

# Reparametrizing the two-parameter Gnedin-Fisher partition model in a Bayesian perspective (Reparamétrisation du modèle de partition Gnedin-Fisher à deux paramètres dans une perspective Bayésienne)

Cerquetti, Annalisa

Sapienza University of Rome, Department MeMoTEF

Via del Castro Laurenziano, 9

Rome (00161), Italy

E-mail: annalisa.cerquetti@gmail.com

## 1. Introduction

Gnedin (2010) introduces a two parameter family of exchangeable partition probability functions (EPPF) belonging to the Gibbs class of genus  $\alpha = -1$  (Gnedin and Pitman, 2006), by suitably mixing Poisson-Dirichlet  $(-1, \xi)$  models (Pitman and Yor, 1997; Fisher *et al.*, 1943), for  $\xi = 1, 2, 3, \dots$ , over  $\xi$  with

$$(1) \quad \mathbb{P}_{\gamma, \zeta}(\Xi = \xi) = \frac{\Gamma(z_1 + 1)\Gamma(z_2 + 1) \prod_{i=1}^{\xi-1} (i^2 - \gamma i + \zeta)}{\Gamma(\gamma) \xi!(\xi - 1)!},$$

for some complex  $z_1$  and  $z_2$ , for  $\gamma \geq 0$  and (i) either  $i^2 - \gamma i + \zeta$  (strictly) positive for all  $i \in \mathbb{N}$  or (ii) the quadratic is positive for  $i \in \{1, \dots, i_0 - 1\}$  and has a root at  $i_0$ . The resulting  $(\gamma, \zeta)$  Gnedin-Fisher species sampling model has EPPF of the Gibbs type  $p(n_1, \dots, n_k) = V_{n,k} \prod_{j=1}^k (1 - \alpha)_{n_j-1}$  in the specific form

$$(2) \quad p_{\gamma, \zeta}(n_1, \dots, n_k) = \frac{(\gamma)_{n-k} \prod_{i=1}^{k-1} (i^2 - \gamma i + \zeta)}{\prod_{l=1}^{n-1} (l^2 + \gamma l + \zeta)} \prod_{j=1}^k n_j!.$$

The fundamental result in Gnedin and Pitman (2006, cfr. Th. 12) establishes that the EPPF of each Gibbs partition of genus  $\alpha \in (-\infty, 1)$  corresponds to a mixture of *extreme* partition probability functions, which differ for  $\alpha \in (-\infty, 0)$ ,  $\alpha = 0$  and  $\alpha \in (0, 1)$ . For  $\zeta = 0$  then  $\gamma \in (0, 1)$  (cfr. Gnedin, 2010 Sect. 6),(2) reduces to

$$(3) \quad p_{\gamma}(n_1, \dots, n_k) = \frac{(k-1)! (1-\gamma)_{k-1} (\gamma)_{n-k}}{(n-1)! (1+\gamma)_{n-1}} \prod_{j=1}^k n_j!,$$

and the law of the number of occupied blocks  $K_n$  is obtained by the general formula for Gibbs partitions,  $\mathbb{P}(K_n = k) = V_{n,k} S_{n,k}^{-1, -\alpha}$ , for  $S_{n,k}^{-1, -\alpha}$  generalized Stirling numbers. For  $\alpha = -1$  those reduce to Lah numbers  $S_{n,k}^{-1, 1} = \binom{n-1}{k-1} \frac{n!}{k!}$  hence

$$(4) \quad \mathbb{P}_{\gamma}(K_n = k) = \binom{n}{k} \frac{(1-\gamma)_{k-1} (\gamma)_{n-k}}{(1+\gamma)_{n-1}}.$$

As from Gnedin (2010, cfr. eq. (9) and (10)), the mixing law yielding the one-parameter model (3) arises from (4) for  $n \rightarrow \infty$  by the standard asymptotics  $\Gamma(n + \alpha)/\Gamma(n + b) \sim n^{a-b}$  and corresponds to

$$(5) \quad \mathbb{P}_{\gamma}(\Xi = \xi) = \frac{\gamma(1-\gamma)_{\xi-1}}{\xi!},$$

for  $\xi = 1, 2, \dots$ , and  $\gamma \in (0, 1)$ , while, thinking of (5) as a *prior*, a *posterior* distribution, for  $1 \leq k \leq n$ , for  $\Xi$  results

$$(6) \quad \mathbb{P}_{\gamma}(\Xi = \xi | K_n = k) = \frac{(n-1)! \Gamma(\gamma + n)}{(k-1)! \Gamma(\gamma + n - k)} \frac{(k-\gamma)_{\xi} \Gamma(k + \xi)}{\Gamma(\xi + 1) \Gamma(k + \xi + n)}.$$

## 2. Main results

Gnedin (2010, cfr. Sect. 2) points out that the Gibbs weights in (2) can be split in linear factors by factoring the quadratics as

$$x^2 + \gamma x + \zeta = (x + z_1)(x + z_2), \text{ and } x^2 - \gamma x + \zeta = (x + s_1)(x + s_2),$$

thus providing the alternative *five* parameters representation

$$(7) \quad V_{n,k}^{\gamma,\zeta} = \frac{(\gamma)_{n-k}(s_1 + 1)_{k-1}(s_2 + 1)_{k-1}}{(z_1 + 1)_{n-1}(z_2 + 1)_{n-1}}$$

for some complex  $z_1, z_2, s_1, s_2$ , such that  $z_1 + z_2 = \gamma$ ,  $z_1 z_2 = \zeta$ ,  $s_1 + s_2 = -\gamma$ ,  $s_1 s_2 = \zeta$ . In the next theorem we show those constraints limit admissible values for the four parameters  $s_1, s_2, z_1$  and  $z_2$  yielding an interesting alternative *two* parameters representation of the model.

**Theorem 1.** For  $\psi \in \mathbb{C}$  with  $Re(\psi) = \gamma/2$ , for  $0 < \gamma < 2$ , and  $\gamma - \psi$  its complex conjugate, the EPPF of the two-parameter  $(\gamma, \zeta)$ -Gnedin-Fisher species sampling model (2) admits the following alternative representation

$$(8) \quad p_{\gamma,\psi}(n_1, \dots, n_k) = \frac{(\gamma)_{n-k}(1 - \psi)_{k-1}(1 - \gamma + \psi)_{k-1}}{(1 + \psi)_{n-1}(1 + \gamma - \psi)_{n-1}} \prod_{j=1}^k n_j!$$

For  $\psi \in (-1, 1)$  then  $0 < \gamma < \psi + 1$  and the model has real parameters. For  $\psi = 0$ , then  $\gamma \in (0, 1)$  and (8) yields the one-parameter model (3).

*Proof:* For  $z_1 + z_2 = \gamma$ ,  $z_1 z_2 = \zeta$ ,  $s_1 + s_2 = -\gamma$  and  $s_1 s_2 = \zeta$  vectors  $(z_1, z_2)$  and  $(s_1, s_2)$  must be the roots (complex or real) of the following quadratic polynomials

$$z_1^2 - \gamma z_1 + \zeta = 0 \quad \text{and} \quad z_2^2 - \gamma z_2 + \zeta = 0, \quad s_1^2 + \gamma s_1 + \zeta = 0 \quad \text{and} \quad s_2^2 + \gamma s_2 + \zeta = 0.$$

For  $\gamma^2 - 4\zeta > 0$  admissible *real* solutions are

$$z_1 = \frac{\gamma \pm \sqrt{\gamma^2 - 4\zeta}}{2} \quad \text{and} \quad z_2 = \frac{\gamma \pm \sqrt{\gamma^2 - 4\zeta}}{2}, \quad s_1 = \frac{-\gamma \pm \sqrt{\gamma^2 - 4\zeta}}{2} \quad \text{and} \quad s_2 = \frac{-\gamma \pm \sqrt{\gamma^2 - 4\zeta}}{2}.$$

For  $\gamma^2 - 4\zeta < 0$  admissible *complex* solutions are

$$z_1 = \frac{\gamma \pm i\sqrt{4\zeta - \gamma^2}}{2} \quad \text{and} \quad z_2 = \frac{\gamma \pm i\sqrt{4\zeta - \gamma^2}}{2}, \quad s_1 = \frac{-\gamma \pm i\sqrt{4\zeta - \gamma^2}}{2} \quad \text{and} \quad s_2 = \frac{-\gamma \pm i\sqrt{4\zeta - \gamma^2}}{2}.$$

Now let indifferently  $A = i\sqrt{4\zeta - \gamma^2}/2$  or  $A = \sqrt{\gamma^2 - 4\zeta}/2$ . Then, regardless of the solutions being real or complex, possible vectors satisfying the constraints  $z_1 + z_2 = \gamma$ ,  $s_1 + s_2 = -\gamma$ ,  $z_1 z_2 = \zeta$  and  $s_1 s_2 = \zeta$  must be as follows

$$(z_1, z_2) = \left(\frac{\gamma}{2} + A, \frac{\gamma}{2} - A\right) \text{ or } \left(\frac{\gamma}{2} - A, \frac{\gamma}{2} + A\right)$$

and

$$(s_1, s_2) = \left(-\frac{\gamma}{2} + A, -\frac{\gamma}{2} - A\right) \text{ or } \left(-\frac{\gamma}{2} - A, -\frac{\gamma}{2} + A\right),$$

which shows admissible solutions reduce to  $z_1 = -s_1$  and  $z_2 = -s_2$  or  $z_1 = -s_2$  and  $z_2 = -s_1$ . Since (7) is invariant to permutations of  $(z_1, z_2)$  and  $(s_1, s_2)$ , the five parameters in (7) reduce to  $\psi$  and  $\gamma$  for  $z_1 = \psi$  and  $z_2 = \gamma - \psi$ , and  $s_1 = -\psi$ ,  $s_2 = \psi - \gamma$  thus yielding (8). For  $\psi \in \mathbb{C}$  then the existence and the positiveness of the complex Gamma function implies  $Re(1 - \psi) > 0$  hence  $0 < \gamma < 2$ . For  $\psi \in \mathbb{R}$

then positiveness of the numerator in (8) implies  $1 - \psi > 0$  and  $1 - \gamma + \psi > 0$ . For  $\psi = 0$ ,  $0 < \gamma < 1$  and (8) reduces to (3) by standard combinatorial calculus.  $\square$

**Remark 2.** Rewriting the *prior* (5) and the *posterior* (6) of the one parameter model as

$$\mathbb{P}_\gamma(\Xi = \xi) = \frac{(1)_{\xi-1} (1-\gamma)_{\xi-1} (\gamma)_1}{\Gamma(\xi) (1)_\xi} \quad \text{and} \quad \mathbb{P}_\gamma(\Xi = \xi | K_n = k) = \frac{(k)_{\xi-1} (k-\gamma)_{\xi-1} (n+\gamma-k)_k}{\Gamma(\xi) (n)_{k+\xi-1}},$$

for  $\xi = 1, 2, \dots$ , one may notice both belong to the class of *shifted univariate generalized Waring distributions*, (also known as inverse Markov-Polya), a family of distributions on  $\mathbb{N} \cup 0$  (Irwin, 1975; Xekalaki, 1983), with probability mass function

$$(9) \quad \mathbb{P}(N = i) = \frac{(\rho)_\eta (a)_i (\eta)_i}{i! (a+\rho)_{\eta+i}},$$

for  $i = 0, 1, 2, \dots$ , and  $a, \eta, \rho$  positive reals, which arise as *Beta*( $\rho, a$ ) mixtures of a Negative Binomial ( $\eta, p$ ) distributions. It is easy to verify that the prior and the posterior arise respectively as *Beta*( $\gamma, 1-\gamma$ ) mixtures of *NB*( $1, p$ ) and as *Beta*( $n+\gamma-k, k-\gamma$ ) mixtures of *NB*( $k, p$ ). The following result identifies the mixing distribution yielding the reparametrized ( $\gamma, \psi$ ) Gnedin-Fisher model with the *shifted Type II Gaussian hypergeometric distribution* ( $\delta, \alpha, \beta, \lambda$ ) (cfr. Rodriguez et al. 2007, cfr. Table 1), a family of discrete distributions generalizing Waring distributions to complex parameters.

**Theorem 3.** *The EPPF in (8) arises by mixing the family of PD(-1,  $\xi$ ) partition models*

$$(10) \quad p_{\xi,-1}(n_1, \dots, n_k) = \frac{(\xi-1)_{k-1} \uparrow_{-1}}{(\xi+1)_{n-1}} \prod_{j=1}^k n_j!,$$

over  $\xi$  with

$$(11) \quad \mathbb{P}_{\gamma,\psi}(\Xi = \xi) = \frac{(1-\psi)_{\xi-1} (1-\gamma+\psi)_{\xi-1} (\gamma)_{1-\psi}}{\Gamma(\xi) (1+\psi)_{\xi-\psi}},$$

a *shifted Type II Gaussian hypergeometric distribution* with parameters  $\lambda = 1$ ,  $\delta = 2$ ,  $\alpha = \psi$  and  $\beta = \gamma - \psi$ . For  $\psi \in (-1, 1)$  then  $\gamma \in (0, \psi + 1)$  and the mixing law reduces to a *shifted generalized Waring distribution* with real parameters  $a = 1 - \gamma + \psi$ ,  $\eta = 1 - \psi$  and  $\rho = \gamma$ .

*Proof:*

$$p_{\xi,-1}(n_1, \dots, n_k) = \sum_{\xi \geq k} \frac{(\xi-1)!}{(\xi-k)! (\xi+1)_{n-1}} \frac{(1-\psi)_{\xi-1} (1-\gamma+\psi)_{\xi-1} (\gamma)_{1-\psi}}{\Gamma(\xi) (1+\psi)_{\xi-\psi}}$$

which can be rewritten as

$$= (\gamma)_{1-\psi} \sum_{y=0}^{\infty} \frac{\Gamma(y+k+1) (1-\psi)_{y+k-1} (1-\gamma+\psi)_{1-\psi}}{y! \Gamma(y+k+n) (1+\psi)_{y+k-\psi}} =$$

then as

$$= (\gamma)_{1-\psi} \sum_{y=0}^{\infty} \frac{\Gamma(k-\psi+y) \Gamma(1+\psi) \Gamma(k-\gamma+\psi+y)}{y! \Gamma(y+k+n) \Gamma(1-\psi) \Gamma(1-\gamma+\psi)},$$

and by some manipulations we obtain

$$= \frac{(\gamma)_{1-\psi} \Gamma(1+\psi) \Gamma(k-\psi) \Gamma(k-\gamma+\psi) \Gamma(\gamma+n-k)}{\Gamma(1-\psi) \Gamma(1-\gamma+\psi) \Gamma(n+\gamma-\psi) \Gamma(n+\psi)} \sum_{y=0}^{\infty} \frac{(k-\psi)_y (k-\gamma+\psi)_y (n+\gamma-k)_{k-\psi}}{y! (n+\psi)_{k-\psi+y}}.$$

In the sum on the right we recognize (9) for parameters  $a = k - \gamma + \psi$ ,  $\eta = k - \psi$  and  $\rho = n + \gamma - k$  and this completes the proof.  $\square$

By exploiting the mixture representation introduced in Theorem 3. we are able to obtain the *structural distribution*, i.e. the law of the frequency of the first block observed,

$$\tilde{P}_1 := \lim_{n \rightarrow \infty} \frac{\#(B_1 \cap [n])}{n}.$$

**Proposition 4.** *The frequency  $\tilde{P}_1$  of block  $B_1$  of the  $(\gamma, \psi)$  Gnedin-Fisher model has distribution*

$\mathbb{P}_{\gamma, \psi}(\tilde{P}_1 \in dy) = (\gamma)_{1-\psi} \Gamma(1 + \psi) [\delta_1(dy) + (1 - \gamma + \psi)(1 - \psi)y {}_2F_1(2 - \psi, 2 - \gamma + \psi, 2; 1 - y)] dy$   
for  ${}_2F_1(a, b, c; x)$  the Gauss hypergeometric function.

*Proof:* By the mixture representation of Theorem 3.

$$\mathbb{P}_{\gamma, \psi}(\tilde{P}_1 \in dy) = \sum_{\xi=1}^{\infty} \mathbb{P}_{\gamma, \psi}(\Xi = \xi) \mathbb{P}(\tilde{P}_{\xi, 1} \in dy).$$

By the theory of the symmetric Dirichlet model, it is known that  $\tilde{P}_{\xi, i} \stackrel{d}{=} \text{Beta}(2, \xi - i)$ , therefore, since  $Be(2, 0) = \delta_1(dy)$ ,

$$\mathbb{P}_{\gamma, \psi}(\tilde{P}_1 \in dy) = (\gamma)_{1-\psi} \Gamma(1 + \psi) \delta_1(dy) + \sum_{\xi=2}^{\infty} \frac{(1 - \psi)_{\xi-1} (1 - \gamma + \psi)_{\xi-1} (\gamma)_{1-\psi} \Gamma(\xi + 1)}{\Gamma(\xi) (1 + \psi)_{\xi-\psi}} \frac{\Gamma(\xi + 1)}{\Gamma(\xi - 1)} y(1 - y)^{\xi-2}.$$

By the change of variable  $\xi - 2 = z$

$$\mathbb{P}_{\gamma, \psi}(\tilde{P}_1 \in dy) = (\gamma)_{1-\psi} \Gamma(1 + \psi) \delta_1(dy) + \sum_{z=0}^{\infty} \frac{(1 - \psi)_{z+1} (1 - \gamma + \psi)_{z+1} (\gamma)_{1-\psi} \Gamma(z + 3)}{\Gamma(z + 1) \Gamma(z + 2) (1 + \psi)_{z+2-\psi}} y(1 - y)^z =$$

and by standard combinatorial calculus

$$= (\gamma)_{1-\psi} \Gamma(1 + \psi) \left[ \delta_1(dy) + (1 - \psi)(1 - \gamma + \psi) \sum_{z=0}^{\infty} \frac{(2 - \psi)_z (2 - \gamma + \psi)_z}{\Gamma(z + 1) \Gamma(z + 2)} y(1 - y)^z \right],$$

and the result follows.  $\square$

The structural distribution may be useful to compute expected values of symmetric statistics of the frequencies of the kind  $\sum_{j=1}^k f(\tilde{P}_j)$ . It is an easy task for example to derive the following result

$$\mathbb{E} \left( \sum_{j=1}^k \tilde{P}_j^n \right) = \mathbb{E}(\tilde{P}_1^{n-1}) = \frac{n! (\gamma)_{n-1}}{(1 + \psi)_{n-1} (1 + \gamma - \psi)_{n-1}}.$$

### 3. Distributional results in a Bayesian perspective

In Lijoi *et al.* (2007, 2008) a Bayesian nonparametric treatment of species sampling problems has been proposed by means of conditional analysis of exchangeable Gibbs partitions. In this setting one assumes that sampling from an unknown population of an unknown number of different species provides an observation from an exchangeable partition model by the multiplicities of different species

observed in the first  $n$  observations. Interest usually lies in making inference on *species richness*, the number of different species in the population. In a Bayesian perspective this sums up to calculate a posterior distribution on the number of new species observed in an additional sample, conditional on having  $k$  different species in a basic sample already observed. Here we derive the posterior distribution for the Gnedin-Fisher two-parameter model relying on the alternative parametrization. The following result extends the one in Gnedin (2010, Eq. 10) for the one-parameter model and shows that (12) is still in the class of shifted Type II Gaussian hypergeometric distributions, for updated parameters  $a = k - \gamma + \psi$ ,  $\eta = k - \psi$  and  $\rho = n + \gamma - k$ , for  $\gamma < k < n + \gamma$ .

**Proposition 5.** *The posterior distribution for  $\Xi$  for the reparametrized  $(\gamma, \psi)$  Gnedin-Fisher model is given by*

$$(12) \quad \mathbb{P}_{\psi, \gamma}(\Xi = \xi | K_n = k) = \frac{(k - \psi)_{\xi-1} (k - \gamma + \psi)_{\xi-1} (n + \gamma - k)_{k-\psi}}{\Gamma(\xi) (n + \psi)_{k-\psi+\xi-1}}.$$

*Proof:* By the general form, for Gibbs models, of the posterior of the number of new species arising in a new sample of dimension  $m$  (cfr. Lijoi *et al.* 2007, eq. (4)), expressed in terms of non-central generalized Stirling numbers, we know that

$$(13) \quad \mathbb{P}(K_m^* = k^* | K_n = k) = \frac{V_{n+m, k+k^*}}{V_{n, k}} S_{m, k^*}^{-1, -\alpha, -(n-\alpha k)}.$$

Inserting the specific weights in (8), and exploiting the definition of non-central Lah numbers  $S_{n, k}^{-1, 1, r} = \frac{n!}{k!} \binom{n-r-1}{n-k}$ , (13) yields

$$(14) \quad \mathbb{P}_{\gamma, \psi}(K_m = k^* | K_n = k) = \binom{m}{k^*} (\gamma + n - k)_{m-k^*} (n + k + k^*)_{m-k^*} \frac{(k - \psi)_{k^*} (k - \gamma + \psi)_{k^*}}{(n + \psi)_m (n + \gamma - \psi)_m}.$$

For  $m \rightarrow \infty$ , by standard Stirling approximations, (12) arises.

**Remark 6.** Distribution (14) belongs to a family of discrete distributions arising by the following summation result for the  ${}_3F_2(1)$  generalized hypergeometric functions

$${}_3F_2(1) = \sum_{k=0}^n \frac{(-n)_k}{k!} \frac{(a)_k (b)_k}{(d)_k (1 + a + b - d - n)_k} = \frac{(d - a)_n (d - b)_n}{(d)_n (d - a - b)_n},$$

yielding after some manipulations

$$\sum_{k=0}^n \binom{n}{k} \frac{(-1)^k (a)_k (b)_k (d - a - b)_{n-k} (1 - n - d)_{n-k}}{(d - b)_n (1 - n - d + a)_n} = 1.$$

(14) arises for  $a = k - \psi$ ,  $b = k - \gamma + \psi$ ,  $d = n + k$  and corresponds to Type IV models in Table 1. of Rodríguez-Avi *et al.* (2008).

We conclude deriving a Bayesian estimator for the probability of discovering a new species at the  $(n + m + 1)$ th draw, given the basic sample  $(n_1, \dots, n_k)$ , under the two-parameter Gnedin-Fisher model by an application of Eq. (6) in Lijoi *et al.* (2007)

$$\tilde{D}_m^{n, k} = \sum_{k=0}^m \frac{(\gamma + n - k)_{m-k^*} (k - \psi)_{k^*+1} (k - \gamma + \psi)_{k^*+1}}{(\psi + n)_{m+1} (n + \gamma - \psi)_{m+1}} \binom{m}{k^*} \frac{m + n + k - 1!}{n + k + k^* - 1!}$$

which corresponds to

$$\tilde{D}_m^{n, k} = \frac{1}{(\psi + n + m)(n + \gamma - \psi + m)} \mathbb{E}[(K^* + k - \psi)(K^* + k - \gamma + \psi)]$$

where  $K^*$  has distribution (14).

Additionally a posterior distribution for the total number  $S_m$  of new observations belonging to new species is obtained according to eq. (11) in Lijoi *et al.* (2008) as follows

$$\mathbb{P}_{\gamma,\psi}(S_m = s | K_n = k) = \binom{m}{s} \frac{(n+k)_{m-s}}{(n+\psi)_m (n+\gamma-\psi)_m} \sum_{k^*=1}^s \binom{s}{k^*} \frac{s-1!}{k^*-1!} (\gamma+n-k)_{m-k^*} (k-\psi)_{k^*} (k-\gamma+\psi)_{k^*}$$

and the corresponding Bayesian estimator under quadratic loss function is given by

$$\mathbb{E}(S_m | K_n = k) = m \frac{(k-\psi)(k-\gamma+\psi)}{(n+\gamma)(n+\gamma-\psi)}.$$

## REFERENCES (RÉFÉRENCES)

- FISHER, R.A., CORBET, A.S. AND WILLIAMS, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Animal. Ecol.*, 12, 42–58.
- GNEDIN, A. (2010) A species sampling model with finitely many types. *Elect. Comm. Probab.*, 15, 79–88.
- GNEDIN, A. AND PITMAN, J. (2006) Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci*, 138, 3, 5674–5685.
- IRWIN, J.O. (1975) The Generalized Waring distribution. Part I. *J. R. Statist. Soc. A*, 138, p.18–31.
- LIJOI, A., MENA, R.H. AND PRÜNSTER, I. (2007) Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika*, 94, 769–786.
- LIJOI, A., PRÜNSTER, I. AND WALKER, S.G. (2008) Bayesian nonparametric estimator derived from conditional Gibbs structures. *Annals of Applied Probability*, 18, 1519–1547.
- PITMAN, J. AND YOR, M. (1997) The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25:855–900.
- RODRÍGUEZ-AVI, J., CONDE-SÁNCHEZ, A., SÁEZ-CASTILLO, A.J. AND OLMO-JIMÉNEZ, M.J. (2007) A new generalization of the Waring distribution. *Computational Statistics and Data Analysis*, 51, 6138–6150.
- RODRÍGUEZ-AVI, J., OLMO JIMÉNEZ, M. J., CONDE-SÁNCHEZ, A. AND SÁEZ-CASTILLO, A.J. (2008) The  ${}_3F_2$  with complex parameters as generating function of discrete distribution. *Communications in Statistics – Theory and Methods*, 37, 3009–3022.
- XEKALAKI, E. (1983) Infinite divisibility, completeness and regression properties of the univariate generalized Waring distribution. *Ann. Inst. Statist. Math.*, 35, 279–289

## ABSTRACT (RÉSUMÉ)

*In modern Bayesian statistics exchangeable Gibbs partitions (Gnedin and Pitman, 2006), are playing a central role both in the construction and properties of almost surely discrete nonparametric priors and in the nonparametric treatment of species sampling problems. Here, by means of an alternative parametrization of the two-parameter Gnedin-Fisher species sampling model (Gnedin, 2010), a family of Gibbs partitions arising by mixtures of Fisher’s models, we obtain additional results for the partition model itself and a posterior analysis in a Bayesian nonparametric perspective. In particular, identifying the prior mixing distribution on the number of blocks with the shifted generalized Waring distribution (a family of probability laws arising by Beta mixtures of Negative Binomial distributions), we provide a direct construction of the exchangeable partition probability function, obtain the conjugate posterior mixing, and derive the structural distribution and its moments. On the Bayesian side, in the spirit of Lijoi et al. (2008), we provide distributional results for an additional sample, conditional on a basic observed sample, for quantities of statistical interest in species sampling problems.*