

A Poisson approximation for the Dirichlet law, the Ewens sampling formula and the Griffiths–Engen–McCloskey law by the Stein–Chen coupling method

ULRICH MARTIN HIRTH

Mathematisch-Geographische Fakultät, Katholische Universität Eichstätt, Ostenstraße 26-28, D-85071 Eichstätt, Germany. e-mail: ulrich.hirth@ku-eichstaett.de

We consider the random number of (Griffiths–Engen–McCloskey (GEM))–(Poisson–Dirichlet) components which are greater than ϵ . In two alternative and similar ways, letting Dirichlet laws and Ewens sampling formula laws respectively converge to the GEM–(Poisson–Dirichlet) law and using the Stein–Chen coupling method, we prove the Poisson approximation with respect to the total variation metric of the satisfactory order of magnitude $1/\text{expectation}$.

Keywords: coupling; Dirichlet distribution; Ewens sampling formula; GEM distribution; Poisson approximation; Poisson–Dirichlet distribution; Stein–Chen method; total variation metric

1. Introduction

The *Griffiths–Engen–McCloskey (GEM) law with parameter $\theta > 0$* is the distribution of a random sequence $X = (X_1, X_2, X_3, \dots)$ with

$$X_1 := Z_1, X_2 := (1 - Z_1)Z_2, \quad X_k := Z_k \prod_{j=1}^{k-1} (1 - Z_j) \quad (1.1)$$

and where Z_1, Z_2, Z_3, \dots are independent beta $(1, \theta)$ random variables.

The *Poisson–Dirichlet law with parameter $\theta > 0$* is obtained from the GEM law with the same parameter θ by applying the non-increasing order statistic to X . The GEM–(Poisson–Dirichlet) laws have a long history (see the references).

We are interested in the random number

$$M_\epsilon := \sum_{k=1}^{\infty} 1(X_k > \epsilon) \quad (1.2)$$

of components which are greater than ϵ ($0 < \epsilon < 1$). This number of components is the same under either the GEM or the Poisson–Dirichlet measure. It was conjectured by Hirth (1997,

Section 4) that M_ϵ is approximately Poisson distributed, asymptotically for $\epsilon \searrow 0$, where only the case $\theta = 1$ was focused on.

Here we provide a statement and two alternative similar proofs of Poisson approximation for all M_ϵ , for $\theta > 0$. The result is as follows.

Theorem.

$$\begin{aligned} d_{TV}(P \circ M_\epsilon^{-1}, Po(\lambda)) &\leq \lambda^{-1}(1 - e^{-\lambda})\{E[M_\epsilon] - \text{var}(M_\epsilon)\} \\ &\leq K(\theta)\lambda^{-1}(1 - e^{-\lambda}) \\ &\leq K(\theta) \min(1, \lambda^{-1}), \end{aligned}$$

where d_{TV} denotes the total variation metric, $Po(\lambda)$ the Poisson law with parameter λ , here $\lambda := E[M_\epsilon]$, and for the special case $\theta = 1$ the constant $K(1)$ can be chosen explicitly as

$$(\log 2)^2 + \sum_{j=1}^{\infty} 2^{-j+1} j^{-2}. \tag{1.3}$$

Here the total variation distance between probability measures μ and ν on \mathbb{N}_0 is given by

$$\begin{aligned} d_{TV}(\mu, \nu) &:= \sup \{|\mu(A) - \nu(A)| : A \subseteq \mathbb{N}_0\} \\ &= \frac{1}{2} \sum_{j \in \mathbb{N}_0} |\mu\{j\} - \nu\{j\}| \end{aligned}$$

as defined and used by Barbour *et al.* (1992, p. 1, (1.3)).

According to Barbour *et al.* (1992) we should not expect that a total variation estimate with order of magnitude better than $1/E[M_\epsilon]$ is true, because $1/\lambda$ is typically the true order of magnitude in the case of ‘negative dependence’. For the two proofs of our theorem we shall let Dirichlet laws and Ewens sampling formula (ESF) laws respectively converge weakly to the GEM–(Poisson–Dirichlet) law, thereby using the invariance of our counting random variable M_ϵ under permutations of the components. More precisely the symmetric Dirichlet laws with all the $d(d \in \mathbb{N} \setminus \{1\})$ parameters equal to $a > 0$ and components ordered by size non-increasingly converge to the Poisson–Dirichlet law with parameter θ as $d \nearrow \infty$, $a \searrow 0$ and $d \cdot a \rightarrow \theta$; we shall have $d \cdot a = \theta$ (Kingman 1975, 1993, Chapter 9, Section 9, p. 93).

The ESF law on the other hand is the probability measure on the set \mathcal{S}_n of all permutations of the set $\{1, \dots, n\}$, $n \in \mathbb{N} := \{1, 2, 3, \dots\}$, whose density with respect to the uniform probability distribution is proportional to θ^k ($\theta > 0$ being the same parameter θ as above), where k is the number of cycles in the permutation. Under the ESF (θ) law, the joint distribution of the normalized (i.e., divided by n) cycle lengths, ordered by size non-increasingly, converges weakly to the Poisson–Dirichlet law (Kingman 1977; Ewens 1990; Hoppe 1987). The ESF analogue to M_ϵ is the random number $K_{b,n}$ of permutation cycles whose length is larger than b , with $b = [n \cdot \epsilon]$ (integer floor of $n \cdot \epsilon$), $b \in \{0, \dots, n - 1\}$.

Each one of the two proofs of our theorem consists of three parts: Stein–Chen coupling, weak-convergence argument and moment calculations. The second and third parts are common to both proofs whereas each of the proofs uses its own Stein–Chen coupling. As a first hint towards the Poisson approximation we present the moment calculations for M_ϵ under the GEM–(Poisson–Dirichlet) law in Section 2, followed by the weak convergence argument in Section 3, the Stein–Chen coupling for symmetric Dirichlet laws in Section 4 and the Stein–Chen coupling for the ESF laws in Section 5.

2. $E[M_\epsilon]$ and $\text{var}(M_\epsilon)$ under the (Griffiths–Engen–McCloskey)–(Poisson–Dirichlet) law

For the special case $\theta = 1$, Hirth (1997, Section 4) calculated

$$E[M_\epsilon] = \log\left(\frac{1}{\epsilon}\right) \tag{2.1}$$

for all $\epsilon \in (0, 1)$ and

$$E[M_\epsilon] - \text{var}(M_\epsilon) = -c_2\left(\frac{1}{\epsilon}\right), \tag{2.2}$$

where $c_2: [2, \infty) \rightarrow [-(\log 2)^2, c_2(\infty))$ is a negative decreasing bounded function of $t := 1/\epsilon$ with $|c_2(2)| = (\log 2)^2$ and

$$|c_2(\infty)| = (\log 2)^2 + \sum_{j=1}^{\infty} 2^{-j+1} j^{-2} \tag{2.3}$$

(compare with (1.3) in the statement of our theorem), which is obviously finite. Note that the result (2.2) is only for $t \geq 2$, i.e., for $\epsilon \leq \frac{1}{2}$, but this restriction is no problem for us.

Here we treat expectation and variance of M_ϵ for general $\theta > 0$. As in Hirth (1997, Section 4) we easily obtain the equality in distribution

$$M_{1/t} \approx 1\left(B > \frac{1}{t}\right) + M_{1/\{(1-B)t\}},$$

where B is an independent beta $(1, \theta)$ random variable and the second summand is zero for $(1 - B)t \leq 1$. Defining $f(t) := E[M_{1/t}]$, we have $f(t) = 0$ for $t \leq 1$ and, for $t \geq 1$,

$$\begin{aligned} f(t) &:= E[M_{1/t}] = P\left[B > \frac{1}{t}\right] + E[M_{1/\{(1-B)t\}}] \\ &= \theta \int_{1/t}^1 (1-x)^{\theta-1} dx + \theta \int_0^1 x^{\theta-1} f(tx) dx \\ &= \left(1 - \frac{1}{t}\right)^\theta + \theta t^{-\theta} \int_1^t x^{\theta-1} f(x) dx, \end{aligned}$$

since $f(t) = 0$ for $t \leq 1$.

Differentiating $t^\theta f(t)$ yields, for $t > 1$, $t^\theta f'(t) = \theta(t-1)^{\theta-1}$. So we have

$$\begin{aligned} \mathbb{E}[M_\epsilon] &= f\left(\frac{1}{\epsilon}\right) \\ &= \int_1^{1/\epsilon} \frac{\theta}{x} \left(1 - \frac{1}{x}\right)^{\theta-1} dx \\ &= \int_\epsilon^1 \frac{\theta}{y} (1-y)^{\theta-1} dy. \end{aligned} \tag{2.4}$$

From (2.4) we easily see that $\mathbb{E}[M_\epsilon]$ tends to infinity as $\epsilon \searrow 0$, for any fixed $\theta > 0$. Next we look at the second moment of M_ϵ ; define and calculate

$$\begin{aligned} g(t) &:= \mathbb{E}[M_{1/t}^2] \\ &= P\left[B > \frac{1}{t}\right] + 2\mathbb{E}\left[M_{1/\{(1-B)t\}} 1\left(B > \frac{1}{t}\right)\right] + \mathbb{E}[M_{1/\{(1-B)t\}}^2] \\ &= \left(1 - \frac{1}{t}\right)^\theta + 2\theta \int_{1/t}^1 (1-x)^{\theta-1} f((1-x)t) dx \\ &\quad + \theta \int_0^1 (1-x)^{\theta-1} g((1-x)t) dx \\ &= \left(1 - \frac{1}{t}\right)^\theta + 2\theta t^{-\theta} \int_1^{t-1} y^{\theta-1} f(y) dy + \theta t^{-\theta} \int_1^t y^{\theta-1} g(y) dy. \end{aligned}$$

Differentiating $t^\theta g(t)$ yields, for $t > 2$, $t^\theta g'(t) = \theta(t-1)^{\theta-1}\{1 + 2f(t-1)\}$. So we have

$$\begin{aligned} g(t) &= f(t) + 2\theta \int_2^t x^{-\theta} (x-1)^{\theta-1} f(x-1) dx \\ &= f(t) + 2\theta^2 \int_2^t dx \int_1^{x-1} dy (xy)^{-\theta} (x-1)^{\theta-1} (y-1)^{\theta-1} \\ &=: f(t) + 2\theta^2 \int_G h, \end{aligned}$$

where $G := \{(x, y): 1 < y \leq x-1 \leq t-1\}$ and $h(x, y) := (xy)^{-\theta} (x-1)^{\theta-1} (y-1)^{\theta-1}$.

Note the symmetry $h(x, y) = h(y, x)$. For comparison we write

$$\begin{aligned}
 f(t)^2 &= \theta^2 \int_1^t dx \int_1^t dy (xy)^{-\theta} (x-1)^{\theta-1} (y-1)^{\theta-1} \\
 &= 2\theta^2 \int_1^t dx \int_1^x dy h(x, y) \\
 &=: 2\theta^2 \int_F h,
 \end{aligned}$$

where

$$F := \{(x, y): 1 < y \leq x \leq t\}.$$

We have $G \subset F$ and

$$\begin{aligned}
 E[M_{1/t}] - \text{var}(M_{1/t}) &= f(t) - g(t) + f(t)^2 \\
 &= 2\theta^2 \int_{F \setminus G} h,
 \end{aligned}$$

where, for $t \geq 2$, $F \setminus G = H \cup J$ with

$$H := \{(x, y): 1 < y \leq x \leq 2\}, \quad J := \{(x, y): 1 < x - 1 < y \leq x \leq t\}.$$

For $(x, y) \in J$ we estimate that

$$h(x, y) \leq \begin{cases} x^{-\theta} (x-1)^{-1} (x-2)^{\theta-1}, & 0 < \theta \leq 1, \\ (x-1)^{-2}, & \theta > 1, \end{cases}$$

so $\int_J h$ is bounded uniformly in t .

3. The weak-convergence argument

Let $(\mu_n)_{n \in \mathbb{N}}$ be a sequence of probability measures which converges weakly to the GEM–(Poisson–Dirichlet) law μ . Think of the measures μ_n as the Dirichlet laws or as the normalized ESF laws as mentioned in Section 1 and, as explained there, we do not pay attention to the ordering of the components of μ_n and μ because the counting random variable M_ϵ (or $K_{b,n}$ respectively with $b = [n \cdot \epsilon]$) we are interested in is invariant under permutations of the components.

It was proved by Hirth (1997, Section 2, Lemmas 3–5 plus corollary) that the distributions $\mu_n \circ M_\epsilon^{-1}$ converge weakly to $\mu \circ M_\epsilon^{-1}$ also. The same is true with $K_{b,n}$ ($b = [n \cdot \epsilon]$) replacing M_ϵ , and from here on we unify these two random elements notationally by writing M_ϵ only.

Since we have always $M_\epsilon < 1/\epsilon$ (at least almost surely, with respect to μ_n (for all $n \in \mathbb{N}$) and μ), we conclude immediately that the expectations $\int M_\epsilon d\mu_n$ converge to the expectation $\int M_\epsilon d\mu$ as $n \rightarrow \infty$. For the same reason the variances of M_ϵ with respect to

μ_n converge to the variance of M_ϵ with respect to μ . Furthermore the convergence $\int M_\epsilon d\mu_n \rightarrow \int M_\epsilon d\mu$ implies elementarily the weak convergence of the Poisson laws with parameters $\int M_\epsilon d\mu_n$ to the Poisson law with parameter $\int M_\epsilon d\mu$.

We recall that the total variation distance d_{TV} metrizes weak convergence of probability measures on \mathbb{N}_0 . All these arguments and the triangle inequality for the total variation metric d_{TV} show that the inequality

$$d_{TV}(\mu_n \circ M_\epsilon^{-1}, Po(\lambda_n)) \leq \lambda_n^{-1}(1 - e^{-\lambda_n})\{\lambda_n - \text{var}_{\mu_n}(M_\epsilon)\} \tag{3.1}$$

with $\lambda_n := \int M_\epsilon d\mu_n$ implies the inequality

$$d_{TV}(\mu \circ M_\epsilon^{-1}, Po(\lambda)) \leq \lambda^{-1}(1 - e^{-\lambda})\{\lambda - \text{var}_\mu(M_\epsilon)\} \tag{3.2}$$

by convergence as $n \rightarrow \infty$ and continuity. So it remains to show (3.1) to complete the proof of our theorem, and we are going to do this in two (Sections 4 and 5) alternative ways, both of which use the Stein–Chen coupling method in similar ways.

4. Stein–Chen coupling for Dirichlet laws

Let $Y = (Y_1, \dots, Y_d)$, $d \in \mathbb{N} \setminus \{1\}$, be a random vector with values in the $(d - 1)$ -dimensional simplex

$$\Delta_d := \left\{ (y_1, \dots, y_d) \in [0, 1]^d : \sum_{j=1}^d y_j = 1 \right\},$$

whose distribution is the symmetric Dirichlet law with all the d parameters equal to $a > 0$, $a := \theta/d$. We are interested in the distribution of the random variable

$$M_\epsilon := \sum_{j=1}^d 1(Y_j > \epsilon), \tag{4.1}$$

the number of components which are greater than ϵ .

Note that we have $Y_1 \neq 1$ almost surely; so we can define

$$Y' = (Y'_1, \dots, Y'_{d-1}) := \left(\frac{Y_2}{1 - Y_1}, \frac{Y_3}{1 - Y_1}, \frac{Y_4}{1 - Y_1}, \dots, \frac{Y_d}{1 - Y_1} \right),$$

and for $d \in \mathbb{N} \setminus \{1, 2\}$ this (Δ_{d-1}) -valued random vector Y' has the symmetric Dirichlet distribution with $d - 1$ parameters a , $a = \theta/d$ remaining the same as above. Moreover $\{Y_1, Y'\}$ is stochastically independent. With this notation we write

$$M_\epsilon = 1(Y_1 > \epsilon) + \sum_{j=1}^{d-1} 1\left(Y'_j > \frac{\epsilon}{1 - Y_1}\right).$$

The following claim is proved by an explicit construction which is used in the sequel.

Claim. The distribution of Y_1 conditional on the event $\{Y_1 > \epsilon\}$ is stochastically greater (\geq) than the unconditioned beta $(a, (d - 1)a)$ distribution of Y_1 , for each $\epsilon \in (0, 1)$.

Proof. Let $\{B, B', Y'\}$ be stochastically independent, B having the same beta $(a, (d - 1)a)$ distribution as Y_1 (unconditioned), B' having the $\{Y_1 > \epsilon\}$ -conditional distribution of Y_1 , and define

$$C := B 1(B > \epsilon) + B' 1(B \leq \epsilon).$$

We obviously have $C \geq B$ in the usual sense of ordering real numbers, and C is equal in distribution to B' . The claim is proved. \square

We now write

$$M_\epsilon = 1(B > \epsilon) + \sum_{j=1}^{d-1} 1\left(Y'_j > \frac{\epsilon}{1-B}\right)$$

and represent the conditional distribution of $M_\epsilon - 1$ given the event $\{B := Y_1 > \epsilon\}$ by

$$\sum_{j=1}^{d-1} 1\left(Y'_j > \frac{\epsilon}{1-C}\right),$$

for which we have

$$\sum_{j=1}^{d-1} 1\left(Y'_j > \frac{\epsilon}{1-C}\right) \leq \sum_{j=1}^{d-1} 1\left(Y'_j > \frac{\epsilon}{1-B}\right) \leq M_\epsilon. \tag{4.2}$$

We have just shown the existence of a *monotone coupling* as required by Barbour *et al.* (1992) for their equation (0.1) on p. 21 (beginning of Chapter 2). This inserted into their Theorem 1.B on p. 11 in Chapter 1, combined with their Lemma 1.1.1 on p. 7, leads to our (3.1) which implies (3.2). So our theorem is proved. \square

An alternative proof for our theorem is given next.

5. Stein–Chen coupling for the Ewens sampling formula

The ESF random permutation can be constructed by the Feller coupling as in Arratia *et al.* (1992, p. 523, Section 3). Using their notation, we have (empty sums being interpreted as zero and empty products as one as usual)

$$K_{b,n} = \sum_{l=1}^{n-b} \xi_l \underbrace{\prod_{j=l+1}^{l+b} (1 - \xi_j)}_{=: I_l}, \tag{5.1}$$

with $(\xi_m)_{m \in \mathbb{N}}$ being an independent sequence satisfying

$$P[\xi_m = 1] = 1 - P[\xi_m = 0] = \frac{\theta}{\theta + m - 1} \quad (5.2)$$

for each $m \in \mathbb{N}$.

So each indicator I_l , $l \in \{1, \dots, n - b\}$, is independent of $(I_j: j \in \{1, \dots, n - b\} \text{ and } |l - j| > b)$, and the event $\{I_l = 1\}$ implies the event $\{I_j = 0 \text{ for all } j \in \{1, \dots, n - b\} \text{ with } 0 < |j - l| \leq b\}$.

Thus the distribution of $K_{b,n} - 1$ conditional on the event $\{I_l = 1\}$ is represented by the partial sum $\sum_{\substack{j=1 \\ |j-l|>b}}^{n-b} I_j$ which is less than or equal to $K_{b,n}$.

The remainder of this proof for our theorem is word by word the same as the paragraph following (4.2). \square

Acknowledgements

The author expresses his sincere thanks to his new institution and to his promotors Professor N.H. Bingham (London) and Professor P. Ressel (Eichstätt), and also to his parents for his two years in London and to Royal Holloway and Bedford New College for financial support and the Department for its supportive research environment, where his tuition fees were covered by Science and Engineering Research Council—Engineering and Physical Sciences Research Council ‘EC-Fees Only’ Award 93001492. Thanks are also due to Professor Andrew D. Barbour, Dr Mihael Perman and the referees for helpful comments.

References

- Arratia, R., Barbour, A.D. and Tavaré, S. (1992) Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Probab.*, **2**, 519–535.
- Barbour, A.D., Holst, L. and Janson, S. (1992) *Poisson Approximation*. Oxford: Clarendon.
- Ewens, W.J. (1990) Population genetics theory—the past and the future. In S. Lessard (ed.), *Mathematical and Statistical Developments of Evolutionary Theory*, pp. 177–227. Dordrecht: Kluwer.
- Hirth, U.M. (1997) Probabilistic number theory, the GEM/Poisson–Dirichlet distribution and the arc–sine law. *Combinatorics, Probability Computing*, **6**, 57–77.
- Hoppe, F.M. (1987) The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.*, **25**, 123–159.
- Kingman, J.F.C. (1975) Random discrete distributions. *J. Roy. Statist. Soc., Ser. B*, **37**, 1–22.
- Kingman, J.F.C. (1977) The population structure associated with the Ewens sampling formula. *Theor. Population Biol.*, **11**, 274–283.
- Kingman, J.F.C. (1993) *Poisson Processes*. Oxford: Clarendon.

Received February 1996 and revised August 1996