



General Type Token Distribution

[Shohei Hidaka](#)

(Submitted on 2 May 2013)

The problem of estimating the number of unique types or distinct species in a group occurs in many fields, but is not a straightforward one. Given an arbitrary probabilistic distribution of entries to be sampled, this study shows, in exact and approximated forms, the conditional probabilistic distribution of unique entries sampled independently given the number of samples. The numerical evaluation validates the proposed approximation. Finally, we discuss a theory that unifies the related approaches regarding estimation of the number of classes.

Comments: 15 pages, 2 figures

Subjects: **Methodology (stat.ME)**; Probability (math.PR)

MSC classes: 60C05, 62P99, 62F10

Cite as: [arXiv:1305.0328](#) [stat.ME]

(or [arXiv:1305.0328v1](#) [stat.ME] for this version)

Submission history

Download:

- [PDF](#)
- [PostScript](#)
- [Other formats](#)

Current browse context:

stat.ME

[< prev](#) | [next >](#)

[new](#) | [recent](#) | [1305](#)

Change to browse by:

[math](#)

[math.PR](#)

[stat](#)

References & Citations

- [NASA ADS](#)

Bookmark [\(what is this?\)](#)

