

Estimating the number of unseen species: How many words did Shakespeare know?

BY BRADLEY EFRON AND RONALD THISTED

Department of Statistics, Stanford University, California

SUMMARY

Shakespeare wrote 31 534 different words, of which 14 376 appear only once, 4343 twice, etc. The question considered is how many words he knew but did not use. A parametric empirical Bayes model due to Fisher and a nonparametric model due to Good & Toulmin are examined. The latter theory is augmented using linear programming methods. We conclude that the models are equivalent to supposing that Shakespeare knew at least 35 000 more words.