Asymptotic analysis of parameter estimation for Ewens–Pitman partition

Takuya Koriyama

University of Chicago

September 4, 2024

Table of content

Partition of integers

$\{U_1, U_2, \dots, U_k\}$ is said to be a partition of $[n] = \{1, \dots n\}$ into k blocks if

$$U_i \neq \emptyset, \quad U_i \cap U_j = \emptyset, \quad \cup_{i=1}^{k} U_k = [n].$$

Letting $\mathcal{P}_{n,k}$ be the set of partitions of $[n]$ into k, define $\mathcal{P}_n$ as

$$\mathcal{P}_n = \cup_{k=1}^{n} \mathcal{P}_{n,k}.$$

Example

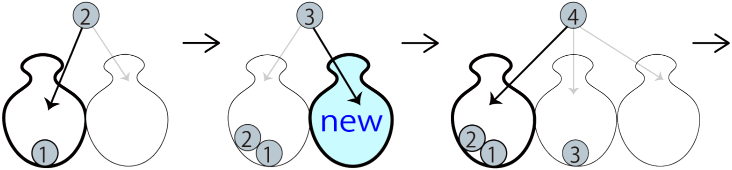
$$\begin{aligned}
&\mathcal{P}_1 = \Big\{ \big\{\{1\}\big\} \Big\}, \qquad \mathcal{P}_2 = \Big\{ \big\{\{1,2\}\big\},\ \big\{\{1\},\{2\}\big\} \Big\}, \\
&\mathcal{P}_3 = \Big\{ \big\{\{1,2,3\}\big\}, \\
&\qquad \big\{\{1,2\},\{3\}\big\},\ \big\{\{2,3\},\{1\}\big\},\ \big\{\{1,3\},\{2\}\big\}, \\
&\qquad \big\{\{1\},\{2\},\{3\}\big\} \Big\}.
\end{aligned}$$

We denote the element of $\mathcal{P}_n$ by Π_n.

Sequential partitions of integers

Starting from $\Pi_1 = \{\{1\}\}$, we consider a sequence of partitions.

$$\Pi_2 = \big\{\{1,2\}\big\} \to \Pi_3 = \big\{\{1,2\},\{3\}\big\} \to \Pi_4 = \big\{\{1,2,4\},\{3\}\big\} \to$$



Ewens–Pitman partition is a stochastic process over the sets of partitions $(\mathcal{P}_n)_{n=1}^{\infty}$.

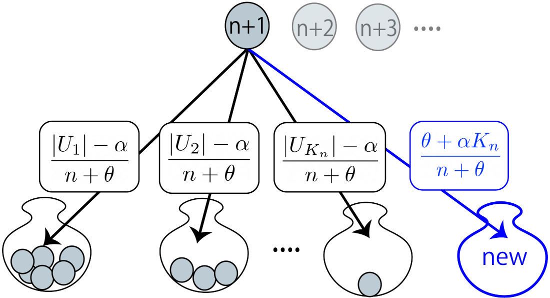
Ewens–Pitman partition (α, θ)

Ewens–Pitman partition is a stochastic process over $(\mathcal{P}_n)_{n=1}^{\infty}$.

- $\Pi_1 = \{1\}$.
- Given $\Pi_n \in \mathcal{P}_n$, letting K_n the number of blocks in Π_n, $(n+1)$-th ball is assigned into the existing blocks $\{U_1, \dots, U_{K_n}\}$ or an empty blocks according to

$$U_i \text{ for } i \in \{1, \dots, K_n\} \qquad \text{with probability } \frac{|U_i| - \alpha}{n + \theta}$$

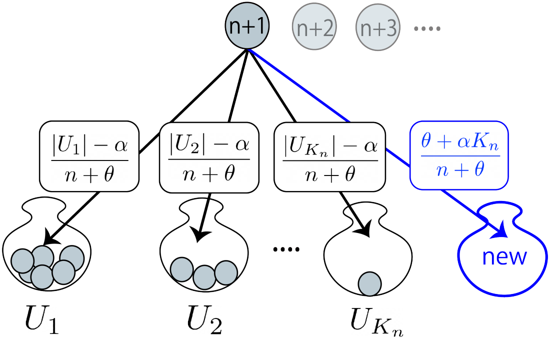
$$\text{Empty block} \qquad \text{with probability } \frac{\theta + \alpha K_n}{n + \theta}$$



Example: $n = 1$

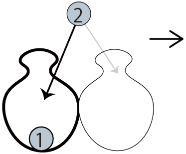
Let us start with $\Pi_1 = (\{1\})$. Then

$$\text{2nd ball belongs to} \begin{cases} \{1\} & \text{with prob. } (1-\alpha)/(1+\theta) \\ \text{new urn} & \text{with prob. } (\theta+\alpha)/(1+\theta). \end{cases}$$



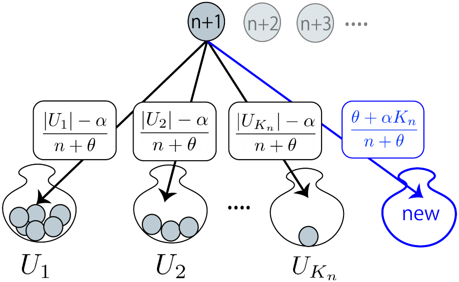
Example: $n = 2$

Suppose 2 was assigned to $\{1\}$



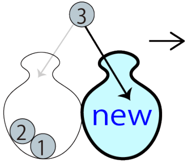
so that $\Pi_2 = \{\{1,2\}\}$. Then

$$\text{3rd ball belongs to} \begin{cases} \{1,2\} & \text{with prob. } (2-\alpha)/(2+\theta) \\ \text{new urn} & \text{with prob. } (\theta+\alpha)/(2+\theta) \end{cases}$$



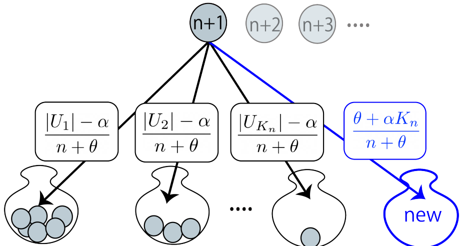
Example $n = 3$

Suppose 3 was assigned to a new block



so that $\Pi_3 = \{\{1,2\},\{3\}\}$. Then

$$\text{4th ball belongs to} \begin{cases} U_1 = \{1,2\} & \text{with prob. } (2-\alpha)/(3+\theta) \\ U_2 = \{3\} & \text{with prob. } (1-\alpha)/(3+\theta) \\ \text{new urn} & \text{with prob. } (\theta+2\alpha)/(3+\theta) \end{cases}$$



Asymptotics of Ewens–Pitman partition as $n \to \infty$

For the partition $\Pi_n = (U_1, U_2, \dots,)$, we define

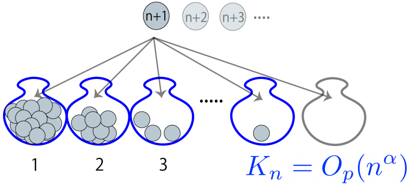
$$S_{n,j} := \sum_{i \geq 1} \mathbf{1}\{|U_i| = j\} \quad \text{"Number of urns of size } j\text{"}$$

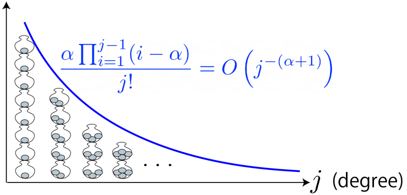
$$K_n := \sum_{j \geq 1} S_{n,j} \quad \text{"Number of non empty urns"}.$$

e.g.) $\Pi_4 = (\{1,4\},\{2\},\{3\}) \Rightarrow S_{4,1} = 2,\ S_{4,2} = 1,\ S_{4,j} = 0\ (\forall j \geq 3)$.

Theorem (Asymptotics when $0 < \alpha < 1, \theta > -\alpha$) [Pit06]

1. $n^{-\alpha} K_n \xrightarrow{\text{a.s.}} M_{\alpha\theta}$, where $M_{\alpha\theta}$ is a non degenerate random variable.
2. $\forall j \in \mathbb{N},\ S_{n,j}/K_n \xrightarrow{\text{a.s.}} p_\alpha(j) := \frac{\alpha \prod_{i=1}^{j-1}(i-\alpha)}{j!} = O(j^{-(\alpha+1)})$.

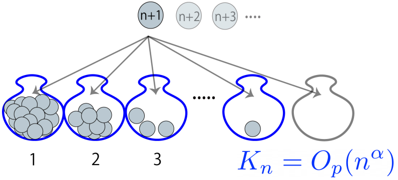


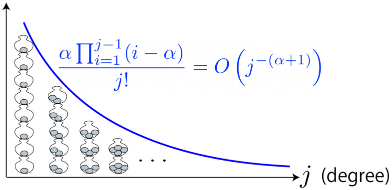


Application

Estimation of α is of particular interest.

	K_n	$S_{n,j}$
Ecology[1]	Species	Species j times observed
Network Analysis[2]	Vertices	Vertices with j edges





[1] [BFN22, FN21, FLMP09, Sib14], [FPR21, Hos01], [CCV22]
[2] [CD18, NRC21]

Connection to Nonparametric Bayesian Inference

- For (α, θ) and a non atomic measure F, Poisson Dirichlet prior $P = \mathrm{PD}(\alpha, \theta, F)$ is a discrete random measure defined by

$$\mathrm{PD}(\alpha, \theta, F) := \sum_{i=1}^{\infty} p_i \delta_{y_i}, \text{where}$$

$$y_i \stackrel{\text{iid}}{\sim} F \text{ and } p_i = v_i \prod_{j=1}^{i-1} (1 - v_j) \text{ with } v_i \sim \mathrm{Beta}(1 - \alpha, \theta + j\alpha).$$

- If $X_i | P \stackrel{\text{iid}}{\sim} P$, $(X_i)_{i=1}^n$ induces a partition Π_n by the equivalence relation $i \sim j$ iff $X_i = X_j$.

Theorem ([Pit06])

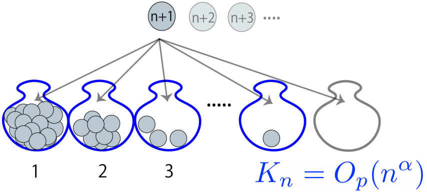
Π_n induced by conditional iid sample from PD(α, θ, F) has the same law as Π_n generated by Ewens–Pitman partition (α, θ).

- Estimation of (α, θ) is the hyper-parameter tuning of $\mathrm{PD}(\alpha, \theta, F)$

Naive Estimation of α

- Recall there exists a positive random variable $M_{\alpha\theta}$ s.t.

$$K_n/n^\alpha \xrightarrow{\text{a.s.}} M_{\alpha\theta}$$

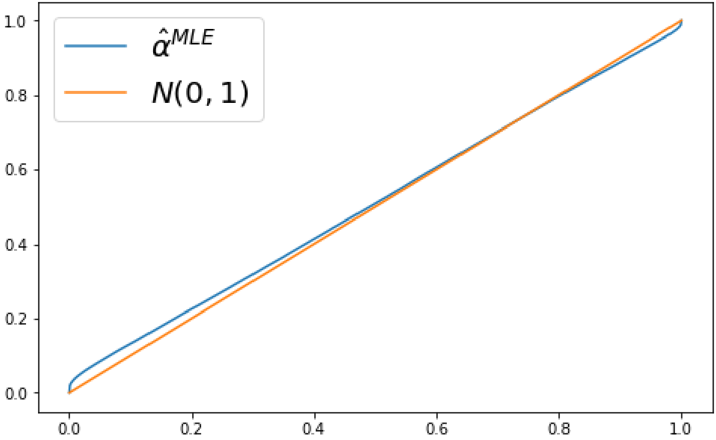


- Define $\hat{\alpha}_n^{\text{naive}} := \log K_n / \log n$. Then

$$\log n \cdot (\hat{\alpha}_n^{\text{naive}} - \alpha) = \log K_n - \alpha \log n = \log(K_n/n^\alpha) \xrightarrow{\text{a.s.}} \log M_{\alpha\theta}$$

- $\hat{\alpha}_n^{\text{naive}}$ is $\log n$-consistent.

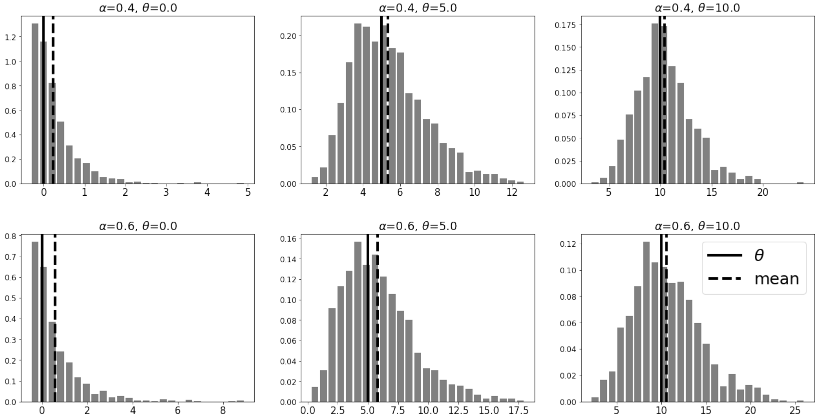
Maximum Likelihood Estimator of α



QQ plot of $\hat{\alpha}_n^{\text{MLE}}$ with $(\alpha, \theta) = (0.8, 0)$, $n = 2^{19}$, replicate $=10^5$

- $\hat{\alpha}_n^{\text{MLE}}$ is not asymptotically normal.

Maximum Likelihood Estimator of θ



Histogram of $\hat{\theta}_n^{\mathsf{MLE}}$ with $n = 2^{16}$ and replicate $= 1000$.

- $\hat{\theta}_n^{\mathsf{MLE}}$ does not concentrate on θ.
- Limit distribution is not normal.

Contribution

We derive the exact asymptotic distribution of $(\hat{\alpha}_n^{\mathsf{MLE}}, \hat{\theta}_n^{\mathsf{MLE}})$:

$$\sqrt{n^\alpha I_\alpha} \cdot (\hat{\alpha}_n^{\mathsf{MLE}} - \alpha) \to \mathcal{N}(0, {M_{\alpha\theta}}^{-1}), \quad \hat{\theta}_n^{\mathsf{MLE}} \to \alpha \cdot f_\alpha^{-1}(\log M_{\alpha\theta}),$$

from which we conclude

- $\hat{\alpha}_n^{\mathsf{MLE}}$ is $n^{\alpha/2}$-consistent, faster than the rate $\log n$ of $\hat{\alpha}_n^{\mathsf{naive}}$.
- $\mathcal{N}(0, {M_{\alpha\theta}}^{-1})$ is a variance mixture of centered normals due to the randomness of $M_{\alpha\theta}$
- $\hat{\theta}_n^{\mathsf{MLE}}$ is not consistent.

We also propose a confidence interval for α.

Table of content

Likelihood Formula

For the partition $\Pi_n = (U_1, U_2, \dots,)$, we define

$$S_{n,j} := \sum_{i \geq 1} \mathbf{1}\{|U_i| = j\} \qquad \text{"Number of urns of size } j\text{"}$$

$$K_n := \sum_{j \geq 1} S_{n,j} \qquad \text{"Number of non empty urns"}.$$

e.g.) $\Pi_4 = (\{1,4\},\{2\},\{3\}) \Rightarrow S_{4,1} = 2,\ S_{4,2} = 1,\ S_{4,j} = 0\ (\forall j \geq 3)$.

Theorem (Ewens–Pitman Sampling Formula) [Pit06]

Likelihood $\mathcal{L}(\Pi_n; \alpha, \theta)$ of Ewens–Pitman partition (α, θ) can be written by

$$\mathcal{L}(\Pi_n; \alpha, \theta) = \frac{\prod_{i=1}^{K_n - 1}(\theta + i\alpha)}{\prod_{i=1}^{n-1}(\theta + i)} \prod_{j=2}^{n} \left\{ \prod_{i=1}^{j-1} (-\alpha + i) \right\}^{S_{n,j}}.$$

Therefore, $(S_{n,j})_{j \geq 1}$ is sufficient statistic.

Asymptotic analysis of Fisher Information

Derive leading terms ($n \to \infty$) of Fisher Information defined as

$$I_{\alpha\alpha}^{(n)} := \mathbb{E}[-\partial_{\alpha\alpha}^2 \log \mathcal{L}(\Pi_n; \alpha, \theta)],$$
$$I_{\alpha\theta}^{(n)} := \mathbb{E}[-\partial_{\alpha\theta}^2 \log \mathcal{L}(\Pi_n; \alpha, \theta)],$$
$$I_{\theta\theta}^{(n)} := \mathbb{E}[-\partial_{\theta\theta}^2 \log \mathcal{L}(\Pi_n; \alpha, \theta)].$$

Useful notation:

- $I_\alpha :=$ Fisher Information of the distribution with pmf $p_\alpha(j)$, i.e.,

$$I_\alpha := -\sum_{j=1}^{\infty} p_\alpha(j) \cdot \partial_\alpha^2 \log p_\alpha(j) \text{ with } p_\alpha(j) = \frac{\alpha \prod_{i=1}^{j-1}(i-\alpha)}{j!}$$

- Function[3] $f_\alpha : (-1, \infty) \to \mathbb{R}$ defined by (don't need to memorize)

$$f_\alpha : z \mapsto \psi(1+z) - \alpha\psi(1+\alpha z) \text{ with } \psi(x) = \Gamma'(x)/\Gamma(x)$$

[3] f_α is bijective, strictly increasing, and convex.

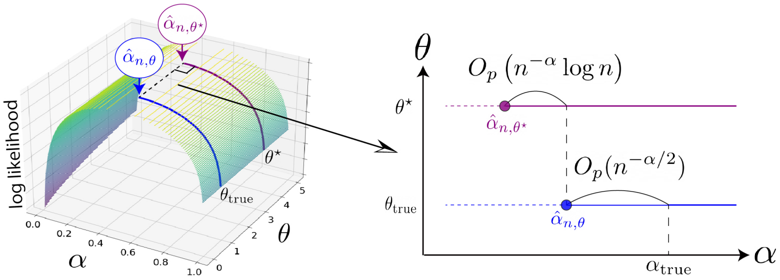
Asymptotic analysis of Fisher Information

Lemma (Leading terms of Fisher Information)

As $n \to \infty$, we have

$$I_{\alpha\alpha}^{(n)} \sim n^{\alpha} I_\alpha \, \mathbb{E}[M_{\alpha\theta}], \; I_{\theta\alpha}^{(n)} \sim \alpha^{-1} \log n, \; I_{\theta\theta}^{(n)} \to \alpha^{-2} f'_\alpha(\theta/\alpha) < +\infty.$$

- Non-identifiability of θ.
- The optimal convergence rate of estimators of α is $n^{-\alpha/2}$
- (α, θ) are asymptotically orthogonal.



Maximum Likelihood Estimator

Given a partition $\Pi_n = (U_1, U_2, \dots)$, define $(\hat{\alpha}_n^{\mathsf{MLE}}, \hat{\theta}_n^{\mathsf{MLE}})$ by

$$(\hat{\alpha}_n^{\mathsf{MLE}}, \hat{\theta}_n^{\mathsf{MLE}}) \in \underset{\alpha \in [\epsilon, 1-\epsilon], \theta > -\alpha}{\arg\max} \frac{\prod_{i=1}^{K_n - 1}(\theta + i\alpha)}{\prod_{i=1}^{n-1}(\theta + i)} \prod_{j=2}^{n} \left\{ \prod_{i=1}^{j-1}(-\alpha + i) \right\}^{S_{n,j}}$$

where $S_{n,j} = \sum_{i \geq 1} \mathbf{1}\{|U_i| = j\}$, and $K_n = \sum_{j \geq 1} S_{n,j}$.

Lemma (Existence and Uniqueness of MLE)

If $\alpha \in [\epsilon, 1-\epsilon]$, $(\hat{\alpha}_n^{MLE}, \hat{\theta}_n^{MLE})$ uniquely exists with high probability.

- Since $\{\alpha \in [\epsilon, 1-\epsilon], \theta > -\alpha\}$ is not compact, this is not obvious.
- We can relax ϵ to a slowly decreasing array.

Asymptotic distribution of the MLE

Theorem (when $0 < \alpha < 1, \theta > -\alpha$)

Let $M_{\alpha\theta} = \lim_{n\to\infty} K_n/n^\alpha$, which is a positive random variable. Then

$$\sqrt{n^\alpha I_\alpha} \cdot (\hat{\alpha}_n^{MLE} - \alpha) \xrightarrow{\text{stable}} \mathcal{N}(0, {M_{\alpha\theta}}^{-1}),$$
$$\hat{\theta}_n^{MLE} \xrightarrow{\text{p}} \alpha \cdot f_\alpha^{-1}(\log M_{\alpha\theta}),$$

where $I_\alpha = -\sum_{j=1}^\infty p_\alpha(j) \cdot \partial_\alpha^2 \log p_\alpha(j)$ with $p_\alpha(j) = \frac{\alpha \prod_{i=1}^{j-1}(i-\alpha)}{j!}$ and $f_\alpha(z) := \psi(1+z) - \alpha\psi(1+\alpha z)$ $(\forall z > -1)$

1. $\hat{\alpha}_n^{\text{MLE}}$ is $n^{\alpha/2}$-consistent, faster than the rate $\log n$ of $\hat{\alpha}_n^{\text{naive}}$.
2. $\mathcal{N}(0, {M_{\alpha\theta}}^{-1})$ is a variance mixture of centered normals. However we can construct a confidence interval for α
3. $\hat{\theta}_n^{\text{MLE}}$ is not consistent, and converges to a non-standard distribution.

Asymptotic mixed normality of $\hat{\alpha}_n$

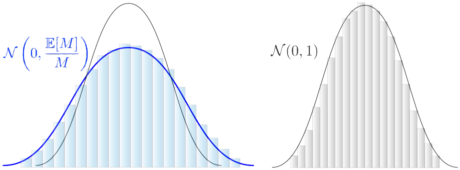
For $I^{(n)}_{\alpha\alpha} := \mathbb{E}[-\partial^2_{\alpha\alpha} \log \mathcal{L}(\Pi_n; \alpha, \theta)]$ and $\hat{\alpha}^{\mathsf{MLE}}_n$, we have shown

$$I^{(n)}_{\alpha\alpha} \sim n^{\alpha}\, \mathbb{E}[M_{\alpha\theta}] I_{\alpha}, \quad \sqrt{n^{\alpha} I_{\alpha}} \cdot (\hat{\alpha}^{\mathsf{MLE}}_n - \alpha) \to \mathcal{N}(0, {M_{\alpha\theta}}^{-1}),$$

which implies

$$\begin{aligned} \sqrt{I^{(n)}_{\alpha\alpha}} \cdot (\hat{\alpha}^{\mathsf{MLE}}_n - \alpha) &\sim \sqrt{n^{\alpha}\, \mathbb{E}[M_{\alpha\theta}] I_{\alpha}} \cdot (\hat{\alpha}^{\mathsf{MLE}}_n - \alpha) \\ &= \sqrt{\mathbb{E}[M_{\alpha\theta}]} \times \sqrt{n^{\alpha} I_{\alpha}} \cdot (\hat{\alpha}^{\mathsf{MLE}}_n - \alpha) \\ &\to \sqrt{\mathbb{E}[M_{\alpha\theta}]} \times \mathcal{N}(0, {M_{\alpha\theta}}^{-1}) \\ &= \mathcal{N}(0, \mathbb{E}[M_{\alpha\theta}]/M_{\alpha\theta}), \end{aligned}$$

where the variance of the normal is random (asymptotic mixed normality).



Confidence Interval for α

For the number of urns K_n and $\hat{\alpha}_n^{\mathsf{MLE}}$, it holds that

$$K_n/n^\alpha \xrightarrow{\text{a.s.}} M_{\alpha\theta}, \quad \sqrt{n^\alpha I_\alpha} \cdot (\hat{\alpha}_n^{\mathsf{MLE}} - \alpha) \xrightarrow{\mathsf{stable}} \mathcal{N}(0, {M_{\alpha\theta}}^{-1}),$$

which implies

$$\begin{aligned}\sqrt{K_n I_\alpha} \cdot (\hat{\alpha}_n^{\mathsf{MLE}} - \alpha) &= \sqrt{\frac{K_n}{n^\alpha}} \cdot \sqrt{n^\alpha I_\alpha}(\hat{\alpha}_n - \alpha) \\ &\to \sqrt{M_{\alpha\theta}} \cdot \mathcal{N}(0, {M_{\alpha\theta}}^{-1}) = \mathcal{N}(0,1),\end{aligned}$$

where the random variable $M_{\alpha\theta}$ is cancelled out.[4]

- Normalizing by K_n, $\hat{\alpha}_n^{\mathsf{MLE}}$ converges to normal distribution
- $[\hat{\alpha}_n^{\mathsf{MLE}} \pm 1.96/\sqrt{K_n I_{\hat{\alpha}_n}}]$ is 95% confidence interval for α.

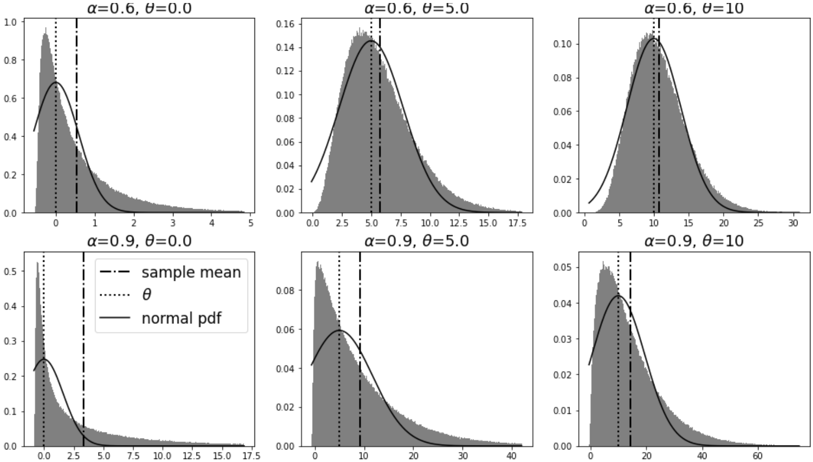
[4] We use an extended Slutzky's lemma for stable convergence.

Non-standard asymptotics of $\hat{\theta}_n$

For $I_{\theta\theta}^{(n)} := \mathbb{E}[-\partial_{\theta\theta}^2 \log \mathcal{L}(\Pi_n; \alpha, \theta)]$ and $\hat{\theta}_n^{\mathsf{MLE}}$, we have shown

$$I_{\theta\theta}^{(n)} \to \alpha^{-2} f_\alpha'(\theta/\alpha), \quad \hat{\theta}_n^{\mathsf{MLE}} \to \alpha \cdot f_\alpha^{-1}(\log M_{\alpha\theta}) \quad \text{(in probability)}$$

Compare $f_\alpha^{-1}(\log M_{\alpha\theta})$ and $\mathcal{N}(\theta, (\lim_{n\to\infty} I_{\theta\theta}^{(n)})^{-1}) = \mathcal{N}(\theta, \alpha^2 / f_\alpha'(\theta/\alpha))$.



Sketch of proof

- Asymptotically orthogonality of $(\alpha, \theta) \Rightarrow$ Coordinate-wise analysis
- Applying Martingale (Stable) CLT for log-likelihood
- Define the random/deterministic measure $\mathbb{P}_n/\mathbb{P}$ on $\mathbb{N}$ by

$$\forall j \in \mathbb{N},\ \mathbb{P}_n(j) := \frac{S_{n,j}}{K_n},\ \mathbb{P}(j) := \frac{\alpha \prod_{i=1}^{j-1}(i-\alpha)}{j!},$$

and, for suitable set of functions $\mathcal{F}$ on $\mathbb{N}$, show

$$\sup_{f \in \mathcal{F}} |\mathbb{E}_n f - \mathbb{E} f| \to^P 0.$$

Table of content

Summary

We derive the exact asymptotic distribution of $(\hat{\alpha}_n^{\mathsf{MLE}}, \hat{\theta}_n^{\mathsf{MLE}})$ as

$$\sqrt{n^\alpha I_\alpha} \cdot (\hat{\alpha}_n^{\mathsf{MLE}} - \alpha) \to \mathcal{N}(0, {M_{\alpha\theta}}^{-1}), \quad \hat{\theta}_n^{\mathsf{MLE}} \to \alpha \cdot f_\alpha^{-1}(\log M_{\alpha\theta}),$$

from which we conclude

- $\hat{\alpha}_n^{\mathsf{MLE}}$ is $n^{\alpha/2}$-consistent, faster than the rate $\log n$ of $\hat{\alpha}_n^{\mathsf{naive}}$.
- $\hat{\alpha}_n^{\mathsf{MLE}}$ is asymptotically mixed normal due to the randomness of $M_{\alpha\theta}$.
- $\sqrt{K_n I_\alpha} \cdot (\hat{\alpha}_n^{\mathsf{MLE}} - \alpha) \to \mathcal{N}(0, 1)$, which leads to confidence interval.
- $\hat{\theta}_n^{\mathsf{MLE}}$ is not consistent, and the limit distribution is positively skewed.

Future research 1: Hypothesis testing of $\alpha = 0$ or not

- We showed

$$\forall \alpha \in (0,1), \ \sqrt{K_n I_\alpha} \cdot (\hat{\alpha}_n^{\mathsf{MLE}} - \alpha) \to \mathcal{N}(0,1)$$

We can test $H_0 : \alpha < \alpha_0$ vs $H_1 : \alpha_0 < \alpha < 1$ for $\alpha_0 \in (0,1)$.

- There is a transition at $\alpha = 0$.

	$\alpha = 0$	$0 < \alpha < 1$
K_n	$O_p(\log n)$	$O_p(n^\alpha)$
MLE $\hat{\theta}_n$	Consistent	Inconsistent
Nonparametric Bayes	Dirichlet prior	Poisson Dirichlet prior
Network data	Dense	Sparse
⋮	⋮	⋮

- How to test $H_0 : \alpha = 0$ vs $H_1 : 0 < \alpha < 1$?

Testing of $H_0 : \alpha = 0$ vs $H_1 : 0 < \alpha < 1$

	$\alpha = 0$	$\alpha = 1/\log\log n$	$0 < \alpha < 1$
K_n	$O_p(\log n)$	$O_p(\frac{\log n^2}{\log\log n})$	$O_p(n^\alpha)$
Limit of $S_{n,j}$	$\to^P$ Poisson (θ/j)	?	$\sim K_n p_\alpha(j)$
$\hat{\theta}_n^{\text{MLE}}$	Consistent	?	Inconsistent
$\hat{\alpha}_n^{\text{MLE}}$		?	$n^{\alpha/2}$-consistent

We can think of

$$H_0 : \alpha = 0, \; H_1 : \alpha = 1/\log\log n$$

and find some criteria R_n and law F such that $R_n \to F$ under H_1.

Future research 2: Prediction of unseen

- For $m \in \mathbb{N}$, predict the law $\mathbb{P}^{n,m}_{\alpha,\theta}$ of $K_{n+m} - K_n$ given a partition Π_n of $[n]$. For example, if $m = 1$,

$$\mathbb{P}^{n,1}_{\alpha,\theta}(1) = \Pr(K_{n+1} - K_n = 1 | \Pi_n) = \frac{\theta + \alpha K_n}{n + \theta}$$

- Plug-in/Bayesian predictive distribution $\mathbb{P}^{n,m}_{\mathsf{MLE}}/\mathbb{P}^{n,m}_{\pi}$ is

$$\mathbb{P}^{n,m}_{\mathsf{MLE}}(\cdot) := \mathbb{P}^{n,m}_{\hat{\alpha}^{\mathsf{MLE}}_n, \hat{\theta}^{\mathsf{MLE}}_n}(\cdot), \quad \mathbb{P}^{n,m}_{\pi}(\cdot) := \int_{\alpha,\theta} \mathbb{P}^{n,m}_{\alpha,\theta}(\cdot) d\pi(\alpha, \theta | \Pi_n).$$

- Compare Plug-in/Bayesian risk $R^{n,m}_{\mathsf{MLE}}/R^{n,m}_{\pi}$, defined by

$$R^{n,m}_{\mathsf{MLE}} := \mathbb{E}^{n}_{\alpha,\theta}\left[\mathsf{KL}\left(\mathbb{P}^{n,m}_{\alpha,\theta} \mid\mid \mathbb{P}^{n,m}_{\mathsf{MLE}}\right)\right], R^{n,m}_{\pi} := \mathbb{E}^{n}_{\alpha,\theta}\left[\mathsf{KL}\left(\mathbb{P}^{n,m}_{\alpha,\theta} \mid\mid \mathbb{P}^{n,m}_{\pi}\right)\right]$$

  Existing works ([FN21, FLMP09]) use $\mathbb{P}^{n,m}_{\mathsf{MLE}}$, but we expect $R^{n,m}_{\mathsf{MLE}} \gtrsim R^{n,m}_{\pi}$ in a regime like $m \gtrsim n$.
- Require BvM, asymptotic expansion of KL, Ibragimov–Has'minski Theory [IHM13], etc.

Reference I

[BFN22] Cecilia Balocchi, Stefano Favaro, and Zacharie Naulet.
Bayesian nonparametric inference for" species-sampling" problems.
arXiv preprint arXiv:2203.06076, 2022.

[CCV22] Giulia Cereda, Fabio Corradi, and Cecilia Viscardi.
Learning the two parameters of the poisson–dirichlet distribution with a forensic application.
Scandinavian Journal of Statistics, pages 1–22, 2022.

[CD18] Harry Crane and Walter Dempsey.
Edge Exchangeable Models for Interaction Networks.
Journal of the American Statistical Association, 113(523): 1311–1326, 2018.

[FLMP09] Stefano Favaro, Antonio Lijoi, Ramsés H Mena, and Igor Prünster.
Bayesian non-parametric inference for species variety with a two-parameter poisson–dirichlet process prior.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71(5): 993–1008, 2009.

[FN21] Stefano Favaro and Zacharie Naulet.
Near-optimal estimation of the unseen under regularly varying tail populations.
arXiv preprint arXiv:2104.03251, 2021.

Reference II

[FPR21] Stefano Favaro, Francesca Panero, and Tommaso Rigon.
Bayesian nonparametric disclosure risk assessment.
Electronic Journal of Statistics, 15(2): 5626–5651, 2021.

[HL15] Erich Häusler and Harald Luschgy.
Stable Convergence and Stable Limit Theorems.
Springer, 2015.

[Hos01] Nobuaki Hoshino.
Applying Pitman's sampling formula to microdata disclosure risk assessment.
Journal of Official Statistics, 17(4): 499–520, 2001.

[IHM13] Ildar Abdulovich Ibragimov and Rafail Zalmanovich Has' Minskii.
Statistical estimation: asymptotic theory, volume 16.
Springer Science & Business Media, 2013.

[NRC21] Zacharie Naulet, Judith Rousseau, and François Caron.
Asymptotic analysis of statistical estimators related to multigraphex processes under misspecification.
arXiv preprint arXiv:2107.01120, 2021.

[Pit06] Jim Pitman.
Combinatorial Stochastic Processes: Ecole d'eté de probabilités de saint-flour xxxii-2002.
Springer, 2006.

Reference III

[Sib14] Masaaki Sibuya.
Prediction in Ewens–Pitman sampling formula and random samples from number partitions.
Annals of the Institute of Statistical Mathematics, 66(5): 833–864, 2014.

Table of content

α-diversity and power-law of EP partition

For $\alpha \in (0, 1)$, define the probability mass function $p_\alpha(j)$ on $\mathbb{N}$ by

$$\forall j \in \mathbb{N}, \quad p_\alpha(j) = \frac{\alpha \prod_{i=1}^{j-1}(i - \alpha)}{j!}.$$

Stirling formula implies, as $j \to \infty$,

$$p_\alpha(j) = \frac{\alpha}{\Gamma(1-\alpha)} \cdot \frac{\Gamma(j-\alpha)}{\Gamma(j+1)} \sim \frac{\alpha}{\Gamma(1-\alpha)} j^{-(1+\alpha)} = O(j^{-(1+\alpha)}).$$

We call it Sibuya distribution of parameter α, denoted by $\mathrm{Sib}(\alpha)$.

α-diversity and power-law of EP partition

- For each $\alpha \in (0,1)$, let S_α be the positive random variable characterized by

$$\lambda \geq 0,\ E[S_\alpha^\lambda] = e^{-\lambda^\alpha}.$$

- Mittag-Leffler distribution (α) is the law of M_α defined as

$$M_\alpha := (S_\alpha)^{-\alpha}$$

- For each $\theta > -\alpha$, Generalized Mittag-Leffler distribution (α, θ), denoted by $\mathrm{GMtLf}(\alpha, \theta)$, is the distribution with its p.d.f. $g_{\alpha\theta}$ characterized by

$$\forall x > 0,\ g_{\alpha\theta}(x) \propto x^{\theta/\alpha} g_\alpha(x),$$

where $g_\alpha(x)$ is the p.d.f. of Mittag-Leffler distribution (α).

When $\alpha = 0, \theta > 0$

Suppose we partitioned n balls into $\{U_1, U_2, \ldots, U_{K_n}\}$. Then $(n+1)$-th ball will be assigned to

- urn U_i with prob. $|U_i|/(n+\theta)$.
- a new urn with prob. $\theta/(n+\theta)$.

Suppose n balls are partitioned. Then, likelihood is expressed by

$$\frac{\theta^{K_n - 1}}{\prod_{i=1}^{n-1}(\theta + i)} \prod_{j=2}^{n} \{\Gamma(j)\}^{S_{n,j}},$$

which implies K_n is sufficient for θ.

When $\alpha = 0, \theta > 0$

K_n can be represented as the sum of independent Bernoulli as

$$K_n = \sum_{m=1}^{n} X_m, \ X_m \sim \text{Bernoulli}\left(\frac{\theta}{m-1+\theta}\right).$$

We can easily show that

$$\frac{K_n}{\log n} \to \theta \text{ (a.s.)}$$

$$\frac{K_n - \theta \log n}{\sqrt{\theta \log n}} \to \mathcal{N}(0,1) \text{ (weakly)}$$

For $\tilde{\theta}_n := K_n / \log n$, we get

$$\sqrt{\frac{\log n}{\theta}}(\tilde{\theta}_n - \theta) \to \mathcal{N}(0,1) \text{ (weakly)}$$

The above asymptotics also holds for MLE $\hat{\theta}_n$.

Stable convergence

$(\Omega, \mathcal{F}, P)$: A probability space
$C_b(\mathcal{X})$: The set of continuous, bounded functions on $\mathcal{X}$.

Definition (Stable convergence)

For a sub σ-field $\mathcal{G} \subset \mathcal{F}$, a sequence of $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$-valued random variables $(X_n)_{n\geq 1}$ is said to converge $\mathcal{G}$-stably to X if

$$\forall f \in \mathcal{L}^1, \forall h \in C_b(\mathcal{X}), \quad \lim_{n\to\infty} \mathbb{E}[f\,\mathbb{E}[h(X_n)|\mathcal{G}]] = \mathbb{E}[f\,\mathbb{E}[h(X)|\mathcal{G}]].$$

If X is independent of $\mathcal{G}$, X_n is said to converge $\mathcal{G}$-mixing to X.

- $X_n \to X$ $\mathcal{G}$-mixing $\Rightarrow X_n \to X$ $\mathcal{G}$-stably $\Rightarrow X_n \xrightarrow{\mathrm{d}} X$.
- When $\mathcal{G} = \{\emptyset, \Omega\}$, these convergences are equivalent.

Generalization of Slutzky's lemma to Stable convergence

Lemma ([HL15])

For $(\mathcal{X}, \mathcal{B}(\mathcal{X})), (\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, a pair of some separable metrizable spaces, let $(X_n)_{n\geq 1}$ be a sequence of $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$-valued random variables and $(Y_n)_{n\geq 1}$ be a sequence of $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$-valued random variables. Assume that there exists a certain random variable X such that $X_n \to X$ $\mathcal{G}$-stably. Then, the following statements hold.

1. *Let $\mathcal{X} = \mathcal{Y}$. If $d(X_n, Y_n) \xrightarrow{\mathrm{p}} 0$, $Y_n \to X$ $\mathcal{G}$-stably.*
2. *If $Y_n \xrightarrow{\mathrm{p}} Y$ and Y is $\mathcal{G}$-**measurable**, $(X_n, Y_n) \to (X, Y)$ $\mathcal{G}$-stably.*
3. *If $g : \mathcal{X} \to \mathcal{Y}$ is $(\mathcal{B}(\mathcal{X}), \mathcal{B}(\mathcal{Y}))$-measurable and continuous P^X-almost surely, then $g(X_n) \to g(X)$ $\mathcal{G}$-stably.*

Stable Martingale Central Limit Theorem

Lemma ([HL15])

Let $(X_k)_{k\geq 1}$ be a martingale difference sequence with respect to $\mathscr{F}$ and let $(a_n)_{n\geq 1}$ be a sequence of positive real number with $a_n \to \infty$. Assume $(X_k)_{k\geq 1}$ satisfies the following two conditions.

1. $\frac{1}{a_n^2}\sum_{k=1}^n \mathbb{E}[X_k^2|\mathcal{F}_{k-1}] \xrightarrow{\text{p}} \eta^2$ *for some random variable* $\eta \geq 0$.
2. $\frac{1}{a_n^2}\sum_{k=1}^n \mathbb{E}[X_k^2 \mathbb{1}\{|X_k| \geq \epsilon a_n\}|\mathcal{F}_{k-1}] \xrightarrow{\text{p}} 0$ *for all* $\epsilon > 0$.

Then,

$$\frac{1}{a_n}\sum_{k=1}^n X_k \to \eta N \ \mathcal{F}_\infty\text{-stably},$$

where $N \sim \mathcal{N}(0,1)$ and N is independent of $\mathcal{F}_\infty$.