Universität Bamberg

Bootstrap vs. Jackknife Variance Estimation for the Gini Coefficient

A Monte Carlo Study

Niklas Ippisch   Moritz Kolbe   Justus Mennerich

Institute of Statistics - Otto-Friedrich-University Bamberg

Motivation

- Inference for the Gini is challenging due to its **nonlinearity** and its **sensitivity to tail observations**.
- Point estimation is straightforward; **variance estimation** in finite samples is **substantially more difficult**.
- Closed-form variance expressions are often **unavailable**, **analytically cumbersome**, or **impractical** for empirical settings (Walter et al., 2022; Kreutzmann et al., 2019).
- In applied work, uncertainty quantification therefore relies almost exclusively on **resampling procedures** like **Bootstrap** and **Jackknife**.
- This raises the methodological question of how well resampling procedures behave under design-based inference and across **varying income distributions** and **domain sizes**.

Simulation Design

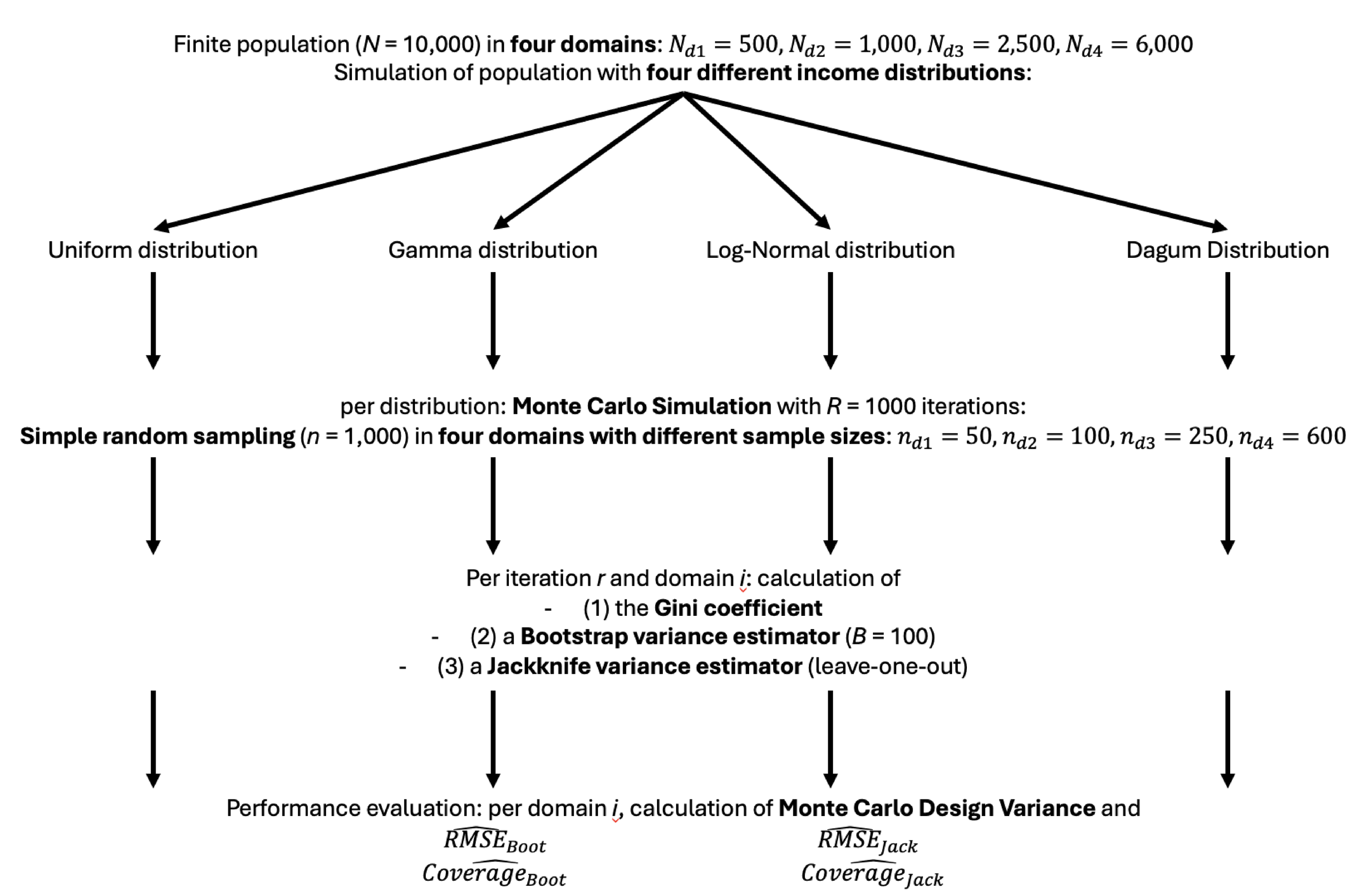

Figure 1: Procedure for the present simulation study.

Variance Estimators

- **Non-Parametric Bootstrap** variance estimator (R package `emdi`):

$$\widehat{Var}_{Boot,r}(G) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{G}_b^* - \hat{G}^*)^2 \qquad \text{with } \hat{G}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{G}_b^*$$

- **Leave-one-out Jackknife** variance estimator (own implementation):

$$\widehat{Var}_{Jack,r}(G) = \frac{A-1}{A}\sum_{a=1}^{A}(\hat{G}_j^* - \hat{G}^*)^2 \qquad \text{with } \hat{G}^* = \frac{1}{A}\sum_{a=1}^{A}\hat{G}_j^*$$

Income Distributions

- **Uniform Distribution**
  - Reference distribution for a (hypothetical) case without inequality
  - $f(x) = \frac{1}{U-L}, L \leq x \leq U, U = 2550, L = 2450$
- **Gamma Distribution**
  - Often chosen to model income distributions with a long tail (i.e., outliers to the top)
  - $f(x;\alpha,\beta) = \frac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}, \alpha = 0.5, \beta = \frac{1}{5000}$
- **Log-Normal Distribution**
  - Also often used to model income distributions in research applications
  - $f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}}, x > 0, \mu = 7.735, \sigma = 0.312$
- **Dagum Distribution**
  - Part of bigger distributional family of Generalized Beta Distributions (Kleiber, 2008)
  - Shown to provide best model fit for income distribution in most European countries (Bandourian et al., 2002)
  - $f(x) = \frac{apx^{ap-1}}{b^{ap}[1+(x/b)^a]^{p+1}}, x > 0, a = 4.413, p = 0.337, b = 3507.611$

Evaluation Metrics

- **Relative Root Mean Squared Error (RMSE)**
  - RMSE measures the distance between the estimated variance (Bootstrap or Jackknife) and the design-based variance per distribution, domain, and method m.

$$\widehat{RMSE}_m = \sqrt{\tfrac{1}{R}\textstyle\sum_{r=1}^{R}(\widehat{Var}_{m,r} - \widehat{Var}_{MC})^2}; \qquad \text{Relative RMSE: } \widehat{RRMSE}_m = \frac{\widehat{RMSE}_m}{\widehat{Var}_{MC}}$$

- **Coverage**
  - 95%-Confidence-Intervals constructed with the estimated variance per domain *d* and method *m*:

$$\widehat{Cov}_m = \frac{1}{R}\sum_{r=1}^{R} I\left(G \in CI = [\hat{G}_r \pm 1.96 \cdot \sqrt{\widehat{Var}_{m,r}}]\right)$$

Relative RMSE

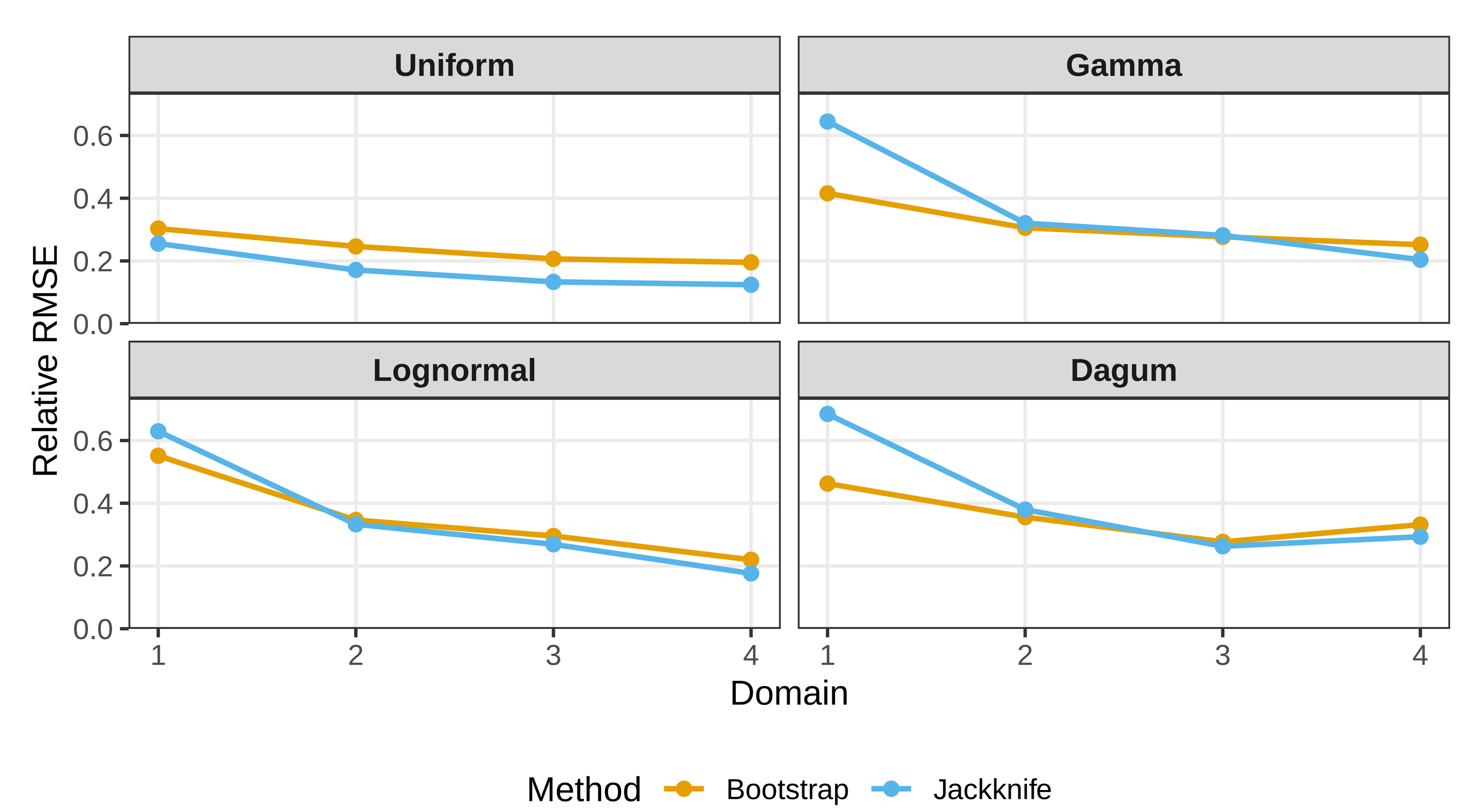

Figure 2: Relative RMSE over different domains per distribution.

Coverage of 95% confidence intervals

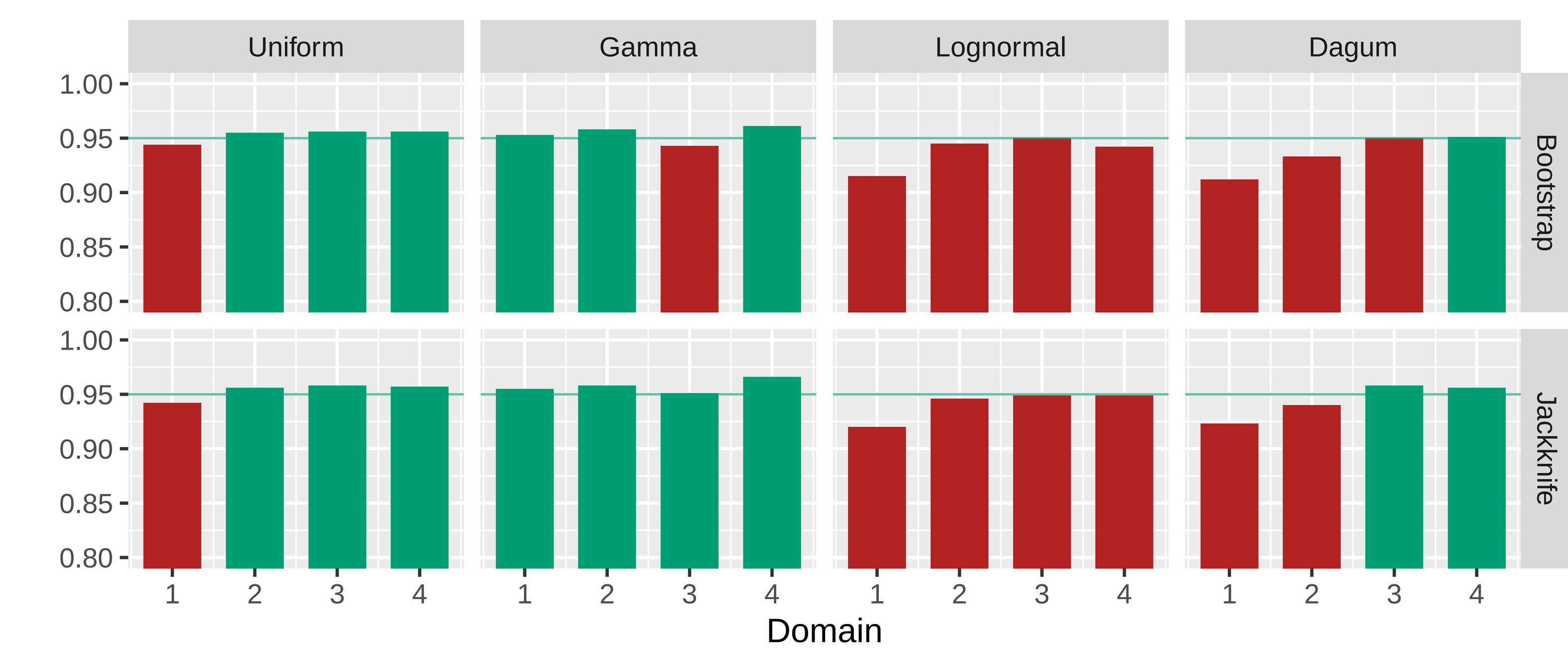

Figure 3: Coverage over domains per distribution;
Note: Y-axis was adjusted to highlight small differences in coverage
(red: $\widehat{Cov}_m < 0.95$; green: $\widehat{Cov}_m \geq 0.95$).

Discussion and Limitations

- **Discussion**
  - With respect to relative RMSE, **bootstrap is more robust** in heavy-tailed settings for very small domain sizes (e.g., $n = 50$). Jackknife performs better under uniform distributions and **improves markedly with increasing domain size**, eventually narrowing the performance gap in heavy-tailed scenarios.
  - Jackknife benefits substantially from larger domains due to **decreasing influence of individual extreme observations**, while **Bootstrap improves only moderately**.
  - Combining relative RMSE with empirical coverage enables a **more differentiated assessment** of the inferential performance of both resampling methods.
- **Limitations**
  - Bootstrap iterations *B* were **kept constant** whereas Jackknife has an increasing amount of data for increasing domain sizes.
  - No **alternative Bootstrap variants** (e.g., BCa, studentized, or m-out-of-n) were evaluated. (Banerjee Chakrabarty, 2025; Panichkitkosolkul, 2023)
  - Only **Wald-type confidence intervals** were considered. Since Wald intervals rely on a normal approximation and plug-in variance estimation, they may exhibit undercoverage in small samples and under heavy-tailed distributions
  - Just **simple random sampling** and no survey weights, stratification, or clustering were incorporated in the setting.

References

[1] Bandourian, R., McDonald, J. B. & Turley, R. S. (2002). A Comparison of Parametric Models of Income Distribution Across Countries and Over Time *Estadística*, 55, 135–152.

[2] Banerjee, I. & Chakrabarty, S. (2025). CLT and Edgeworth Expansion for m-out-of-n Bootstrap Estimators of the Studentized Median *arXiv preprint*, arXiv:2505.11725.

[3] Kleiber, C. (2008). A Guide to the Dagum Distributions. In: D. Chotikapanich (Eds.). *Modeling Income Distributions and Lorenz Curves* (pp. 97-117)

[4] Kreutzmann, A.-K., Pannier, S., Rojas-Perilla, N., Schmid, T., Templ, M., and Tzavidis, N. (2019). The R package `emdi` for estimating and mapping regionally disaggregated indicators. *Journal of Statistical Software*, 91(7), 1–33.

[5] Panichkitkosolkul, W. (2023). Bootstrap Methods for Estimating the Confidence Interval for the Parameter of the Zero-Truncated Poisson-Sujatha Distribution and Their Applications *Sains Malaysiana*, 52(6), 1879–1888.

[6] Walter, P., Groß, M., Schmid, T., and Weimer, K. (2022). Iterative kernel density estimation applied to grouped data: Estimating poverty and inequality indicators from the German Microcensus. *Unpublished manuscript*.

Reproduction Materials available on GitHub

Small Area Estimation

{niklas-matthias.ippisch; moritz.kolbe; justus-lasse.mennerich}@stud.uni-bamberg.de