



Asymptotic behavior of the number of distinct values in a sample from the geometric stick-breaking process

Pierpaolo De Blasi¹ · Ramsés H. Mena² · Igor Prünster³

Received: 2 September 2020 / Revised: 19 January 2021 / Accepted: 1 February 2021 /
Published online: 3 April 2021
© The Institute of Statistical Mathematics, Tokyo 2021

Abstract

Discrete random probability measures are a key ingredient of Bayesian nonparametric inference. A sample generates ties with positive probability and a fundamental object of both theoretical and applied interest is the corresponding number of distinct values. The growth rate can be determined from the rate of decay of the small frequencies implying that, when the decreasingly ordered frequencies admit a tractable form, the asymptotics of the number of distinct values can be conveniently assessed. We focus on the geometric stick-breaking process and we investigate the effect of the distribution for the success probability on the asymptotic behavior of the number of distinct values. A whole range of logarithmic behaviors are obtained by appropriately tuning the prior. A two-term expansion is also derived and illustrated in a comparison with a larger family of discrete random probability measures having an additional parameter given by the scale of the negative binomial distribution.

Keywords Bayesian nonparametrics · Random probability measure · Geometric stick-breaking process · Asymptotic growth rate · Occupancy problem

✉ Pierpaolo De Blasi
pierpaolo.deblasi@unito.it

¹ ESOMAS Department, University of Torino and Collegio Carlo Alberto, Piazza Arbarello 8, Turin, Italy

² Departamento de Probabilidad y Estadística, IIMAS Universidad Nacional Autónoma de México, Apartado Postal 20-726, Coyoacán 01000, Mexico

³ Department of Decision Sciences, Bocconi University and BIDSa, Via Roentgen 1, Milan, Italy