Ann. Inst. Statist. Math.
Vol. 57, No. 3, 597–616 (2005)

UNORDERED AND ORDERED SAMPLE FROM DIRICHLET DISTRIBUTION

THIERRY HUILLET

Laboratoire de Physique Théorique et Modélisation, CNRS-UMR 8089 et Université de Cergy-Pontoise, 5 mail Gay-Lussac, 95031, Neuville sur Oise, France,
e-mail: Thierry.Huillet@ptm.u-cergy.fr

(Received March 2, 2004; revised September 1, 2004)

Abstract. Consider the random Dirichlet partition of the interval into n fragments with parameter $\theta > 0$. Explicit results on the statistical structure of its size-biased permutation are recalled, leading to (unordered) Ewens and (ordered) Donnelly-Tavaré-Griffiths sampling formulae from finite Dirichlet partitions. We use these preliminary statistical results on frequencies distribution to address the following sampling problem: what are the intervals between new sampled categories when sampling is from Dirichlet populations? The results obtained are in accordance with the ones found in sampling theory from random proportions with GEM(γ) distribution. These can be obtained from Dirichlet model when considering the Kingman limit $n \uparrow \infty$, $\theta \downarrow 0$ while $n\theta = \gamma > 0$.

Key words and phrases: Random discrete distribution, Dirichlet partition, size-biased permutation, GEM, Ewens and Donnelly-Tavaré-Griffiths sampling formulae, intervals between new sampled species.