(\tilde{T}_{j+1} - 1), \quad j \geqslant 1,$$

with the V_js, in such a way that $0 \leqslant V_j - \tilde{V}_j \leqslant n^{-1}$ for as long as possible. To start with, sample $U_1 = V_1$, and set $\tilde{V}_1 = n^{-1}[nV_1]$, where $[x]$ denotes the integer part of x. Then $0 \leqslant V_1 - \tilde{V}_1 \leqslant n^{-1}$, and V_1 and $\tilde{V}_1$ have the right distributions. Set $J_1 = 0$, and, if $\tilde{V}_1 = 0$, set $K_0 = 1$.

The construction now proceeds inductively. Given $\{(V_i, \tilde{V}_i, J_i),\ 1 \leqslant i < j\}$, sample U_j to give $V_j = U_j V_{j-1}$. If $\tilde{V}_{j-1} = 0$, set $\tilde{V}_j = 0$. If $\tilde{V}_{j-1} > 0$ and $J_{j-1} = 0$, set $\tilde{V}_j = n^{-1}[nV_j]$ and $J_j = 0$ if

$$V_j \in n^{-1}(0, [nV_{j-1}]) = (0, \tilde{V}_{j-1}).$$

Otherwise, set $J_j = 1$ and sample $\tilde{V}_j$ uniformly from $n^{-1}\{1, 2, \ldots, n\tilde{V}_{j-1} - 1\}$. If $\tilde{V}_{j-1} > 0$ and $\tilde{V}_j = 0$, set $K_0 = j$. As a result of this construction, $0 \leqslant V_j - \tilde{V}_j \leqslant n^{-1}$ for as long as $J_j = 0$, and the sequences $(V_j,\ j \geqslant 1)$ and $(\tilde{V}_j,\ 1 \leqslant j < K_0)$ have the correct distributions, generating realizations of Θ and $\Theta_P^{(n)}$ respectively. Furthermore,

$$d_G(\Theta_P^{(n)}, \Theta) \leqslant \inf\{\epsilon > \tau : n^{-1} N_\epsilon < \epsilon\}, \tag{5·16}$$

where $\tau = \max_{j: J_j = 1} V_j$ and $N_\epsilon = \#\{j : V_j > \epsilon\}$.

Now, in view of the first construction of Section 2, we have

$$\mathbb{P}\left[\tau \geqslant \frac{r+1}{n}\right] \leqslant \sum_{j \geqslant r} j^{-1}\, \mathbb{E} I_j = \sum_{j \geqslant r} j^{-2} \leqslant \frac{1}{r-1}, \tag{5·17}$$

where we define

$$I_j = I[j \in n\{\tilde{V}_1, \tilde{V}_2, \ldots\}];$$

and $N_\epsilon \sim \mathrm{Po}\,(\log(1/\epsilon))$. Thus, from (5·16),

$$\begin{aligned}\mathbb{P}[d_G(\Theta_P^{(n)}, \Theta) > n^{-1}(r+2)] &\leqslant \mathbb{P}[\tau \geqslant n^{-1}(r+2)] + \mathbb{P}[n^{-1} N_{n^{-1}(r+2)} \geqslant n^{-1}(r+2)] \\ &\leqslant r^{-1} + \mathbb{P}[\mathrm{Po}\,(\log(n/r)) \geqslant r] \leqslant r^{-1} + n^{-1},\end{aligned}$$

because, for $r \geqslant 4 \log n$, from Barbour, Holst and Janson[4], Proposition A 2·3,

$$\mathbb{P}[\mathrm{Po}\,(\log(n/r)) \geqslant r] \leqslant \mathbb{P}[\mathrm{Po}\,(\log n) \geqslant 4 \log n] \leqslant n^{-1}. \quad \blacksquare$$

To compare $\Phi_P^{(n)}$ with Φ, we combine the coupling of Theorem 5·11 with the argument of Theorem 5·10.

THEOREM 5·12. *It is possible to construct Φ and $\Phi_P^{(n)}$ on the same probability space, in such a way that, if n and r satisfy $8 \log^2 n + 2 \log n \leqslant r < n$, then*

$$\mathbb{P}[d_G(\Phi_P^{(n)}, \Phi) > r/n] \leqslant 2r^{-\frac{1}{2}} + 3r^{-1} + n^{-1}.$$

Hence also

$$\mathbb{E}[d_G(\Phi_P^{(n)}, \Phi)] = O(n^{-\frac{1}{2}}).$$

Proof. Construct $\Phi_P^{(n)}$ and Φ from $\Theta_P^{(n)}$ and Θ of the previous theorem. Then observe that, for $4 \leqslant k \leqslant r < n$,

$$\mathbb{P}\left[d_G(\Phi_P^{(n)}, \Phi) > \frac{r}{n}\right] \leqslant \mathbb{P}\left[\tau \geqslant \frac{k-1}{n}\right] + \mathbb{P}\left[\frac{1}{2n} N_{k/n}(N_{k/n}+1) \geqslant \frac{r}{n}\right] + \mathbb{P}\left[\min_{1 \leqslant i \leqslant K_r} \Delta_i < k-1\right],$$

where the middle term arises because, although the matched intervals in the $\Theta_P^{(n)}$ and Θ processes on $(\tau \vee (k/n), 1)$ differ in length by at most n^{-1}, the re-ordering to $\Phi_P^{(n)}$ and Φ can result in cumulative differences in the positions of their atoms. For $r \geqslant 8\log^2 n + 2\log n$, the middle term is less than n^{-1} as before, and the theorem follows by taking $k = [\sqrt{r}]+4$, and using (5·15) and (5·17). ∎

Remark 5·13. The Wasserstein–Kantorovich metric ρ_G on probability measures over $\mathscr{G}$ can be defined by

$$\rho_G(P, Q) = \inf \mathbb{E} d_G(X, Y),$$

where the infimum is taken over all couplings of random elements $X \sim P$ and $Y \sim Q$ of $\mathscr{G}$: see Rachev [23], Chapters 5 and 6. We have thus shown that

$$\rho_G(\mathscr{L}(\Theta_F^{(n)}), \mathscr{L}(\Theta_P^{(n)})),\ \rho_G(\mathscr{L}(\Theta_P^{(n)}),\ \mathscr{L}(\Theta)), \rho_G(\mathscr{L}(\Theta_F^{(n)}), \mathscr{L}(\Theta)) = O(n^{-1}\log n); \quad (5{\cdot}18)$$

$$\rho_G(\mathscr{L}(\Phi_F^{(n)}), \mathscr{L}(\Phi_P^{(n)})),\ \rho_G(\mathscr{L}(\Phi_P^{(n)}), \mathscr{L}(\Phi)),\ \rho_G(\mathscr{L}(\Phi_F^{(n)}), \mathscr{L}(\Phi)) = O(n^{-\frac{1}{2}}). \quad (5{\cdot}19)$$

Since d_G is topologically equivalent to d as defined in (5·6), these results sharpen the convergence theorems of Shepp and Lloyd [25] and Vershik and Shmidt [26] for random permutations, and of Diaconis, McGrath and Pitman [10] for random polynomials, and also that of Hansen [16] when applied to these structures.

6. *The total number of factors*

In this section, we are concerned with the distribution of K_0', the total number of factors. We begin with sharp estimates of the point probabilities $\mathbb{P}[K_0' = k]$, in the spirit of the estimates obtained by Car [5], using generating functions.

THEOREM 6·1. *Let* $\psi(n+1) = \sum_{j=1}^n j^{-1}$. *Then*

$$\mathbb{P}[K_0' = k] = \mathbb{P}\left[\sum_{j=1}^n C_j = k\right](1+\epsilon_k), \quad k \geqslant 1,$$

where

$$|\epsilon_1| \leqslant 2q^{-n/2}; \quad |\epsilon_k| \leqslant c\frac{(k-1)}{q\psi(n)}\left(1 - \frac{k-1}{\psi(n)}\right)^{-1}, \quad k \geqslant 2,$$

and c is a universal constant.

Remark 6·2. For $k \geqslant 2$,

$$\mathbb{P}\left[\sum_{j=1}^n C_j = k\right] = \frac{\psi(n)^{k-1}}{n(k-1)!}\left\{1 + O\left(\left(\frac{k-1}{\log n}\right)^2\right)\right\} = \frac{\log^{k-1} n}{n(k-1)!}\left\{1 + O\left(\frac{k-1}{\log n}\right)\right\};$$

for $k = 1$, $\mathbb{P}[\sum_{j=1}^n C_j = 1] = n^{-1}$.

Proof. For $k = 1$, the result follows from (2·4). For $k \geqslant 2$, Corollaries 5·2 and 5·4 imply that

$$\left|\mathbb{P}[K_0' = k] - \mathbb{P}\left[\sum_{j=1}^{n} C_j = k\right]\right| \leqslant 2 \sum_{i=1}^{k} \sum_{d_1+\ldots+d_k=n} \frac{q^{-(\frac{1}{2}d_i \vee 1)}}{n(n-d_1)\ldots(n-\sum_{j=1}^{k-1} d_j)}. \tag{6·1}$$

The contribution to the sums from $i = k$ is easily bounded by

$$\frac{1}{n}\sum_{j=1}^{n} j^{-1} q^{-(\frac{1}{2}j \vee 1)} \frac{\psi(n)^{k-2}}{(k-2)!} \leqslant \frac{\psi(n)^{k-2}}{n(k-2)!} q^{-1} c_1, \tag{6·2}$$

where

$$c_1 = \sum_{j=1}^{\infty} q^{-(\frac{1}{2}j-1)^+} = 1 + \frac{1}{1-q^{-\frac{1}{2}}}. \tag{6·3}$$

For $i = k-l$, $1 \leqslant l \leqslant k-2$, let j denote the value of $(n-\sum_{s=1}^{k-l} d_s)$, and bound the contribution to the sum by

$$\frac{\psi(n)^{k-2-l}}{n(k-2-l)!} \sum_{j=1}^{n} \frac{\psi(j)^{l-1}}{j^2(l-1)!} q^{-1} c_1 \leqslant \frac{\psi(n)^{k-2-l}}{n(k-2-l)!} q^{-1} c_1 c_2, \tag{6·4}$$

where

$$1 \leqslant c_2 = \max_{l \geqslant 0} \sum_{j=1}^{\infty} \frac{\psi(j)^l}{l!\, j^2} < \infty. \tag{6·5}$$

For $i = 1$, bound the contribution to the sum by

$$\frac{\psi(n)^{k-2}}{n(k-2)!} \sum_{j=1}^{n-1} \frac{q^{-(\frac{1}{2}j \vee 1)}}{n-j} \sim \frac{\psi(n)^{k-2}}{n^2(k-2)!} q^{-1} c_1. \tag{6·6}$$

The theorem now follows, with $c = 4c_1 c_2$. ∎

Theorem 6·1 and Remark 6·2 show that $\mathscr{L}(K_0'-1)$ is very close to $\mathrm{Po}(\log n)$ in the lower tail factors. We now show that $\mathscr{L}(K_0'-1)$ is close to $\mathrm{Po}(\log n)$ over the whole range. Indeed, taking $t = 1$ in Theorem 4·4 is already enough to show that they are close, of order $O((\log\log n)/\sqrt{(\log n)})$, with respect to Dudley's metric

$$d(P, Q) = \sup_{f:\|f\|\leqslant 1, \|f'\|\leqslant 1} \left|\int f\,dP - \int f\,dQ\right|.$$

However, it is also natural to ask how good total variation approximation of K_0' by $\mathrm{Po}(\psi(n))$ is. This is not only because of Car's sharp tail estimates, but also because a corresponding approximation of the total number of cycles in a random permutation, to order $(\log n)^{-1}$, can be derived using the construction from independent indicators. Here, we use a rather complicated argument, based on the Stein–Chen method and a coupling, to obtain an approximation of order $(\log n)^{-\frac{1}{2}}$ in total variation.

The first step is again to compare certain pairs of distributions. The results of the comparisons are then used to show that a particular coupling is exact with high probability. Let $u, j \geqslant 1$ be such that $j+u \leqslant n$, and let $v = j+u$. The symbol η is used to denote any quantity of order $q^{-j}+q^{-u}$, the implied constants being universal. Thus, for instance, the inequalities

$$q^{-\delta(i)-\delta(l)} \leqslant \mathbb{E}(X_i X_l) \leqslant \mathbb{E}(\tilde{X}_i \tilde{X}_l) = \frac{q^{-\delta(i)-\delta(l)}}{(1-q^{-\delta(i)})(1-q^{-\delta(l)})} \tag{6·7}$$

can be used to infer the statement

$$\mathbb{E}(X_i X_l) = q^{-\delta(i)-\delta(l)}(1+\eta), \tag{6·8}$$

whenever $(\delta(i) \wedge \delta(l)) \geqslant (j \wedge u)$. Let m_α, m_β and m_γ be irreducible polynomials of degrees v, j, u respectively.

LEMMA 6·3. *If* $\{m_2, \ldots, m_k\} \cap \{m_\alpha, m_\beta, m_\gamma\} = \varnothing$ *and* $\sum_{i=2}^k \delta(m_i) + v = n$, *then*

$$\log\left\{\frac{\mathbb{P}[M_2 = m_2, \ldots, M_k = m_k \,|\, M_1 = m_\alpha]}{\mathbb{P}[M_3 = m_2, \ldots, M_{k+1} = m_k \,|\, M_1 = m_\beta, M_2 = m_\gamma]}\right\} = \eta.$$

Note that the constants implied by η, *being universal, are the same for all* m_i *etc.*

Proof. Direct computation shows that under the given conditions, the ratio is just $\mathbb{E}(X_{m_\beta} X_{m_\gamma})/\mathbb{E}X_{m_\alpha}$, from which the assertion follows. ∎

LEMMA 6·4.

$$\mathbb{P}[X_{m_\alpha} \geqslant 2 \,|\, M_1 = m_\alpha] = \eta; \quad \mathbb{P}[X_{m_\beta} + X_{m_\gamma} \geqslant 1 \,|\, M_1 = m_\alpha] = \eta.$$

Proof. The former probability does not exceed $\mathbb{E}\{X_{m_\alpha}(X_{m_\alpha}-1)\}/\mathbb{E}X_{m_\alpha}$, and the latter is no greater than $\mathbb{E}\{(X_{m_\beta}+X_{m_\gamma})X_{m_\alpha}\}/\mathbb{E}X_{m_\alpha}$. ∎

Remark 6·5. Because of Lemmas 6·3 and 6·4, the total variation distance between the distributions of the residual factorization (a) given $M_1 = m_\alpha$, and (b) given $M_1 = m_\beta$ and $M_2 = m_\gamma$, is of order η.

LEMMA 6·6.

$$\mathbb{P}[\delta(M_2) = u \,|\, M_1 = m_\beta] = (n-j)^{-1}(1+O(q^{-j}+q^{-u/2}));$$
$$\mathbb{P}[\delta(M_1) = v] = n^{-1}(1+O(q^{-v/2})).$$

Proof. Direct computation yields

$$\mathbb{P}[M_2 = m \,|\, M_1 = m_\beta] = \frac{u\mathbb{E}(X_m X_{m_\beta})}{(n-j)\,\mathbb{E}X_{m_\beta}} = \frac{uq^{-u}}{n-j}(1+O(q^{-j}+q^{-u}))$$

if $\delta(m) = u$, and that

$$\mathbb{P}[M_1 = m'] = vn^{-1}\,\mathbb{E}X_{m'} = n^{-1}vq^{-v}(1+O(q^{-v}))$$

if $\delta(m') = v$. Adding over the possible choices of m and m', and using (2·6), concludes the proof. ∎

Remark 6·7. As a result of the three lemmas, it is possible to realize degree processes $(\delta(M_i'))_{i\geqslant 1}$ with the unconditional distribution and $(\delta(M_i''))_{i\geqslant 1}$ with the distribution conditional on $M_1 = m_\beta$, in such a way that $\delta(M_i') = \delta(M_{i+1}'')$ for all $i \geqslant 2$ holds, except on a set of probability of order at most

$$O(n^{-1}j + (n-j)^{-1} + q^{-j}). \tag{6·9}$$

THEOREM 6·8. *The distribution of the total number* K_0' *of factors satisfies*

$$d_{TV}(\mathscr{L}(K_0'), \mathrm{Po}\,(\psi(n+1))) \leqslant c(\log n)^{-\frac{1}{2}}.$$

Proof. Take a random monic polynomial π of degree n over F_q, and split it into linear factors over a splitting field. For each irreducible (over the original field) factor

of π, assign a mark to just one of the linear factors which make it up. Label the linear factors $1, 2, \ldots, n$ at random. Set $U_i = 1$ if the ith factor carries a mark, and set $U_i = 0$ otherwise. Then $K_0' = \sum_{i=1}^n U_i$ is the number of irreducible factors in π, and

$$\mathbb{E}U_i = \mathbb{P}[U_i = 1] = \sum_{j=1}^{n} j^{-1}\,\mathbb{P}[\delta(M_1) = j] = n^{-1}(\psi(n+1) + O(1)), \tag{6·10}$$

because of Lemma 6·6. Set $\lambda = \mathbb{E}K_0'$. From Barbour, Holst and Janson [4], Remark 1·1·7, and by symmetry,

$$\begin{aligned} d_{TV}(\mathscr{L}(K_0'), \mathrm{Po}\,(\lambda)) &\leqslant 2\lambda^{\frac{1}{2}} d_{TV}(\mathscr{L}(K_0'+1), \mathscr{L}(K_0' \mid U_1 = 1)) \\ &\leqslant 2\lambda^{\frac{1}{2}} \sum_{j=1}^{n} \mathbb{P}[\delta(M_1) = j \mid U_1 = 1]\, d_{TV}(\mathscr{L}(K_0'+1), \mathscr{L}(K_0' \mid U_1 = 1, \delta(M_1) = j)) \\ &= O\Big((\log n)^{-\frac{1}{2}} \sum_{j=1}^{n} j^{-1} d_{TV}(\mathscr{L}(K_0'+1), \mathscr{L}(K_0' \mid U_1 = 1, \delta(M_1) = j))\Big). \end{aligned}$$

Using Remark 6·7 and (6·9), it thus follows that

$$d_{TV}(\mathscr{L}(K_0'), \mathrm{Po}\,(\lambda)) = O((\log n)^{-\frac{1}{2}}\{1 + n^{-1}\log n + 1\}) = O((\log n)^{-\frac{1}{2}}).$$

Finally, note that $|\lambda - \psi(n+1)| = O(1)$ as $n \to \infty$, in view of (2·6). ∎

Note that, because of the precision of the coupling, the argument could be used to prove Poisson approximation for other quantities, such as the number of factors of even degree.

This work was supported in part by Schweiz. Nationalfonds Grants Nos 21-25579.88 and 20-31262.91, and by NSF Grant DMS 90-05833.

REFERENCES

[1] R. A. Arratia and S. Tavaré. Independent process approximations for random combinatorial structures. *Advances in Mathematics* (1993). (In the press.)

[2] R. A. Arratia, A. D. Barbour and S. Tavaré. Poisson process approximations for the Ewens Sampling Formula. *Ann. Appl. Prob.* **2** (1992), 519–535.

[3] A. D. Barbour. Comment on a paper of Arratia, Goldstein and Gordon. *Statistical Science* (1990), 425–427.

[4] A. D. Barbour, L. Holst and S. Janson. *Poisson approximation.* Oxford University Press, 1992.

[5] M. Car. Factorization dans $\mathbb{F}_q(X)$. *C.R. Acad. Sci. Paris Ser. I.* **294** (1982), 147–150.

[6] M. Car. Ensembles de polynômes irréductibles et théorèmes de densité. *Acta Arith.* **44** (1984), 323–342.

[7] M. Csörgő and P. Révész. *Strong approximation in probability and statistics.* Academic Press, 1981.

[8] J. M. de Laurentis and B. G. Pittel. Random permutations and Brownian motion. *Pacific J. Math.* **119** (1985), 287–301.

[9] P. Diaconis and J. W. Pitman. Unpublished lecture notes. Statistics Department, University of California, Berkeley (1986).

[10] P. Diaconis, M. McGrath and J. W. Pitman. Cycles and descents of random permutations. Preprint (1992).

[11] P. Donnelly and P. Joyce. Continuity and weak convergence of ranked and size-biassed permutations on an infinite simplex. *Stoch. Proc. Applns* **31** (1989), 89–103.

[12] W. Feller. The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.* **51** (1945), 800–832.

[13] W. Feller. *An introduction to probability theory and its applications. Vol I.* Wiley, New York, 1950.

[14] P. FLAJOLET and M. SORIA. Gaussian limiting distributions for the number of components in combinatorial structures. *J. Comb. Th. A* **53** (1990), 165–182.

[15] I. M. GESSEL and C. REUTENAUER. Counting permutations with given cycle structure and descent set. Preprint (1991).

[16] J. HANSEN. Order statistics for decomposable combinatorial structures. Preprint (1991).

[17] V. F. KOLCHIN. *Random mappings*. Optimization Software, Inc., New York, 1986.

[18] J. KOMLÓS, P. MAJOR and G. TUSNÁDY. An approximation of partial sums of independent RV-s, and the sample DF. I, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **32** (1975), 111–131.

[19] T. G. KURTZ. Strong approximation theorems for density dependent Markov chains. *Stoch. Proc. Applns* **6** (1978), 223–240.

[20] R. LIDL and H. NIEDERREITER. *Introduction to finite fields and their applications*. Cambridge University Press, 1986.

[21] N. METROPOLIS and G.-C. ROTA. Witt vectors and the algebra of necklaces. *Adv. Math.* **50** (1983), 95–125.

[22] N. METROPOLIS and G.-C. ROTA. The cyclotomic identity. In *Contemporary Mathematics 34*, (American Mathematical Society, 1984), pp. 19–27.

[23] S. T. RACHEV. *Probability metrics and the stability of stochastic models*. Wiley, 1991.

[24] A. RÉNYI. Théorie des elements saillants d'une suite d'observations. *Coll. Comb. Meth. Prob. Th.* Mathematisk Institut, Aarhus Universitet (1962), 104–115.

[25] L. A. SHEPP and S. P. LLOYD. Ordered cycle lengths in a random permutation. *Trans. Amer. Math. Soc.* **121** (1966), 340–357.

[26] A. M. VERSHIK and A. A. SHMIDT. Limit measures arising in the theory of groups I. *Theor. Prob. Applns* **22** (1977), 79–85.