HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

Score-Adjusted Moment Estimators for the Dirichlet Distribution Family

Ioannis Oikonomidis and Samis Trevezas

Department of Statistics
Athens University of Economics and Business

11.11.2025

Presentation Structure

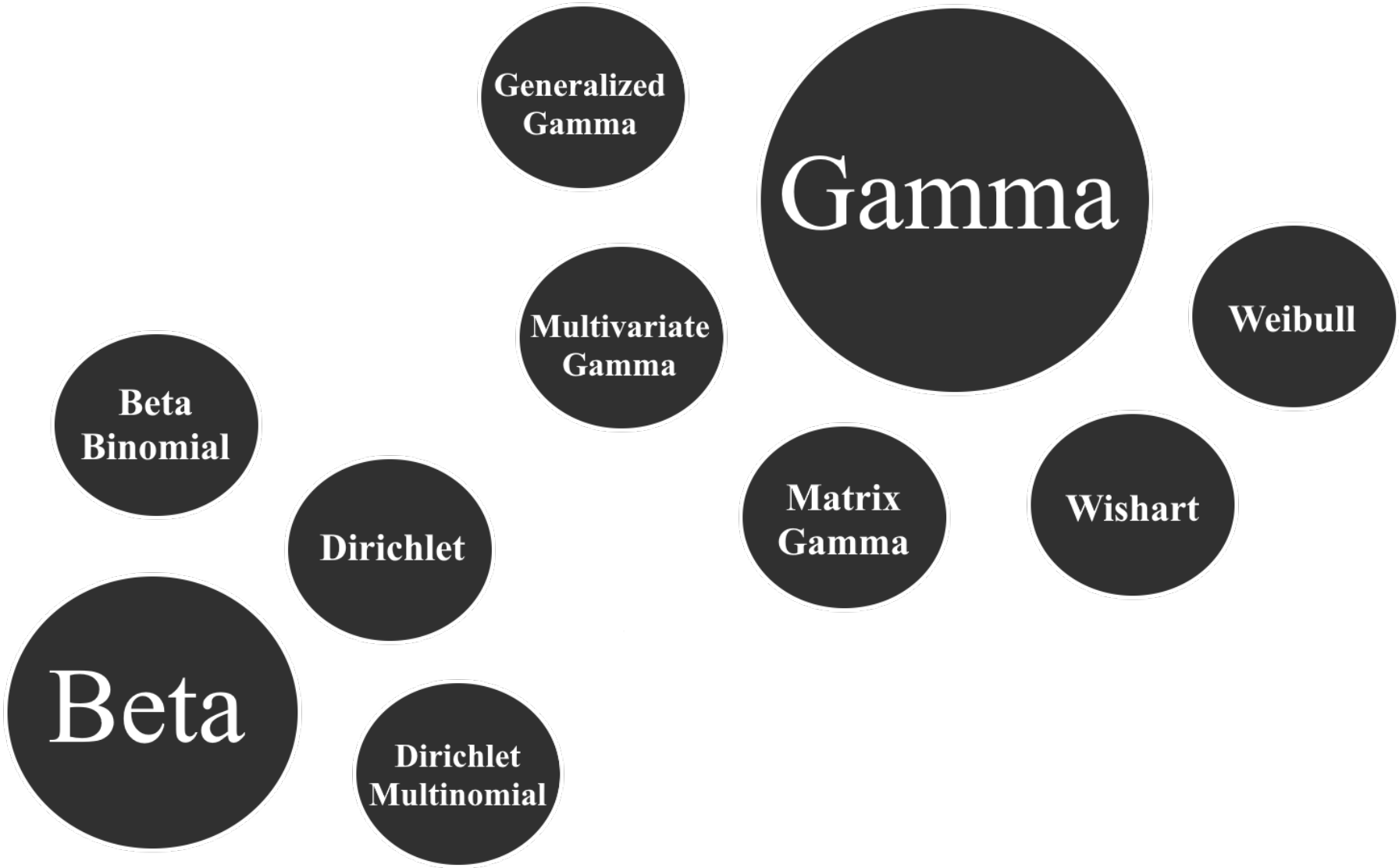
Score-Adjusted Moment Estimators for the Dirichlet Distribution Family

1. Introduction and Study Domain
2. Published Work
   - Moment-Type Estimators for the Dirichlet and the Multivariate Gamma Distributions.
3. Work in Progress
   - An Analytical Comparison of the Dirichlet Moment Estimators.
   - The `joker` Package: Object-Oriented Distribution System and Parameter Estimation in `R`.

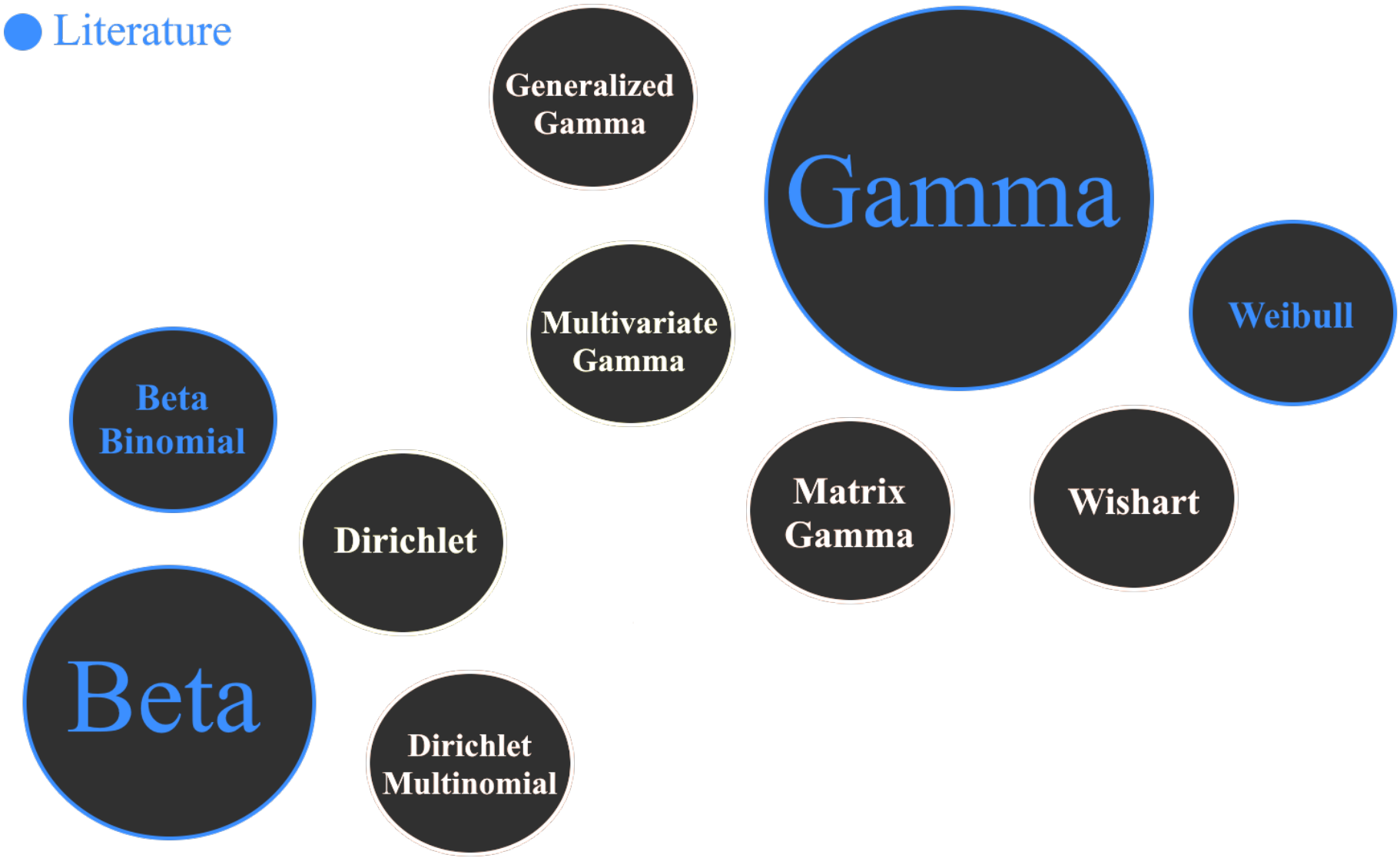
Score-Adjusted Moment Estimators for the Dirichlet Distribution Family

1. Introduction and Study Domain

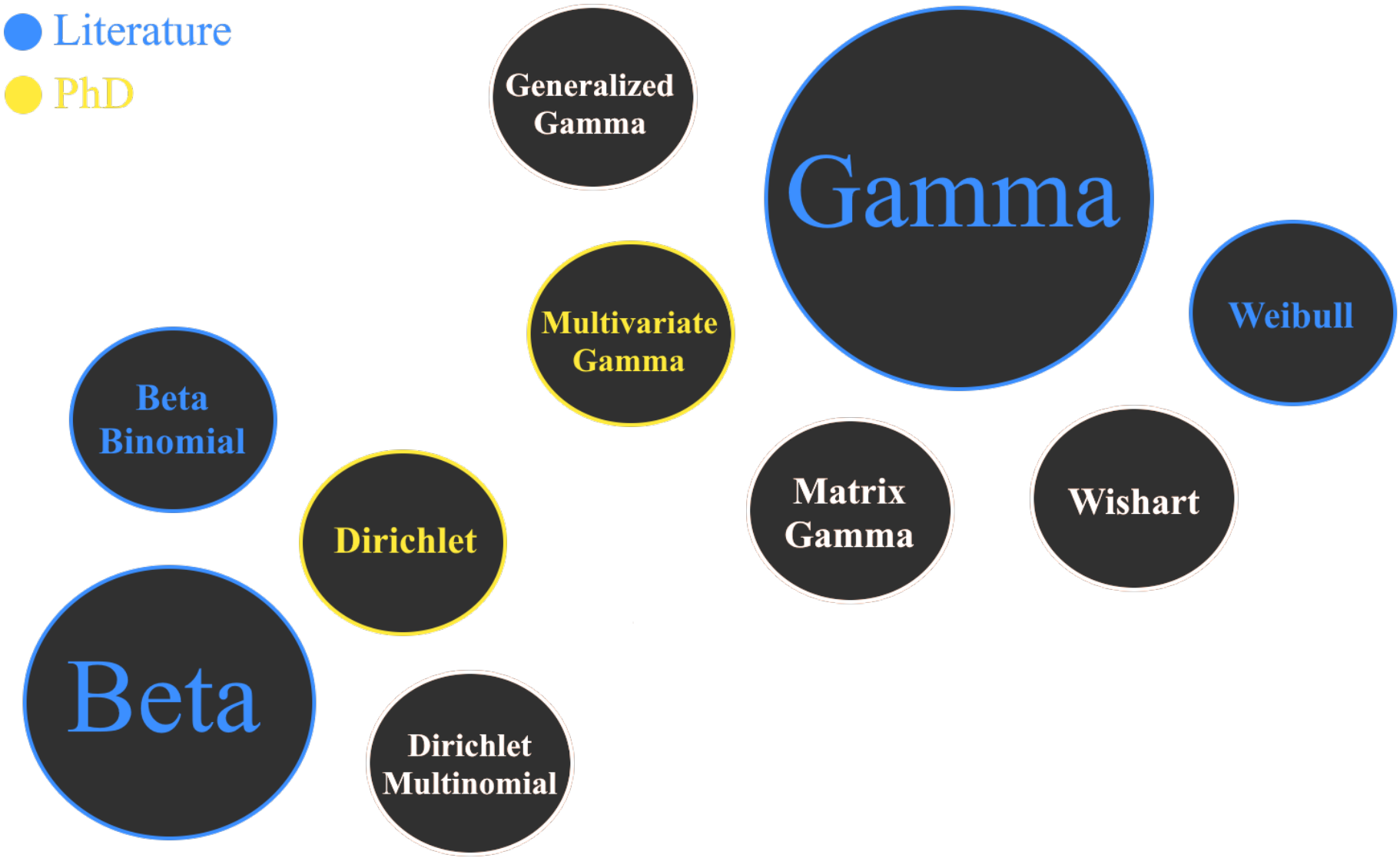
Distribution Families Without Explicit MLE



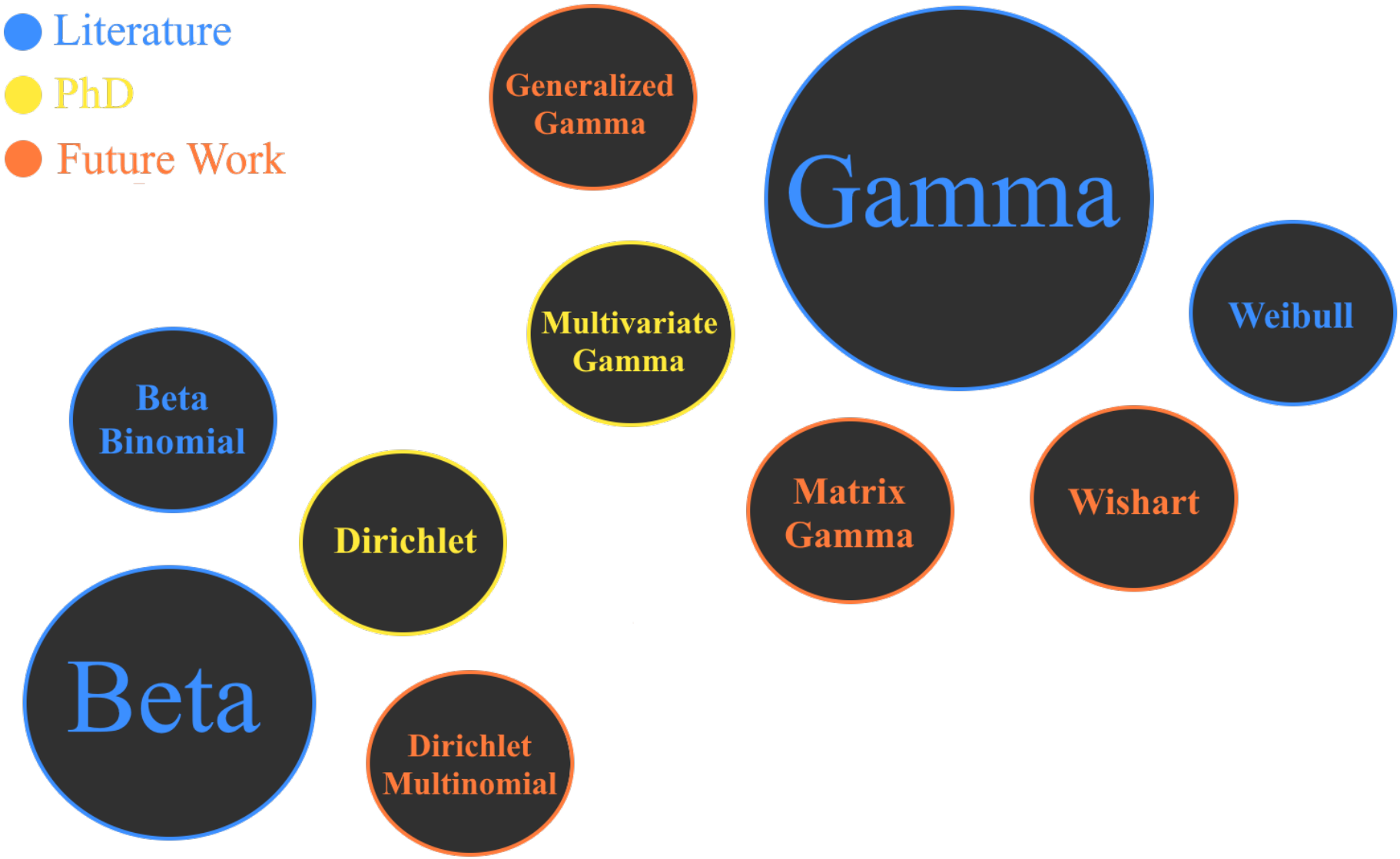
Distribution Families Without Explicit MLE



Distribution Families Without Explicit MLE



Distribution Families Without Explicit MLE



Beta Moment-Type Estimators

Proposition

Let X_n, $n \in [N]$, be a random sample from $\mathcal{B}(\alpha, \beta)$, $Y_n := 1 - X_n$, $n \in [N]$.

- The Moment Estimator can be found by solving the equation system

$$\overline{X} = \frac{\tilde{\alpha}}{\tilde{\alpha} + \tilde{\beta}}, \quad \overline{X^2} - \overline{X}^2 = \frac{\tilde{\alpha}\tilde{\beta}}{(\tilde{\alpha} + \tilde{\beta})^2(\tilde{\alpha} + \tilde{\beta} + 1)}.$$

  The solution of the system is

$$\tilde{\alpha} = \frac{\overline{X}(\overline{X} - \overline{X^2})}{\overline{X^2} - \overline{X}^2}, \quad \tilde{\beta} = \frac{\overline{Y}(\overline{Y} - \overline{Y^2})}{\overline{Y^2} - \overline{Y}^2}.$$

- The Maximum Likelihood Estimator is the unique solution of the system

$$\overline{\ln X} = \psi(\hat{\alpha}) - \psi(\hat{\alpha} + \hat{\beta}), \quad \overline{\ln Y} = \psi(\hat{\beta}) - \psi(\hat{\alpha} + \hat{\beta}).$$

Tamae, H., Irie, K., & Kubokawa, T. (2020). A score-adjusted approach to closed-form estimators for the gamma and beta distributions. Japanese Journal of Statistics and Data Science, 3, 543-561.

Beta Score-Adjusted Moment Estimator

Remark

Let X_n, $n \in [N]$, be a random sample from $\mathcal{B}(\alpha, \beta)$ $Y_n := 1 - X_n$, $n \in [N]$. Then,

$$\mathbb{C}(X_n, \ln X_n) + \mathbb{C}(Y_n, \ln Y_n) = \frac{1}{\alpha + \beta}, \quad n \in [N],$$

which can be estimated by $s_{xy} := \overline{X \ln X} - \overline{X}\,\overline{\ln X} + \overline{Y \ln Y} - \overline{Y}\,\overline{\ln Y}$.

Proposition

Let X_n, $n \in [N]$, be a random sample from $\mathcal{B}(\alpha, \beta)$. The Score - Adjusted Moment Estimator is the solution of the system

$$\overline{X} = \frac{\breve{\alpha}}{\breve{\alpha} + \breve{\beta}}, \quad s_{xy} = \frac{1}{\breve{\alpha} + \breve{\beta}}.$$

The solution of the system is $\breve{\alpha} = \overline{X}/s_{xy}$, $\breve{\beta} = \overline{Y}/s_{xy}$.

Bibliography

1. Ye, Zhi-Sheng, and Nan Chen. "Closed-form estimators for the gamma distribution derived from likelihood equations." The American Statistician 71.2 (2017): 177-181.

2. Tamae, Hiromasa, Kaoru Irie, and Tatsuya Kubokawa. "A score-adjusted approach to closed-form estimators for the gamma and beta distributions." Japanese Journal of Statistics and Data Science 3 (2020): 543-561.

3. Papadatos, Nickos D. "On point estimators for Gamma and Beta distributions." The American Statistician 78.4 (2024): 412-417.

4. Wiens, Douglas P., Julian Cheng, and Norman C. Beaulieu. "A class of method of moments estimators for the two-parameter gamma family." Pakistan Journal of Statistics 19.1 (2003): 129-141.

Bibliography

5. Tamae, Hiromasa, Kaoru Irie, and Tatsuya Kubokawa. "Score-adjusted methods for estimation of shape parameters in Gamma-Poisson and Beta-Binomial distributions." Communications in Statistics-Simulation and Computation (2022): 1-11.

6. Zhao, Jun, SungBum Kim, and Hyoung-Moon Kim. "Closed-form estimators and bias-corrected estimators for the Nakagami distribution." Mathematics and Computers in Simulation 185 (2021): 308-324.

7. Chen, Piao, and Xun Xiao. "Novel closed-form point estimators for the beta distribution." Statistical theory and related fields 9.1 (2025): 12-33.

8. Jang, Yu-Hyeong, et al. "New closed-form efficient estimator for multivariate gamma distribution." Statistica Neerlandica 77.4 (2023): 555-572.

Score-Adjusted Moment Estimators for the Dirichlet Distribution Family

2. Completed Work

ELSEVIER

Journal of Multivariate Analysis

Volume 210, November 2025, 105471

Moment-type estimators for the Dirichlet and the multivariate gamma distributions

Ioannis Oikonomidis, Samis Trevezas

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.jmva.2025.105471

Get rights and content

The Dirichlet Distribution Family

Definition

A random vector $\mathbf{X}$ is said to follow the Dirichlet distribution with parameters $\boldsymbol{\alpha} \in (0, +\infty)^k$, denoted by $\mathbf{X} \sim \mathcal{D}(\boldsymbol{\alpha})$, if it has a probability density function of the form

$$f(\mathbf{x}; \boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} x_i^{\alpha_i - 1} \, \mathbb{I}_{\mathcal{X}}(\mathbf{x}),$$

where $\alpha_0 = \sum_{i=1}^{k} \alpha_i$ and $\mathcal{X} = \{\mathbf{x} \in (0,1)^k : \sum_{i=1}^{k} x_i = 1\}$.

Lemma

The $\mathcal{D}(\boldsymbol{\alpha})$ distribution is a $(k-1)$-variate k-dimensional exponential family with

$$\eta(\boldsymbol{\alpha}) = \boldsymbol{\alpha}, \quad T(\mathbf{X}) = (\ln X_1, \ldots, \ln X_k), \quad A(\boldsymbol{\alpha}) = \sum_{i=1}^{k} \ln \Gamma(\alpha_i) - \ln \Gamma(\alpha_0).$$

Dirichlet Maximum Likelihood Estimator

Proposition

Let $\mathbf{X}_n = (X_{1n}, \ldots, X_{kn})$, $n \in [N]$, be a random sample from $\mathcal{D}_k(\boldsymbol{\alpha})$. The MLE of $\boldsymbol{\alpha}$, derived as the unique solution of the equation system

$$\overline{\ln(X_i)} = \psi(\widehat{\alpha}_i) - \psi(\widehat{\alpha}_0), \qquad i \in [k],$$

is a moment-type estimator with asymptotic variance-covariance matrix $\boldsymbol{\Sigma}$ such that

$$\boldsymbol{\Sigma}_{ij} = \frac{\delta_{ij}}{\psi_1(\alpha_i)} + \frac{\psi_1(\alpha_0)}{\psi_1(\alpha_i)\psi_1(\alpha_j)\left\{1 - \psi_1(\alpha_0)\sum_{m=1}^{k}\psi_1^{-1}(\alpha_m)\right\}}, \quad i, j \in [k],$$

where δ_{ij} is Kronecker's delta, i.e. $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise, $i, j \in [k]$.

Dirichlet Maximum Likelihood Estimator

Proof.

The MLE is a moment-type estimator and its asymptotic variance-covariance matrix takes the form $\boldsymbol{\Sigma} = [\mathbb{V}(\ln \mathbf{X})]^{-1}$. The elements of $\mathbb{V}(\ln \mathbf{X})$ are:

$$\mathbb{C}(\ln X_i, \ln X_j) = \delta_{ij}\psi_1(\alpha_i) - \psi_1(\alpha_0).$$

The matrix can be broken down to $\mathbf{A} + \mathbf{u}\mathbf{w}^\top$, where

$$\mathbf{A} = \operatorname{diag}\{\psi_1(\boldsymbol{\alpha})\}, \quad \mathbf{u} = -\psi_1(\alpha_0)\mathbf{1}_{k\times 1}, \quad \mathbf{w} = \mathbf{1}_{k\times 1}.$$

This form allows the application of the Sherman - Morrison formula:

$$\boldsymbol{\Sigma} = [\mathbb{V}(\ln \mathbf{X})]^{-1} = \left(\mathbf{A} + \mathbf{u}\mathbf{w}^\top\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{w}^\top\mathbf{A}^{-1}}{1 + \mathbf{w}^\top\mathbf{A}^{-1}\mathbf{u}}.$$

The individual parts can easily be calculated, and the result follows. □

Dirichlet Moment Estimator - ME_1

Proposition

Let $\mathbf{X}_n = (X_{1n}, \ldots, X_{kn})$, $n \in [N]$, be a random sample from $\mathcal{D}_k(\boldsymbol{\alpha})$. A Moment Estimator ($\mathrm{ME}_1$) can be found by solving the system

$$\overline{X_i} = \frac{\widetilde{\alpha}_i}{\widetilde{\alpha}_0}, \quad \overline{X_i^2} - \overline{X_i}^2 = \frac{\widetilde{\alpha}_i(\widetilde{\alpha}_0 - \widetilde{\alpha}_i)}{\widetilde{\alpha}_0^2(\widetilde{\alpha}_0 + 1)}, \qquad i \in [k].$$

The solution of the system is

$$\widetilde{\alpha}_i = \frac{\overline{X_i}\left(\overline{X_i} - \overline{X_i^2}\right)}{\overline{X_i^2} - \overline{X_i}^2}, \qquad i \in [k].$$

The asymptotic variance-covariance matrix $\boldsymbol{\Sigma}$ is such that, for $i, j \in [k]$:

$$\boldsymbol{\Sigma}_{ij} = \frac{(\alpha_i + 1)(\alpha_j + 1)(\delta_{ij}\alpha_0 - \alpha_i)\alpha_j\alpha_0}{(\alpha_0 - \alpha_i)(\alpha_0 - \alpha_j)(\alpha_0 + 2)(\alpha_0 + 3)} \left\{ \frac{2(\alpha_0 + 1)^2}{\alpha_j + 1}\delta_{ij} - \frac{2\alpha_0^2 + \alpha_0 + 1}{\alpha_0 + 1} \right\}.$$

Dirichlet Moment Estimator - ME_2

Proposition

Let $\mathbf{X}_n = (X_{1n}, \ldots, X_{kn})$, $n \in [N]$ be a random sample from $\mathcal{D}_k(\boldsymbol{\alpha})$. A Moment Estimator ($\mathrm{ME}_2$) can be found by solving the system

$$\overline{X_i} = \frac{\tilde{\tilde{\alpha}}_i}{\tilde{\tilde{\alpha}}_0}, \quad i \in [k-1], \qquad \sum_{m=1}^{k} \left(\overline{X_m^2} - \overline{X_m}^2 \right) = \frac{\tilde{\tilde{\alpha}}_0^2 - \sum_{m=1}^{k} \tilde{\tilde{\alpha}}_m^2}{\tilde{\tilde{\alpha}}_0^2 (\tilde{\tilde{\alpha}}_0 + 1)}.$$

The solution of the system is:

$$\tilde{\tilde{\alpha}}_0 = \frac{1 - \sum_{m=1}^{k} \overline{X_m^2}}{\sum_{m=1}^{k} \left(\overline{X_m^2} - \overline{X_m}^2 \right)}, \qquad \tilde{\tilde{\alpha}}_i = \tilde{\tilde{\alpha}}_0 \, \overline{X_i} \quad i \in [k].$$

Dirichlet Moment Estimator - ME_2

Proposition

The asymptotic variance-covariance matrix $\boldsymbol{\Sigma}$ of ME_2 is such that, for $i, j \in [k]$:

$$\boldsymbol{\Sigma}_{ij} = \frac{\alpha_0 \alpha_i}{\alpha_0 + 1} \delta_{ij} + \frac{2\alpha_0 \alpha_i \alpha_j (\alpha_i + \alpha_j)}{b(\boldsymbol{\alpha})(\alpha_0 + 2)} + \frac{\alpha_0 \alpha_i \alpha_j p(\boldsymbol{\alpha})}{b^2(\boldsymbol{\alpha})(\alpha_0 + 1)(\alpha_0 + 2)(\alpha_0 + 3)},$$

where

$$b(\boldsymbol{\alpha}) = \alpha_0^2 - \sum_{m=1}^{k} \alpha_m^2,$$

$$c(\boldsymbol{\alpha}) = \alpha_0^3 - \sum_{m=1}^{k} \alpha_m^3,$$

$$p(\boldsymbol{\alpha}) = 4(\alpha_0 + 1)^2(\alpha_0 - 1)c(\boldsymbol{\alpha}) + (2\alpha_0^2 + \alpha_0 + 1)b^2(\boldsymbol{\alpha}) + (-6\alpha_0^4 - 4\alpha_0^3 + 4\alpha_0^2 + 4\alpha_0 + 2)b(\boldsymbol{\alpha}).$$

Dirichlet Score - Adjusted Moment Estimator

Proposition

Let $\mathbf{X}_n = (X_{1n}, \ldots, X_{kn}), n \in [N]$ be a random sample from $\mathcal{D}_k(\boldsymbol{\alpha})$. The Score-Adjusted Moment Estimator can be found by solving the system:

$$\overline{X_i} = \frac{\breve{\alpha}_i}{\breve{\alpha}_0}, \quad i \in [k-1], \qquad \sum_{m=1}^{k} \left(\overline{X_m \ln X_m} - \overline{X_m}\,\overline{\ln X_m} \right) = \frac{k-1}{\breve{\alpha}_0}.$$

The solution of the system is:

$$\breve{\alpha}_i = \frac{(k-1)\overline{X_i}}{\sum_{m=1}^{k} \left(\overline{X_m \ln X_m} - \overline{X_m}\,\overline{\ln X_m} \right)}, \quad i \in [k].$$

The asymptotic variance-covariance matrix $\boldsymbol{\Sigma}$ is such that, for $i, j \in [k]$:

$$\boldsymbol{\Sigma}_{ij} = \frac{\alpha_0 \alpha_i}{\alpha_0 + 1}\delta_{ij} + \alpha_i \alpha_j \sum_{m=1}^{k} \frac{\alpha_m(\alpha_0 - \alpha_m)\psi_1(\alpha_m)}{(\alpha_0+1)(k-1)^2} + \frac{\alpha_i \alpha_j (\alpha_0 + 2) - \alpha_0(\alpha_i + \alpha_j)}{(\alpha_0+1)(k-1)}.$$

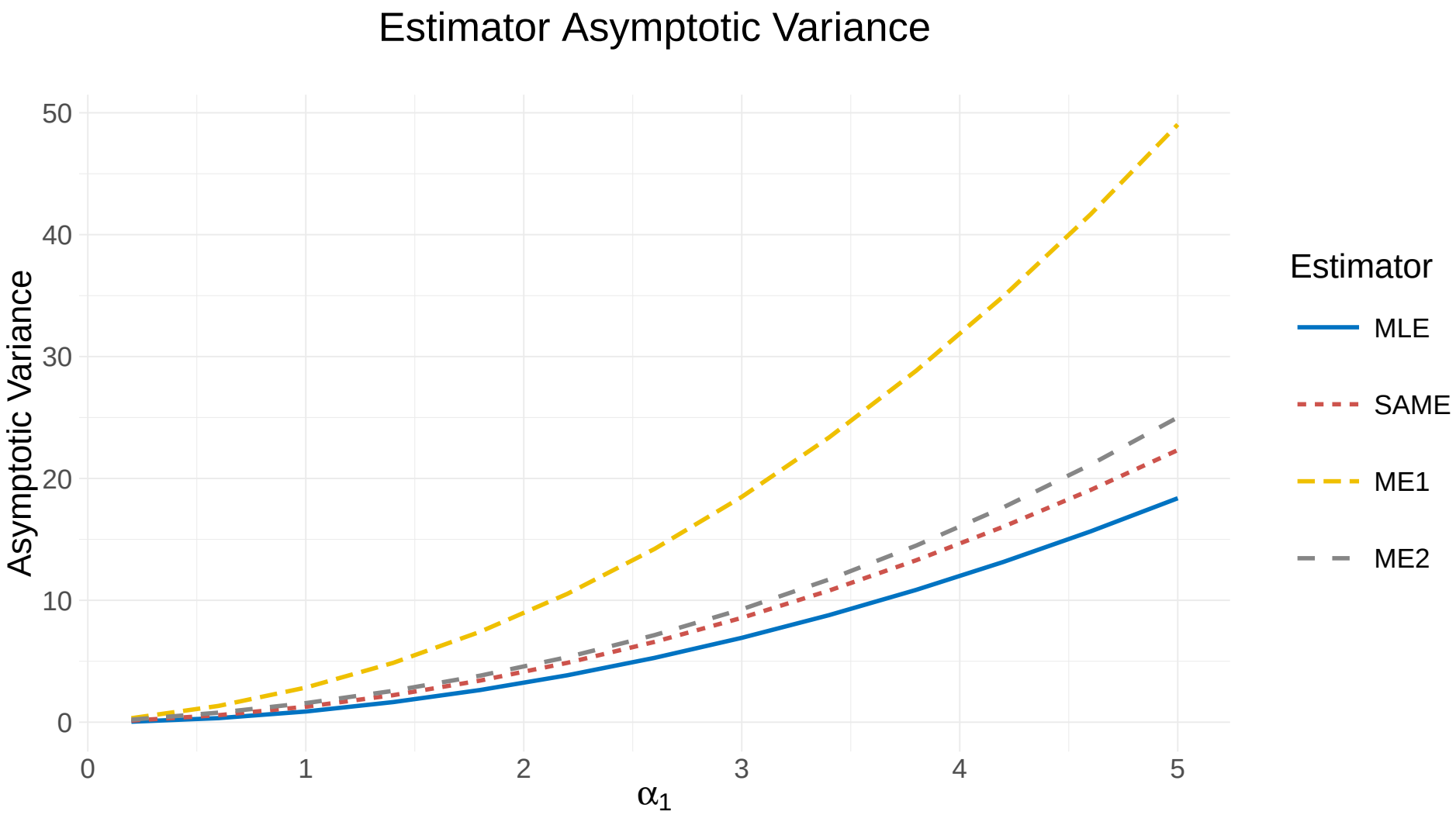
Asymptotic Variance of all α_i Estimators

$$\text{MLE} \quad \operatorname{Avar}\left(\widehat{\alpha}_i\right) = \frac{1}{\psi_1(\alpha_i)} + \frac{\psi_1(\alpha_0)}{\psi_1^2(\alpha_i)\left\{1 - \psi_1(\alpha_0)\sum_{m=1}^{k}\psi_1^{-1}(\alpha_m)\right\}},$$

$$\text{SAME} \quad \operatorname{Avar}\left(\breve{\alpha}_i\right) = \alpha_i^2 \sum_{m=1}^{k} \frac{\alpha_m(\alpha_0 - \alpha_m)\psi_1(\alpha_m)}{(\alpha_0+1)(k-1)^2} + \frac{\alpha_i^2(\alpha_0+2) - 2\alpha_0\alpha_i}{(\alpha_0+1)(k-1)} + \frac{\alpha_0\alpha_i}{\alpha_0+1},$$

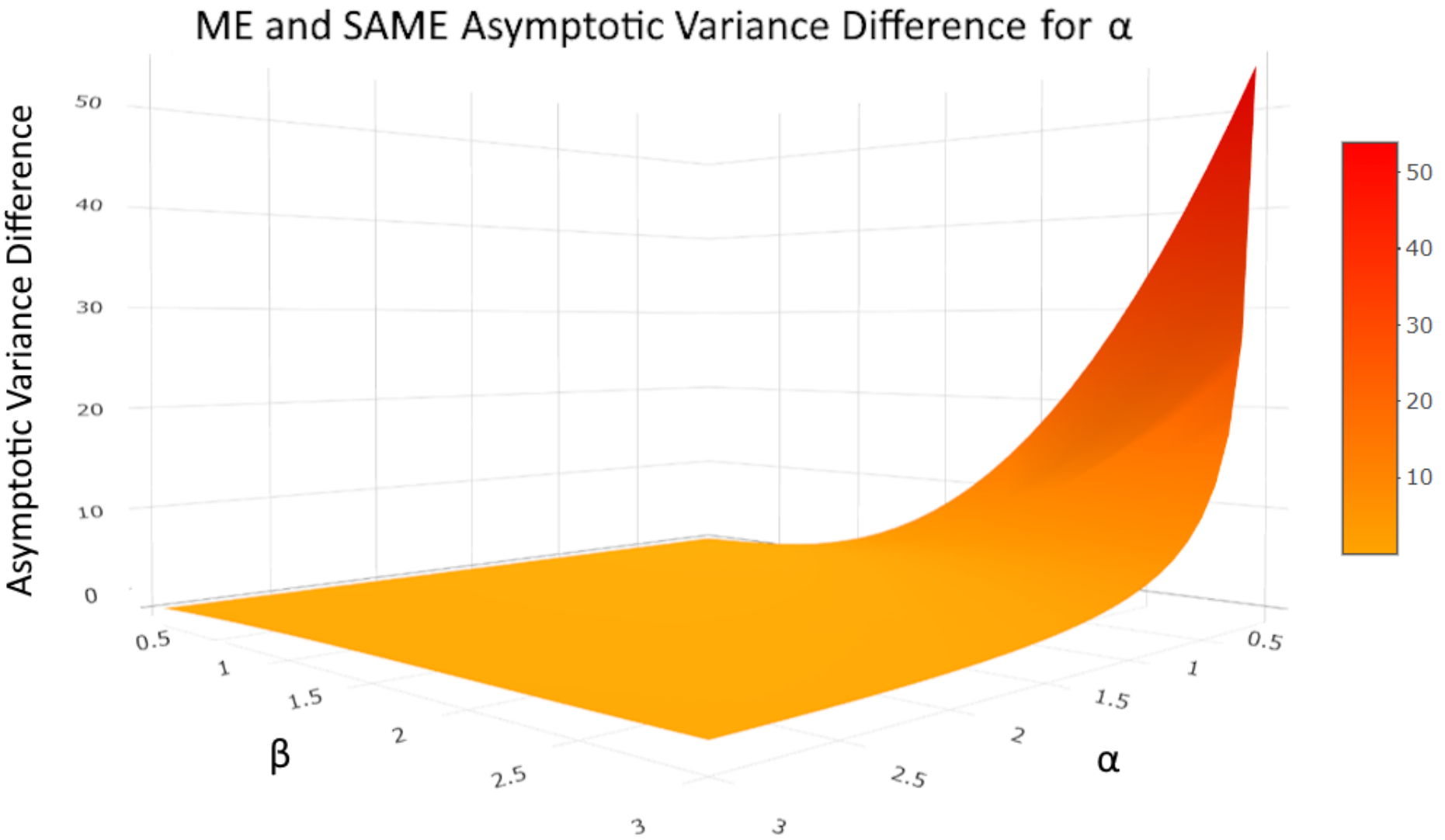
$$\text{ME}_1 \quad \operatorname{Avar}\left(\widetilde{\alpha}_i\right) = \frac{\alpha_0\alpha_i(\alpha_i+1)^2}{(\alpha_0-\alpha_i)(\alpha_0+2)(\alpha_0+3)}\left\{\frac{2(\alpha_0+1)^2}{\alpha_i+1} - \frac{2\alpha_0^2+\alpha_0+1}{\alpha_0+1}\right\},$$

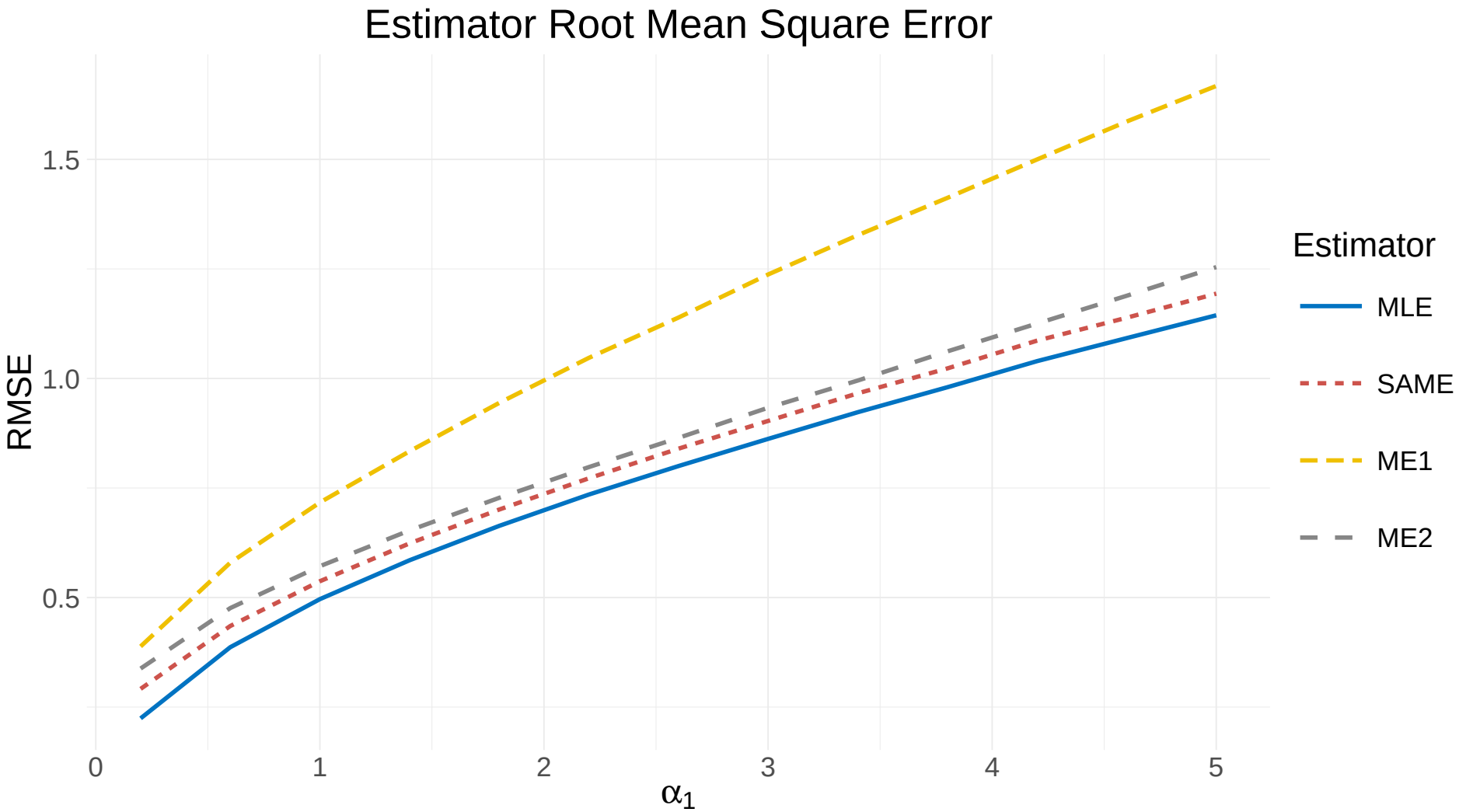
$$\text{ME}_2 \quad \operatorname{Avar}\left(\widetilde{\widetilde{\alpha}}_i\right) = \frac{\alpha_0\alpha_i}{\alpha_0+1} + \frac{4\alpha_0\alpha_i^3}{b(\boldsymbol{\alpha})(\alpha_0+2)} + \frac{\alpha_0\alpha_i^2 p(\boldsymbol{\alpha})}{b^2(\boldsymbol{\alpha})(\alpha_0+1)(\alpha_0+2)(\alpha_0+3)}.$$



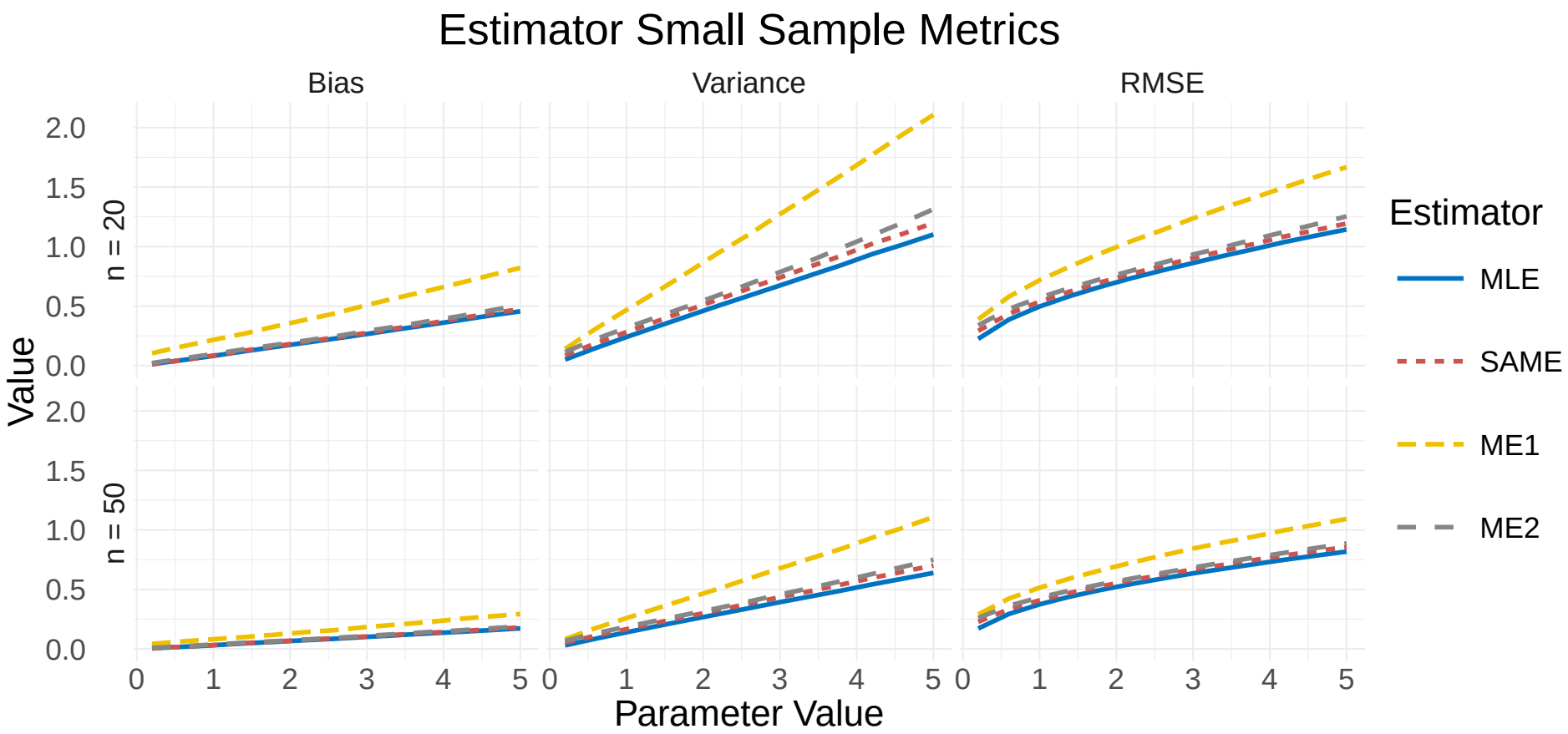
Asymptotic variance of the $\alpha_1 \in [0.2, 5]$ parameter estimators in $\mathcal{D}(\alpha_1, 0.2, 1, 2, 5)$.

$\mathrm{Avar}_{\boldsymbol{\alpha}}(\tilde{\alpha}) - \mathrm{Avar}_{\boldsymbol{\alpha}}(\breve{\alpha})$

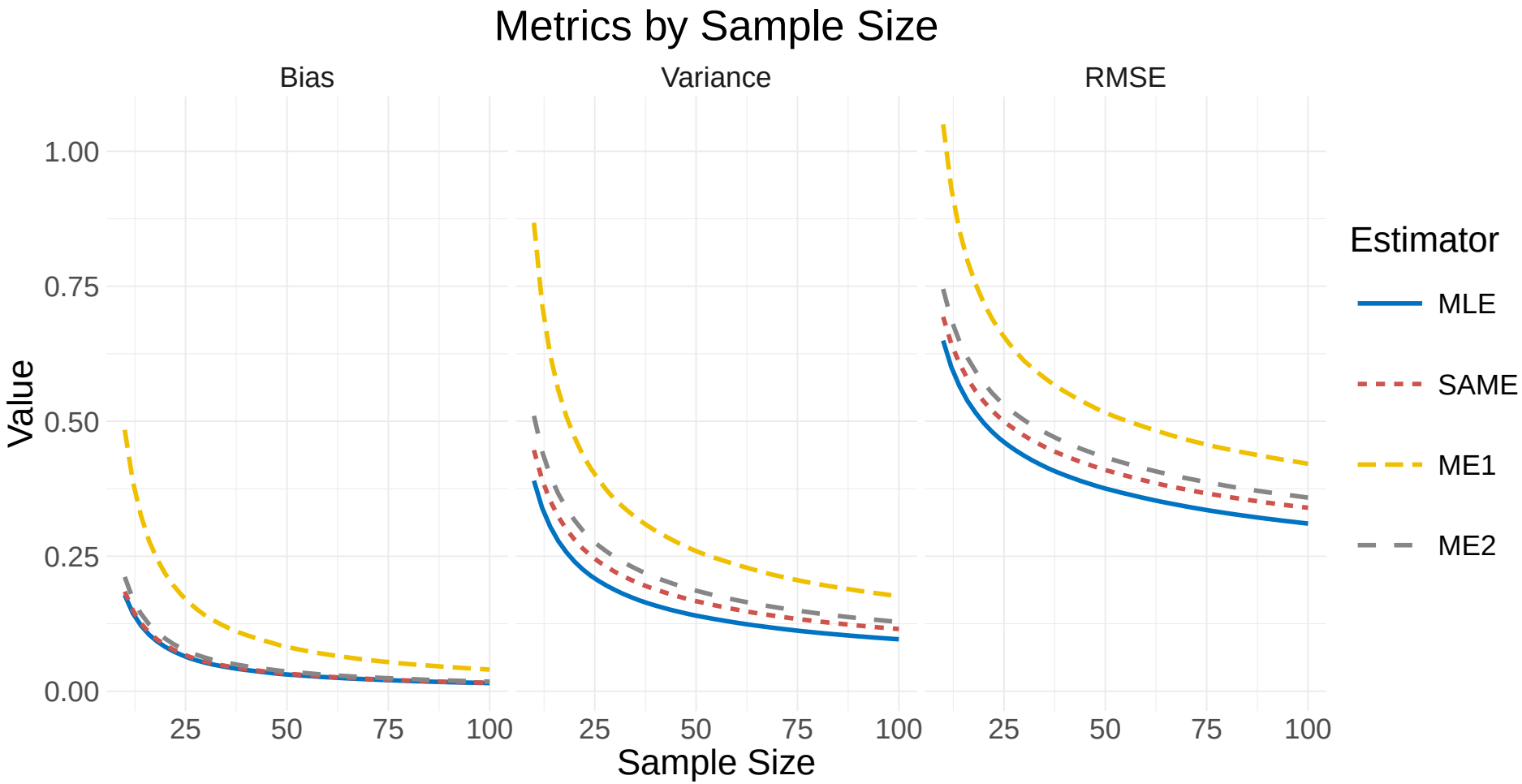




Root Mean Square Error of the $\alpha_1 \in [0.2, 5]$ parameter estimators in $\mathcal{D}(\alpha_1, 0.2, 1, 2, 5)$ with $n = 20$ observations. Estimated with $m = 10^5$ Monte Carlo samples.



Small sample metrics of the $\alpha_1 \in [0.2, 5]$ parameter estimators in $\mathcal{D}(\alpha_1, 0.2, 1, 2, 5)$ with $n = 20$ observations. Estimated with $m = 10^5$ Monte Carlo samples.



Small sample metrics of the α_1 parameter estimators in $\mathcal{D}(0.2, 0.2, 1, 2, 5)$ as a function of sample size. Estimated with $m = 10^5$ Monte Carlo samples.

Large Dimension Comparison

Proposition

Let $\{\alpha_m\}_{m\in\mathbb{N}}$ be a sequence of positive real numbers s.t. $\sum_{m=1}^{k}\alpha_m^{-1} = o\left(k^2\right)$. Then,

$$\lim_{k\to+\infty}\operatorname{Avar}\left(\widetilde{\alpha}_i\right) = 2\alpha_i(\alpha_i+1) > \alpha_i = \lim_{k\to+\infty}\operatorname{Avar}\left(\breve{\alpha}_i\right).$$

Large Dimension Comparison

Proposition

Let $\{\alpha_m\}_{m\in\mathbb{N}}$ be a sequence of positive real numbers s.t. $\sum_{m=1}^{k}\alpha_m^{-1} = o\left(k^2\right)$. Then,

$$\lim_{k\to+\infty}\operatorname{Avar}\left(\widetilde{\alpha}_i\right) = 2\alpha_i(\alpha_i+1) > \alpha_i = \lim_{k\to+\infty}\operatorname{Avar}\left(\breve{\alpha}_i\right).$$

Proof.

By the proposition assumption, it is immediate that $\sum_{m=1}^{\infty}\alpha_m = +\infty$.

$$\operatorname{Avar}\left(\widetilde{\alpha}_i\right) = \frac{\alpha_0\alpha_i(\alpha_i+1)^2}{(\alpha_0-\alpha_i)(\alpha_0+2)(\alpha_0+3)}\left\{\frac{2(\alpha_0+1)^2}{\alpha_i+1} - \frac{2\alpha_0^2+\alpha_0+1}{\alpha_0+1}\right\},$$

$$\operatorname{Avar}\left(\breve{\alpha}_i\right) = \alpha_i^2\sum_{m=1}^{k}\frac{\alpha_m(\alpha_0-\alpha_m)\psi_1(\alpha_m)}{(\alpha_0+1)(k-1)^2} + \frac{\alpha_i^2(\alpha_0+2)-2\alpha_0\alpha_i}{(\alpha_0+1)(k-1)} + \frac{\alpha_0\alpha_i}{\alpha_0+1}.$$

Large Dimension Comparison - Proof

Proof.

Therefore, it suffices to show that:

$$h_k(\boldsymbol{\alpha}) := \frac{\alpha_i^2}{(\alpha_0+1)(k-1)^2} \sum_{m=1}^{k} \alpha_m(\alpha_0-\alpha_m)\psi_1(\alpha_m) \xrightarrow{k\to+\infty} 0.$$

Utilizing the trigamma function bound $\psi_1(x) \leq (x+1)/x^2$, $x>0$ yields

$$0 \leq h_k(\boldsymbol{\alpha}) \leq \frac{\alpha_0\alpha_i^2}{(\alpha_0+1)(k-1)^2} \left\{ \sum_{m=1}^{k} \frac{1}{\alpha_m} + (k-1) - \frac{k}{\alpha_0} \right\}.$$

By the assumption $\sum_{m=1}^{k} \alpha_m^{-1} = o\left(k^2\right)$, the right hand side converges to 0, and the result follows by the squeeze theorem. □

Score-Adjusted Moment Estimators for the Dirichlet Distribution Family

3. Work in Progress

Dirichlet Moment Estimators

Proposition

The asymptotic variance of ME_2 is uniformly smaller than the one of ME_1, i.e.

$$\mathrm{Avar}\left(\widetilde{\alpha}_i\right) > \mathrm{Avar}\left(\widetilde{\widetilde{\alpha}}_i\right), \quad \forall \boldsymbol{\alpha} \in \mathbb{R}_+^k,\ k \geq 3.$$

Proof.

We can rewrite $\mathrm{Avar}\left(\widetilde{\alpha}_i\right) - \mathrm{Avar}\left(\widetilde{\widetilde{\alpha}}_i\right) = \pi_1(\boldsymbol{\alpha})\pi_2(\boldsymbol{\alpha})$, where

$$\pi_1(\boldsymbol{\alpha}) = \frac{\alpha_0\alpha_i}{b(\boldsymbol{\alpha})(\alpha_0+2)(\alpha_0+3)(\alpha_0-\alpha_i)} > 0, \quad \forall \boldsymbol{\alpha} \in \mathbb{R}_+^k,\ k \geq 2,$$

$$\begin{aligned}\pi_2(\boldsymbol{\alpha}) = {} & b^2(\boldsymbol{\alpha})\left[2(\alpha_0+3)\alpha_i + (1-\alpha_0)^2\right] + 4\alpha_i(\alpha_0-\alpha_i)(1-\alpha_0)(1+\alpha_0)c(\boldsymbol{\alpha}) \\ & - 2b(\boldsymbol{\alpha})(\alpha_0-\alpha_i)\alpha_i\left[2(\alpha_0+3)\alpha_i + (1-\alpha_0)^2 + 3\alpha_0(1-\alpha_0)(1+\alpha_0)\right].\end{aligned}$$

$\pi_2(\boldsymbol{\alpha})$ can be expressed as a 4th degree polynomial of α_i.

Dirichlet Moment Estimators

Proof.

$$\pi_2(\boldsymbol{\alpha}) = c_4(\boldsymbol{\alpha}_{-i})\alpha_i^4 + c_3(\boldsymbol{\alpha}_{-i})\alpha_i^3 + c_2(\boldsymbol{\alpha}_{-i})\alpha_i^2 + c_1(\boldsymbol{\alpha}_{-i})\alpha_i + c_0(\boldsymbol{\alpha}_{-i}),$$

with positive coefficients:

$$c_4(\boldsymbol{\alpha}) = 6\alpha_0 b(\boldsymbol{\alpha}),$$

$$c_3(\boldsymbol{\alpha}) = 2\alpha_0 \left[d(\boldsymbol{\alpha}) + 3(2\alpha_0 + 1)b(\boldsymbol{\alpha})\right],$$

$$c_2(\boldsymbol{\alpha}) = (6\alpha_0^3 + 8\alpha_0^2 + 2\alpha_0)b(\boldsymbol{\alpha}) + 3b^2(\boldsymbol{\alpha}) + 4\alpha_0^2 d(\boldsymbol{\alpha}),$$

$$c_1(\boldsymbol{\alpha}) = 2\left[2(\alpha_0 + 1)b^2(\boldsymbol{\alpha}) + \alpha_0 b(\boldsymbol{\alpha})(1 - \alpha_0)^2 - \alpha_0(1 - \alpha_0^2)d(\boldsymbol{\alpha})\right],$$

$$c_0(\boldsymbol{\alpha}) = (1 - \alpha_0)^2 b^2(\boldsymbol{\alpha}),$$

where $b(\boldsymbol{\alpha}) = \alpha_0^2 - \sum_{m=1}^k \alpha_m^2$, $c(\boldsymbol{\alpha}) = \alpha_0^3 - \sum_{m=1}^k \alpha_m^3$, $d(\boldsymbol{\alpha}) = 3\alpha_0 b(\boldsymbol{\alpha}) - 2c(\boldsymbol{\alpha})$.

The `joker` R Package

README | Code of conduct | GPL-3.0 license

joker

CRAN 0.14.2 | R-CMD-check.yaml passing | codecov 90% | pkgcheck passing | PASSED

Introduction

The `joker` R package develops an S4 distribution system and performs parameter estimation in common distribution families, making well-established and state-of-the-art methods more accessible.

Key Features

1. The common d, p, q, r function family for each distribution (e.g. dnorm, pnorm, qnorm, rnorm) is enriched with

- the ll counterpart (e.g. llnorm) that calculates the log-likelihood,
- the e counterpart (e.g. enorm) that performs parameter estimation,

Comparison of `stats` and `joker`

Distribution Handling

Math	**`stats`**
$X \sim \mathcal{B}(a, b)$	
Density Function $f(x)$	`dbeta(x, a, b)`
Cum. Distr. Function $F(x)$	`pbeta(x, a, b)`
Gen. Inverse CDF $F^{-1}(x)$	`qbeta(x, a, b)`
Simulation Function	`rbeta(n, a, b)`

Comparison of `stats` and `joker`

Distribution Handling

Math	**`stats`**	**`joker`**
$X \sim \mathcal{B}(a, b)$		`D <- Beta(a, b)`
Density Function $f(x)$	`dbeta(x, a, b)`	`d(D, x)`
Cum. Distr. Function $F(x)$	`pbeta(x, a, b)`	`p(D, x)`
Gen. Inverse CDF $F^{-1}(x)$	`qbeta(x, a, b)`	`qn(D, x)`
Simulation Function	`rbeta(n, a, b)`	`r(D, n)`

Random Variable Calculations

This OOP approach allows calculations with random variables, as shown in the following examples:

Location - Scale Transformation

If $X \sim \mathcal{N}(2, 1)$, then $Y := 3X + 1 \sim \mathcal{N}(7, 9)$.

```
x <- Norm(2, 1)
y <- 3 * x + 1
```

Sum of Normal Random Variables

If $X_1 \sim \mathcal{N}(1, 1), X_2 \sim \mathcal{N}(2, 1)$, and $X_1 \perp\!\!\!\perp X_2$ then $X_1 + X_2 \sim \mathcal{N}(3, 2)$.

```
x1 <- Norm(1, 1)
x2 <- Norm(2, 1)
x1 + x2
```

New Functionalities

Parametric Functions

```
mean(D)
median(D)
mode(D)
var(D)
sd(D)
skew(D)
kurt(D)
entro(D)
finf(D)
```

New Functionalities

Parametric Functions

```
mean(D)
median(D)
mode(D)
var(D)
sd(D)
skew(D)
kurt(D)
entro(D)
finf(D)
```

Log-likelihood

```
llbeta(x, a, b)
ll(D, x)
```

New Functionalities

Parametric Functions

```
mean(D)
median(D)
mode(D)
var(D)
sd(D)
skew(D)
kurt(D)
entro(D)
finf(D)
```

Log-likelihood

```
llbeta(x, a, b)
ll(D, x)
```

Parameter Estimation

```
ebeta(x, a, b, type = "mle")
e(D, x, type = "mle")
mle(D, x)
```

New Functionalities

Parametric Functions

```
mean(D)
median(D)
mode(D)
var(D)
sd(D)
skew(D)
kurt(D)
entro(D)
finf(D)
```

Log-likelihood

```
llbeta(x, a, b)
ll(D, x)
```

Parameter Estimation

```
ebeta(x, a, b, type = "mle")
e(D, x, type = "mle")
mle(D, x)
```

Asymptotic Variance

```
vbeta(x, a, b, type = "mle")
v(D, x, type = "mle")
avar_mle(D, x)
```

Distribution Families Coverage

Overview of the distributions implemented in the `joker` package, along with their respective class names.

Distribution	Class Name	Distribution	Class Name
Bernoulli	`Bern`	Geometric	`Geom`
Beta	`Beta`	Laplace	`Laplace`
Binomial	`Binom`	Log-Normal	`Lnorm`
Categorical	`Cat`	Multivariate Gamma	`Multigam`
Cauchy	`Cauchy`	Multinomial	`Multinom`
Chi-Square	`Chisq`	Negative Binomial	`Nbinom`
Dirichlet	`Dir`	Normal	`Norm`
Exponential	`Exp`	Poisson	`Pois`
Fisher	`Fisher`	Student	`Stud`
Gamma	`Gam`	Uniform	`Unif`
Generalized Gamma	`Gengam`	Weibull	`Weib`

Package Reliability and Perspectives

- The package utilizes continuous integration, ensuring consistent performance and reducing the likelihood of bugs in production.
- More than 1000 automated tests are integrated in the package.
- Additional standards are incorporated in the package to achieve the rOpenSci golden badge.
- The early release version of the package (0.8.5) had about 7000 CRAN downloads.
- The package will promptly be submitted to the Journal of Statistical Software.

Score-Adjusted Moment Estimators for the Dirichlet Distribution Family

Thank you for your attention!

Appendix