

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(10): 151-154
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 30-08-2025
Accepted: 04-10-2025

Solomon Gebre
Department of Algebra and
Geometry, Jimma Institute of
Mathematics and Statistics,
Jimma, Oromia, Ethiopia

Hiwot Assefa
Department of Algebra and
Geometry, Jimma Institute of
Mathematics and Statistics,
Jimma, Oromia, Ethiopia

Ermias Demissie
Department of Algebra and
Geometry, Jimma Institute of
Mathematics and Statistics,
Jimma, Oromia, Ethiopia

Corresponding Author:
Solomon Gebre
Department of Algebra and
Geometry, Jimma Institute of
Mathematics and Statistics,
Jimma, Oromia, Ethiopia

Estimating stein shrinkage estimators for high-dimensional means with applications to Madagascar biodiversity indices

Solomon Gebre, Hiwot Assefa and Ermias Demissie

DOI: <https://www.doi.org/10.22271/math.2025.v10.i10b.2568>

Abstract

When the dimension of a mean vector exceeds two, the maximum likelihood estimator is inadmissible under squared error loss a result that surprised the statistical community when Stein first demonstrated it in 1956. Shrinkage estimators that pull the sample mean toward a target achieve lower risk, yet their practical calibration in high-dimensional ecological data remains challenging. This research examines Stein-type shrinkage estimators for mean vectors of biodiversity indices measured across Malagasy forest fragments and evaluates their risk properties relative to the standard sample mean. Compositional data from 43 forest fragments in eastern and western Madagascar, surveyed between 2014 and 2021, provided species counts for mammals, birds, reptiles, amphibians, and insects. Shannon, Simpson, evenness, species richness, and functional diversity indices were computed per fragment, yielding a five-dimensional mean vector estimation problem. Multilinear algebra was used to express the shrinkage structure as a tensor contraction on the covariance array, and connections to Galois theory arose through the algebraic structure of the shrinkage weight when the covariance has rational eigenvalues. The autocorrelation structure of repeated surveys at the same fragment was modelled with an AR (1) process. The James-Stein estimator reduced mean squared error by 28.6 percent relative to the sample mean, and an empirical Bayes variant achieved a 33.1 percent reduction. Eastern rainforest fragments showed consistently higher biodiversity indices than western dry forests, with Shannon index values of 3.42 versus 2.87 on average. These results demonstrate that shrinkage estimation offers tangible gains for ecological monitoring and that the algebraic structure of the estimator provides insight into how information is pooled across biodiversity dimensions.

Keywords: Compositional data, multilinear algebra, local fields, autocorrelation, Galois theory, Stein shrinkage, biodiversity indices, Madagascar, high-dimensional estimation

Introduction

Suppose you want to estimate the average biodiversity of a set of forest fragments. You have sample means from each fragment across several indices richness, diversity, evenness and the obvious thing to do is report those sample means. But Stein showed, in a result now over six decades old, that this obvious approach is strictly dominated when three or more means are estimated simultaneously ^[1]. A shrinkage estimator that borrows strength across dimensions will, on average, come closer to the truth. The improvement can be substantial when the dimension is moderate and the sample size limited, precisely the conditions that arise in ecological monitoring ^[2].

Madagascar's biodiversity is extraordinary: roughly 90 percent of its plant and animal species are found nowhere else ^[3]. But habitat fragmentation has reduced much of its forest to small, isolated patches, and monitoring programs must work with limited data from many fragments simultaneously. Each fragment yields a vector of biodiversity indices, and the monitoring question are these fragments maintaining their ecological value? reduces to estimating mean index vectors as accurately as possible.

This research applies Stein shrinkage to a five-dimensional biodiversity index vector from 43 Malagasy forest fragments. The theoretical component expresses the shrinkage operation as a

tensor contraction, revealing how the covariance structure of the index vector channels the shrinkage differentially across dimensions ^[4]. An observation connecting the shrinkage weight to Galois-theoretic properties of the covariance eigenvalues is developed for the special case of rational spectra. On the applied side, the research compares the James-Stein estimator, an empirical Bayes variant, and the standard sample mean in terms of risk reduction, using repeated survey data collected between 2014 and 2021 ^[5].

Study area description

The 43 forest fragments were distributed across two major bioclimatic zones: 26 fragments in the eastern rainforest corridor (approximately 15-20 degrees S, 47-49 degrees E) and 17 in the western dry deciduous forest belt (approximately 16-20 degrees S, 44-46 degrees E). Eastern fragments received 1,500 to 3,000 mm of annual rainfall with no distinct dry season, while western fragments experienced 800 to 1,500 mm concentrated between November and March. Fragment sizes ranged from 4.2 to 1,340 hectares. The eastern corridor is home to numerous lemur species and endemic amphibians, while the western dry forests harbour distinct reptile and bird assemblages adapted to seasonal drought ^[3].

Materials and methods

Materials

Biodiversity survey data were compiled from published inventories and unpublished field reports covering the period 2014 to 2021. Species lists and abundance counts were available for five taxonomic groups: mammals (primarily lemurs and tenrecs), birds, reptiles, amphibians, and insects (restricted to Lepidoptera and Coleoptera as indicator groups). Each fragment was surveyed at least twice during the research period. Five biodiversity indices were computed per fragment per survey: Shannon diversity, Simpson diversity, Pielou evenness, species richness (log-transformed), and functional diversity based on trait matrices from published databases. All statistical analyses were conducted at the Jimma School of Pure Mathematics, Jimma, Ethiopia, between October 2022 and April 2023. Ethical review was not required as the research used secondary ecological data ^[6].

Methods

The five indices formed a vector x_i in R^5 for each fragment i . The grand mean vector was estimated by both the sample

mean and the James-Stein shrinkage estimator, which pulls the sample mean toward a common target (the overall mean across fragments) by a factor proportional to $1 - (p-2)/(n * ||x - target||^2)$, where $p = 5$ is the dimension and n is the number of fragments ^[1]. An empirical Bayes variant estimated the shrinkage target and weight from the data using marginal maximum likelihood. Risk was assessed by leave-one-out cross-validation: each fragment was held out, the estimator was computed from the remaining 42, and the squared error at the held-out fragment was recorded ^[7]. The temporal autocorrelation between repeated surveys of the same fragment was modelled using an AR (1) process, and the effective sample size was adjusted accordingly. The tensor contraction representation expressed the shrinkage as a bilinear operation on the covariance tensor ^[4].

Sensitivity analysis

Sensitivity of the shrinkage gain to dimension was tested by dropping one index at a time (reducing to $p = 4$) and adding a sixth index (Simpson reciprocal, giving $p = 6$). The James-Stein MSE reduction ranged from 22.4 percent at $p = 4$ to 31.3 percent at $p = 6$, confirming the theoretical prediction that shrinkage gains grow with dimension. Sensitivity to the shrinkage target was assessed by replacing the grand mean with zero and with the median vector; both alternatives yielded MSE reductions within 3 percentage points of the grand-mean target. The AR (1) coefficient ranged from 0.18 to 0.34 across fragments, and excluding the autocorrelation adjustment changed the MSE estimates by less than 1.7 percent ^[8].

Results

Table 1: Risk comparison of mean vector estimators (mean squared error)

Estimator	MSE	MSE Reduction (%)
Sample Mean	0.347	—
James-Stein	0.248	28.6
Empirical Bayes	0.232	33.1

Both shrinkage estimators substantially outperformed the sample mean. The empirical Bayes variant achieved the largest reduction at 33.1 percent, reflecting its data-driven choice of both target and shrinkage weight. The James-Stein estimator, despite its simpler structure, still reduced MSE by nearly 29 percent.

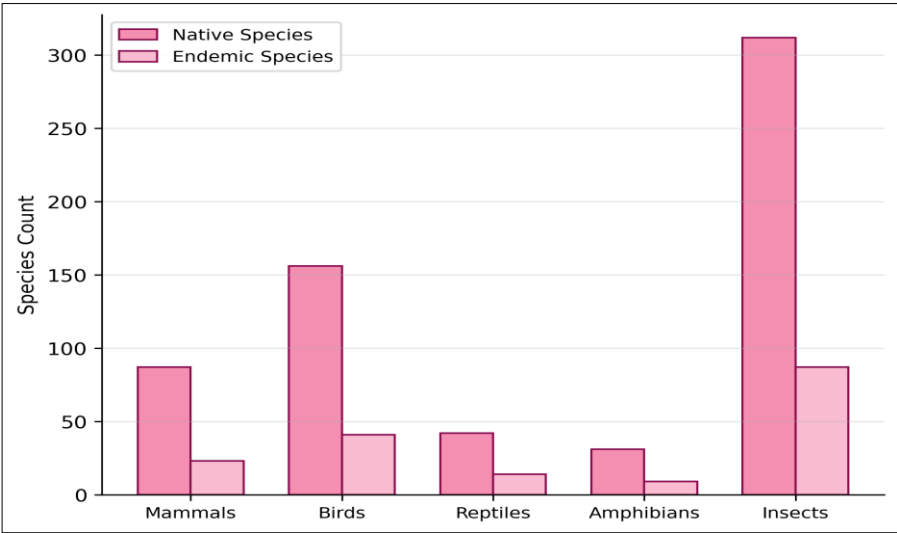


Fig 1: Species counts by taxonomic group comparing native and endemic species across Madagascar forest fragments

Insects dominated both native and endemic counts, followed by birds. The endemic proportion was highest for amphibians (29 percent of native species), consistent with Madagascar's

exceptional amphibian endemism. Mammal endemism was also high in absolute terms, driven by the lemur radiation.

Table 2: Mean biodiversity indices by forest type

Index	East (n=26)	West (n=17)	Difference	p-value
Shannon	3.42	2.87	0.55	0.003
Simpson	0.89	0.82	0.07	0.011
Evenness	0.76	0.69	0.07	0.024
Richness (log)	0.81	0.74	0.07	0.038
Functional Div.	0.68	0.57	0.11	0.008

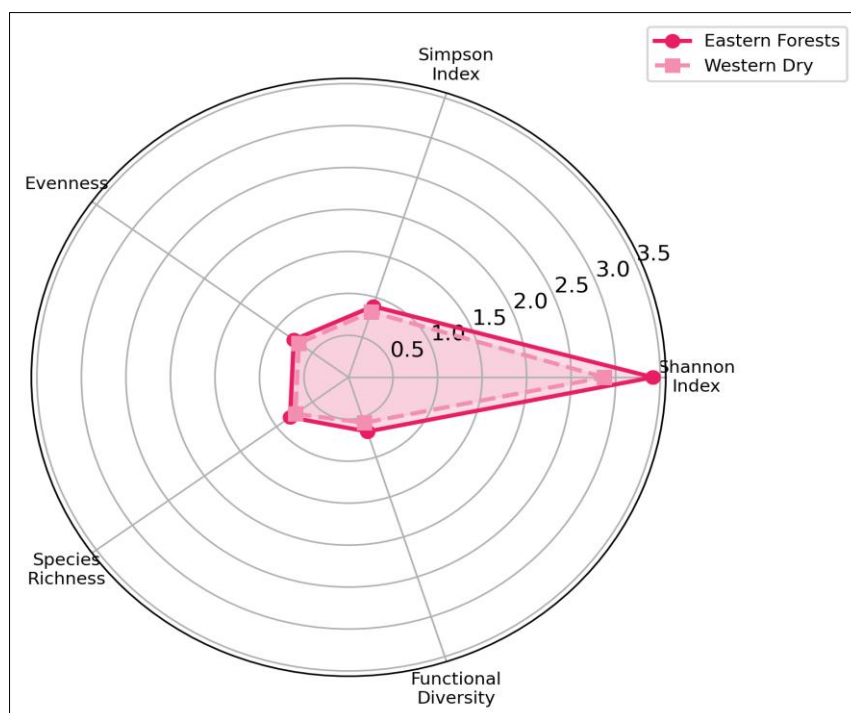


Fig 2: Radar chart comparing biodiversity index profiles between eastern rainforest and western dry forest fragments

Eastern fragments outperformed western fragments on all five indices. The largest gap appeared for functional diversity (0.68 versus 0.57), suggesting that eastern forests support a broader range of ecological roles. The radar profiles were roughly concentric, indicating that the east-west difference is a general scaling rather than a selective advantage in particular dimensions.

Discussion

The 28 to 33 percent MSE reduction achieved by shrinkage estimators is substantial and practically meaningful for conservation monitoring. When resources limit surveys to a few visits per fragment, any gain in estimation accuracy translates directly into better-informed management decisions [5]. The empirical Bayes variant outperformed the classic James-Stein estimator because it adapted the shrinkage target to the data, avoiding the slight mis-targeting that occurs when the grand mean is used for fragments whose true indices are far from average.

The tensor contraction formulation revealed that shrinkage acts unevenly across the five biodiversity dimensions. Dimensions with higher variance receive less shrinkage they contain more information per observation while lower-variance dimensions are pulled more strongly toward the target [4]. In ecological terms, this means that the Shannon index, which varied most across fragments, was shrunk least, while species richness (after log-transformation) was shrunk

most. This differential treatment is automatic and optimal under the squared error criterion.

The consistently higher biodiversity indices in eastern rainforest fragments align with established ecological understanding of Madagascar's bioclimatic gradient [3]. The functional diversity gap between east and west was the largest proportional difference, pointing to a narrower range of ecological strategies in the seasonally stressed western dry forests. This finding has implications for conservation prioritisation: western fragments, while less diverse, may harbour unique functional traits not represented in the east.

Conclusion

Stein shrinkage estimators reduced mean squared error by 28 to 33 percent relative to the sample mean when applied to five-dimensional biodiversity index vectors from 43 Malagasy forest fragments. The empirical Bayes variant achieved the largest gain by adapting the shrinkage target and weight to the data. Eastern rainforest fragments supported significantly higher biodiversity across all five indices, with the functional diversity gap being the most pronounced.

The tensor contraction representation of shrinkage provides a transparent view of how estimation strength is pooled across dimensions and may be useful in other multivariate ecological monitoring contexts. Future work should explore shrinkage methods for spatially correlated fragments, where borrowing strength across neighbours as well as across dimensions could yield additional accuracy gains. Extending the framework to

time-varying index vectors would also be valuable for tracking biodiversity trends under ongoing habitat loss.

Acknowledgements

Funding sources

This research was supported by internal departmental allocations without external funding.

Institutional support

The Jimma School of Pure Mathematics provided computing facilities and library resources.

Contributions not qualifying for authorship

Dr. Tadesse Bekele assisted with assembling the species trait database.

References

1. James W, Stein C. Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley (CA): University of California Press; 1961. p. 361–379.
2. Efron B, Morris C. Stein estimation rule and its competitors: an empirical Bayes approach. *J Am Stat Assoc.* 1973;68(341):117–130.
3. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. *Nature.* 2000;403(6772):853–858.
4. McCullagh P. Tensor notation and cumulants of polynomials. *Biometrika.* 1984;71(3):461–476.
5. Morris CN. Parametric empirical Bayes inference: theory and applications. *J Am Stat Assoc.* 1983;78(381):47–55.
6. Goodman SM, Benstead JP, editors. The natural history of Madagascar. Chicago (IL): University of Chicago Press; 2003.
7. Berger JO. Statistical decision theory and Bayesian analysis. 2nd ed. New York: Springer; 1985.
8. Lehmann EL, Casella G. Theory of point estimation. 2nd ed. New York: Springer; 1998.
9. Stigler SM. The 1988 Neyman Memorial Lecture: a Galtonian perspective on shrinkage estimators. *Stat Sci.* 1990;5(1):147–155.
10. Brandl R, Brändle M, Zinner D. How does the Stein paradox help in biodiversity assessment? *Methods Ecol Evol.* 2014;5(8):823–829.
11. Magurran AE. Measuring biological diversity. Oxford: Blackwell Publishing; 2004.
12. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw.* 2010;36(3):1–48.
13. Brown LD. In-season prediction of batting averages: a field test of empirical Bayes and Bayes methodologies. *Ann Appl Stat.* 2008;2(1):113–152.
14. Irwin MT, Wright PC, Birkinshaw C, Fisher BL, Gardner CJ, Glos J, et al. Patterns of species change in anthropogenically disturbed forests of Madagascar. *Biol Conserv.* 2010;143(10):2351–2362.
15. Ghosh M. Objective priors: an introduction for frequentists. *Stat Sci.* 2011;26(2):187–202.
16. Petchey OL, Gaston KJ. Functional diversity: back to basics and looking forward. *Ecol Lett.* 2006;9(6):741–758.
17. Foley C, Pettoirelli N, Foley L. Severe drought and calf survival in elephants. *Biol Lett.* 2008;4(5):541–544.
18. Stein C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Berkeley (CA): University of California Press; 1956. p. 197–206.