

REVISITED BOX COUNTING TECHNIQUE IN BAYESIAN SENSE

David Blatský¹, Jaromír Kukal¹, Václav Hubata-Vacek¹

¹CTU in Prague, Faculty of Nuclear Sciences and Physical Engineering
Department of Software Engineering
Břehová 7, 115 19 Prague 1
Czech Republic
blatsdav@fjfi.cvut.cz

Abstract: *Fractal patterns appear in a wide variety of sources across nature. The unusual characteristic of fractals is that they entail non-integer dimension. The Box Counting method is one of the often used approach to estimate the fractal dimension of a signal. Thanks to the relationship between entropy and the fractal dimension, it is possible to employ entropy in estimating the fractal dimension. In this paper, we propose to utilize Bayesian estimate of Hartley entropy of a finite sample in fractal dimension estimation. This method was tested on artificial fractals generated by recursive expansion of appropriate matrices.*

Keywords: *unbiased estimation, Hartley entropy, Shannon entropy, Box Counting*

1 Introduction

A fractal is an object whose so-called fractal dimension exceeds its topological dimension and its Hausdorff dimension is non-integer at the same time. The Box Counting method [6] can be used for estimating the fractal dimension due to the relationship

$$\ln C(a) = A - D_0 \ln a, \quad (1)$$

where $a > 0$ is a box size and $C(a)$ is a number of covering elements. Capacity dimension [4] D_0 is estimated as a slope of the line computed by the least square method. These estimates tend to be biased especially for small values of a . We propose to enhance the Box Counting method by Bayesian estimation of Hartley entropy H_0 , which offers better estimate of capacity dimension D_0 .

2 Multinomial Distribution and Naive Entropy Estimates

A multinomial distribution [7] model plays the main role in investigating of point set structures. Let $n \in \mathbb{N}$ be a number of distinguished events. Let $p_j > 0$ be a probability of the j^{th} event for $j = 1, \dots, n$ satisfying $\sum_{j=1}^n p_j = 1$. Then the random variable j has a multinomial distribution $\text{Mul}(p_1, \dots, p_n)$. After realization of multinomial distribution sample of size $N \in \mathbb{N}$, we can count the events and obtain $N_j \in \mathbb{N}_0$ as the number of j^{th} event occurrences for $j = 1, \dots, n$ satisfying $\sum_{j=1}^n N_j = N$. Therefore, we define the number of various events in a sample as $K = \sum_{N_j > 0} 1 \leq \min(n, N)$. Revising Hartley [3] and Shannon [3] entropy definitions

$$H_0 = \ln n, \quad (2)$$

$$H_1 = - \sum_{j=1}^n p_j \ln p_j, \quad (3)$$

we can perform a direct but naive estimation of them as

$$\hat{H}_{0,\text{naive}} = \ln K, \quad (4)$$

$$\hat{H}_{1,\text{naive}} = - \sum_{N_j > 0} \frac{N_j}{N} \ln \frac{N_j}{N}. \quad (5)$$

The main disadvantage of the naive estimates is their biasness. The random variable $K \in \{1, \dots, n\}$ is capped by n , which causes $E\hat{H}_{0,\text{naive}} = E \ln K < E \ln n = \ln n = H_0$. Hence, the naive estimate of Hartley entropy $\hat{H}_{0,\text{naive}}$ is negatively biased. On the other hand, the traditional Box Counting Technique is based on this

estimate. There we plot the logarithm of the covering element number $C(a) \in \mathbb{N}$ against the logarithm of the covering element size $a > 0$ and then estimate their dependency in the linear form $\ln C(a) = A_0 - \hat{D}_{0,\text{naive}} \ln a$. Recognizing equivalence $C(a) = K$ leads to $\ln C(a) = \ln K = \hat{H}_{0,\text{naive}}$ and then $\hat{H}_{0,\text{naive}} = A_0 - \hat{D}_{0,\text{naive}} \ln a$. Defining $\hat{D}_{0,\text{naive}}$ as an estimate of capacity dimension and recognizing the occurrence of $\hat{H}_{0,\text{naive}}$ in the Box Counting procedure [6], we are not surprised to be victims of the bias of Hartley entropy estimate.

A similar situation is the case of Shannon entropy estimation. There are several approaches how to decrease the bias of $\hat{H}_{1,\text{naive}}$ to be closer to a theoretical value of Shannon entropy H_1 . Miller [1] modified the naive estimate $\hat{H}_{1,\text{naive}}$ using a first-order Taylor expansion resulting in

$$\hat{H}_{1,M} = \hat{H}_{1,\text{naive}} + \frac{K-1}{2N}. \quad (6)$$

Lately, Harris [1] improved the formula to

$$\hat{H}_{1,H} = \hat{H}_{1,\text{naive}} + \frac{K-1}{2N} - \frac{1}{12N^2} \left(1 - \sum_{p_j > 0} \frac{1}{p_j} \right). \quad (7)$$

Finally, we can estimate the capacity and information dimensions according to relation

$$\hat{H}_d = A_d - \hat{D}_d \ln a, \quad (8)$$

where \hat{H}_d is any estimate of H_d . Therefore, we can also estimate Hausdorff dimension D_H using inequalities $D_1 \leq D_H \leq D_0$ under the assumption that $\hat{D}_1 \leq D_H \leq \hat{D}_0$ for any “good” estimates \hat{D}_0, \hat{D}_1 of capacity and information dimensions, respectively. The next section is oriented to Bayesian estimation of H_0 and H_1 , which are essential for evaluating \hat{D}_0 and \hat{D}_1 .

3 Bayesian Estimation of Hartley Entropy

We suppose Dirichlet distribution [7] of a random vector $\mathbf{p} = (p_1, \dots, p_n)$ satisfying $p_j \geq 0$, $\sum_{j=1}^n p_j = 1$, with $\alpha_j = \alpha^* > 0$. Using properties of multinomial and its conjugate distribution — the Dirichlet distribution, we can calculate probability estimate $\hat{p}(K|n, N)$ of the random variable $K \in \mathbb{N}$ for $K \leq \min(n, N)$ as

$$\hat{p}(K | n, N) = \text{prob} \left(\sum_{N_j > 0} 1 = K \middle| n, \sum_{j=1}^n N_j = N \right) = \binom{n}{K} \frac{\Gamma(N+1)\Gamma(n\alpha^*)}{\Gamma(N+n\alpha^*)} \sum_{\vec{N} \in \mathbb{D}_{K,N}} \prod_{j=1}^K \frac{\Gamma(N_j + \alpha^*)}{\Gamma(N_j + 1)\Gamma(\alpha^*)}. \quad (9)$$

Derivation of (9) is included in the Appendix A.1. When $N \geq K + 2$, we can calculate

$$S_{K,N} = \sum_{n=K}^{\infty} \hat{p}(K | n, N). \quad (10)$$

When the number of events is constrained as $n \leq n_{\max}$, we apply an alternative formula

$$S_{K,N}^* = \sum_{n=K}^{n_{\max}} \hat{p}(K | n, N). \quad (11)$$

Convergence of the infinite series (10) is proved in the Appendix A.2. Having a knowledge of K, N where $N \geq K + 2$, we can calculate a Bayesian density

$$\hat{p}(n | K, N) = \frac{\hat{p}(K | n, N)}{S_{K,N}}, n \geq K \quad (12)$$

Thereafter, Bayesian estimate of Hartley entropy comes out as

$$\hat{H}_{0,\text{Bayes}} = EH_0 = \sum_{n=K}^{\infty} \hat{p}(n | K, N) \ln n = \sum_{n=K}^{\infty} \frac{\hat{p}(K | n, N) \ln n}{S_{K,N}} = \frac{\sum_{n=K}^{\infty} \hat{p}(K | n, N) \ln n}{\sum_{n=K}^{\infty} \hat{p}(K | n, N)} > \ln K, \quad (13)$$

which is a convergent sum as well. We gain an equivalent formula by substituting $n = K + j$

$$\hat{H}_{0,\text{Bayes}} = \frac{\sum_{j=0}^{\infty} b_j \ln(K+j)}{\sum_{j=0}^{\infty} b_j}, \quad (14)$$

where

$$b_j = \binom{K+j}{j} \frac{B((K+j)\alpha^*, N)}{B(K\alpha^*, N)}. \quad (15)$$

Convergence of the sums in (13) is proved in Appendix A.2. Particular coefficients b_j can also be generated recursively

$$\begin{aligned} b_0 &= 1 \\ b_j &= \frac{K+j}{j} \frac{\Gamma((K+j)\alpha^*)}{\Gamma((K+j)\alpha^* - \alpha^*)} \frac{\Gamma(N + (K+j)\alpha^* - \alpha^*)}{\Gamma(N + (K+j)\alpha^*)} b_{j-1} \\ b_j &= b_{j-1} \frac{K+j}{j} \prod_{u=0}^{N-1} \left(1 - \frac{\alpha^*}{(K+j)\alpha^* + u} \right). \end{aligned} \quad (16)$$

4 Bayesian Estimation of Shannon Entropy

In the case when the number of events n is known, we perform Bayesian estimation of Shannon entropy for arbitrary $\alpha_j = \alpha^* > 0$ as

$$\begin{aligned} & - \sum_{i=1}^M \frac{\Gamma(N + \alpha)}{\Gamma(n_i + \alpha_i)} \frac{\Gamma(n_i + \alpha_i + 1)}{\Gamma(N + \alpha + 1)} \left(\psi^{(0)}(n_i + \alpha_i + 1) - \psi^{(0)}(N + \alpha + 1) \right) \\ \hat{H}_{1,n} = EH_1(K = n) &= - \sum_{j=1}^n \left(\frac{N_j + \alpha^*}{N + n\alpha^*} (\psi(N_j + \alpha^* + 1) - \psi(N + n\alpha^* + 1)) \right), \end{aligned} \quad (17)$$

where ψ is digamma function. However, when the number of events n is unknown, we can use K as a lower estimate of n and perform the final Bayesian estimation as

$$\hat{H}_{1,\text{Bayes}} = \sum_{n=K}^{\infty} p(n | K, N) \hat{H}_{1,n}, \quad (18)$$

which is also a convergent sum for $N \geq K + 2$.

Substituting $n = K + j$, we obtain an adequate formula

$$\hat{H}_{1,\text{Bayes}} = \frac{\sum_{j=0}^{\infty} b_j \hat{H}_{1,K+j}}{\sum_{j=0}^{\infty} b_j}. \quad (19)$$

Unfortunately, asymptotic expansion of (19) depends on individual frequencies N_j . But $\hat{H}_{1,n} \leq \ln n$, accordingly $\hat{H}_{1,K+j} \leq \ln K + j$, which implies the convergence of $\sum_{j=0}^{\infty} b_j \hat{H}_{1,K+j}$ based on majority rule and (14).

5 Revisited Box Counting Method

Let $\mathbb{F} \subset \mathbb{R}^m$ be a set of N points placed into m -dimensional rectangular grid of element size $a > 0$. Let $\hat{H}_{0,\text{Bayes}}$ be an unbiased estimate of Hartley entropy H_0 . Fitting the linear model

$$\hat{H}_{0,\text{Bayes}} = A - \hat{D}_0 \ln a \quad (20)$$

via the method of least squares is called Revisited Box Counting.

Revisited Box Counting can be modified by using $\hat{H}_{1,\text{Bayes}}$ instead of $\hat{H}_{0,\text{Bayes}}$ comes to estimation of information dimension [5] according to

$$\hat{H}_{1,\text{Bayes}} = A - \hat{D}_1 \ln a. \quad (21)$$

6 Experimental Part

The Revisited Box Counting technique will be tested on models of deterministic self-similar 2D fractal sets. They are generated by recursive expansion of binary matrix $\mathbb{G}_{u,v} \in \{0,1\}^{v \times v}$, where u is the number of non-zero elements (units), $v > 1$ is a matrix dimension, and $v < u < v^2$.

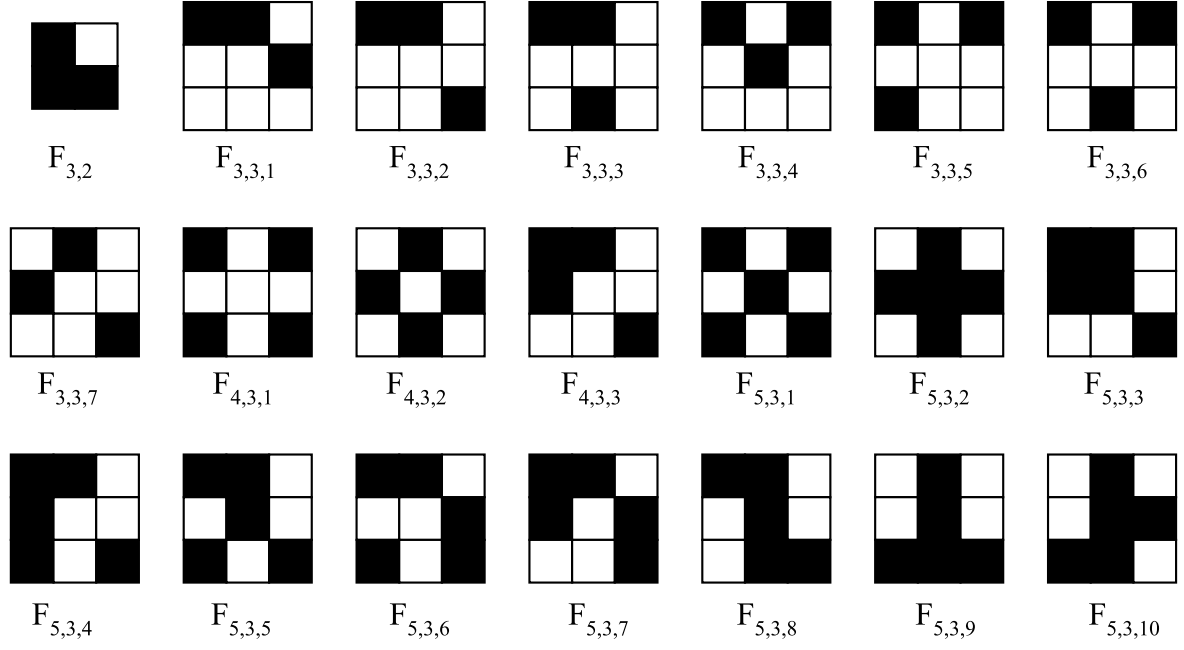


Figure 1: Table of the fractals involved in the research.

Table 1: Particular values of each fractal. α_{model} and p_{value} computed for exponential model

Fractal	h	D_0	$\tilde{D}_{0,\text{naive}}$	α_{opt}	α_{model}	p_{value}
$F_{3,2}$	11	1,585	1,567	0,778	0,749	0,392
$F_{3b,3}$	7	1	0,982	1,885	1,886	0,995
$F_{3c,3}$	7	1	0,977	2,178	1,901	0,131
$F_{3d,3}$	7	1	0,977	2,004	1,901	0,667
$F_{3e,3}$	7	1	0,985	1,958	1,878	0,743
$F_{3f,3}$	7	1	0,976	1,525	1,904	0,129
$F_{3g,3}$	7	1	0,983	1,760	1,883	0,698
$F_{3h,3}$	7	1	0,979	1,873	1,895	0,929
$F_{4a,3}$	7	1,262	1,251	1,302	1,236	0,789
$F_{4b,3}$	7	1,262	1,248	1,196	1,242	0,850
$F_{4c,3}$	7	1,262	1,258	1,502	1,223	0,283
$F_{5a,3}$	7	1,465	1,447	0,871	0,908	0,653
$F_{5b,3}$	7	1,465	1,447	0,830	0,908	0,427
$F_{5c,3}$	7	1,465	1,451	0,941	0,903	0,544
$F_{5d,3}$	7	1,465	1,445	0,886	0,911	0,720
$F_{5f,3}$	7	1,465	1,444	0,864	0,913	0,367
$F_{5g,3}$	7	1,465	1,451	1,046	0,903	0,185
$F_{5h,3}$	7	1,465	1,453	1,061	0,900	0,019
$F_{5i,3}$	7	1,465	1,447	0,841	0,908	0,122
$F_{5j,3}$	7	1,465	1,442	0,755	0,916	0,009
$F_{5k,3}$	7	1,465	1,448	0,860	0,907	0,186

Table 2: Comparison of the models fitting the relationship between $\hat{D}_{0,\text{naive}}$ and α

model	A	B	r
linear	3,905	-2,066	0,646
exponential	2,178	-1,571	0,485
power	0,605	-1,892	0,500

Recursive expansion of $\mathbb{G}_{u,v}$ generates a binary matrix which represents fractal set $\mathbb{F}_{u,v}$ of a similarity dimension $D_S = D_H = D_0 = D_1 = \frac{\log u}{\log v}$. Depth h of recursion depends on v and should be appropriate to computer memory size. The structures involved in the research are depicted in Fig. 1

At first, adequate point sets of given depth h were generated. Then, they were randomly rotated around the origin, and finally they were randomly shifted. Afterwards, a grid of size a was put on the data points and entropy estimates were calculated. Due to physical interpretation of entropy, the estimates were averaged over 20 realizations and mean values of entropy were calculated.

The relationship between \hat{D}_{naive} and optimum value of α was studied on the aforementioned fractals for the grid of size $a = 12, 16, 20, \dots, 480, 500$. Results of optimization are collected in Tab. 1. We suppose, the relationship can be approximated by linear, exponential, or power model as

$$\begin{aligned}\alpha &= A + B\hat{D}_{0,\text{naive}}, \\ \ln \alpha &= A + B\hat{D}_{0,\text{naive}}, \\ \ln \alpha &= A + B \ln \hat{D}_{0,\text{naive}}.\end{aligned}\tag{22}$$

Linearised regression was used for the estimation of unknown parameters. Results are included in Tab. 2. The best model was the exponential and the values α_{model} and corresponding p_{value} from t-test of hypothesis $E\hat{D}_0 = D_0$ are also included in Tab. 1.

7 Conclusion

In this paper we derived the Bayesian estimator $\hat{H}_{0,\text{Bayes}}$ of Hartley entropy for a general value of the parameter α^* . As shown, similar approach can also be used to estimate Shannon entropy $\hat{H}_{1,\text{Bayes}}$. These both can result in less biased estimates of the appropriate entropy when used with proper α^* and the grid size a . According to the relation between fractal (information) dimension and Hartley (Shannon) entropy, we estimated dimensions from experimental data. The results turned out to be significantly more accurate than those estimated naively.

Acknowledgement: The authors acknowledge the funding from the CTU in Prague, Grant SGS14/208/OHK4/3T/14.

References

- [1] Harris, B., *The statistical estimation of entropy in the non-parametric case*. MRC Technical Summary Report, 1975.
- [2] Wendel, J., *Note on the gamma function*. Amer. Math. Monthly, 55(1948), 563-564.
- [3] Renyi, A., *On measures of entropy and information*. In Proc. Fourth Berkeley Symp. Math. Stat. Prob., 1960, volume 1, page 547, Berkeley, 1961. University of California Press.
- [4] Baker, G. L. and Gollub, J. B. *Chaotic Dynamics: An Introduction, 2nd ed.* Cambridge, England: Cambridge University Press, 1996.
- [5] Ott, E. *Chaos in Dynamical Systems*. New York: Cambridge University Press, p. 79, 1993.
- [6] Mandelbrot, B. B. *The Fractal Geometry of Nature*. W.H. Freeman and Company, 1982.
- [7] S. Kotz, N. Balakrishnan, and N. L. Johnson. *Continuous Multivariate Distributions. Volume 1: Models and Applications*. New York: Wiley, 2000. ISBN 0-471-18387-3.
- [8] Bromwich, T. J. I'A. and MacRobert, T. M. *An Introduction to the Theory of Infinite Series, 3rd ed.* New York: Chelsea, p. 39, 1991.

A Appendix

A.1 Derivation of $\hat{p}(K|n, N)$ in (9)

Let $\mathbb{Q}_n = \{\vec{q} \in (\mathbb{R}_0^+)^n \mid \sum_{j=1}^n q_j = 1\}$ be a support set of a Dirichlet-distributed random variable $\vec{p} \in \mathbb{Q}_n$ with parameters α_j , for $j = 1, \dots, n$. The conditional probability of an integer K satisfying $1 \leq K \leq \min(n, N)$ is

$$p(K \mid n, N) = \text{prob} \left(\sum_{N_j > 0} 1 = K \mid n, \sum_{j=1}^n N_j = N \right). \quad (23)$$

The vector of N_j can be reorganized to begin with positive values:

$$p(K \mid n, N) = \binom{n}{K} \text{prob} \left(\forall j = 1, \dots, n : N_j > 0 \Leftrightarrow j \leq K \mid n, \sum_{j=1}^K N_j = N \right). \quad (24)$$

Let $\mathbb{D}_{K,N} = \{\vec{x} \in \mathbb{N}^K \mid \sum_{j=1}^K x_j = N\}$ be the domain of $\vec{N} = (N_1, \dots, N_K) \in \mathbb{D}_{K,N}$. Using the mean value of a multinomial distribution over \mathbb{Q}_n , we obtain an unbiased estimate of $p(K \mid n, N)$ as

$$\hat{p}(K \mid n, N) = \binom{n}{K} \mathbb{E} \left(\sum_{\vec{N} \in \mathbb{D}_{K,N}} \binom{N}{N_1, \dots, N_K} \prod_{j=1}^K p_j^{N_j} \prod_{j=K+1}^n p_j^0 \right) = \binom{n}{K} \sum_{\vec{N} \in \mathbb{D}_{K,N}} \binom{N}{N_1, \dots, N_K} \mathbb{E} \left(\prod_{j=1}^K p_j^{N_j} \right). \quad (25)$$

Using the generalized Beta function

$$B(\vec{x}) = \int_{\vec{p} \in \mathbb{Q}_n} \prod_{j=1}^m p_j^{x_j-1} d\vec{p} = \frac{\prod_{j=1}^m \Gamma(x_j)}{\Gamma(\sum_{j=1}^m x_j)}, \quad (26)$$

we can calculate

$$\mathbb{E} \left(\prod_{j=1}^K p_j^{N_j} \right) = \frac{\int_{\vec{p} \in \mathbb{Q}_n} B(\vec{\alpha})^{-1} \prod_{j=1}^K p_j^{N_j+\alpha_j-1} \prod_{j=K+1}^n p_j^{\alpha_j-1} d\vec{p}}{\int_{\vec{p} \in \mathbb{Q}_n} B(\vec{\alpha})^{-1} \prod_{j=1}^n p_j^{\alpha_j-1} d\vec{p}} = \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)} \prod_{j=1}^K \frac{\Gamma(N_j+\alpha_j)}{\Gamma(\alpha_j)}, \quad (27)$$

where α is the sum of all α_j . Therefore,

$$\hat{p}(K \mid n, N) = \binom{n}{K} \sum_{\vec{N} \in \mathbb{D}_{K,N}} \frac{N!}{\prod_{j=1}^K N_j!} \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)} \frac{\prod_{j=1}^K \Gamma(N_j+\alpha_j)}{\prod_{j=1}^K \Gamma(\alpha_j)} = \binom{n}{K} \frac{\Gamma(N+1)\Gamma(\alpha)}{\Gamma(N+\alpha)} \sum_{\vec{N} \in \mathbb{D}_{K,N}} \prod_{j=1}^K \frac{\Gamma(N_j+\alpha_j)}{\Gamma(N_j+1)\Gamma(\alpha_j)} \quad (28)$$

In this particular paper, we assume $\alpha_j = \alpha^*, \forall j = 1, \dots, n$ which results in a simpler form of Equation 28

$$\hat{p}(K \mid n, N) = \binom{n}{K} \frac{\Gamma(N+1)\Gamma(n\alpha^*)}{\Gamma(N+n\alpha^*)} \sum_{\vec{N} \in \mathbb{D}_{K,N}} \prod_{j=1}^K \frac{\Gamma(N_j+\alpha^*)}{\Gamma(N_j+1)\Gamma(\alpha^*)}. \quad (29)$$

A.2 Convergence of $\sum_{j=0}^{\infty} b_j \ln(K+j)$ in (14) and $\sum_{j=0}^{\infty} b_j$ in (10)

The ratio of coefficients b_j could be expressed as:

$$q_j = \frac{b_j}{b_{j-1}} \frac{\ln(K+j)}{\ln(K+j-1)} = \frac{(K+j)}{j} \frac{\ln(K+j)}{\ln(K+j-1)} \frac{\Gamma((K+j)\alpha^*)}{\Gamma((K+j-1)\alpha^*)} \frac{\Gamma(N+(K+j-1)\alpha^*)}{\Gamma(N+(K+j)\alpha^*)}. \quad (30)$$

Starting with inequality proved by Wendel [2]:

$$\forall d \in [0; 1], \forall x > 0 : \frac{\Gamma(x+d)}{\Gamma(x)} \leq x^d; \quad (31)$$

that can be generalized for $\delta = D + d$ where $D \in \mathbb{N}_0, d \in [0; 1)$ as

$$\frac{\Gamma(x+\delta)}{\Gamma(x)} \leq x^d \prod_{i=0}^{D-1} (x+i+d). \quad (32)$$

We should see the similarity between $\alpha^* = A + a$, where $A \in \mathbb{N}_0, a \in [0; 1)$, and δ leading to

$$q_j = \frac{b_j}{b_{j-1}} \frac{\ln(K+j)}{\ln(K+j-1)} \leq \frac{K+j}{j} \frac{\ln(K+j)}{\ln(K+j-1)} \left(\frac{(K+j-1)\alpha^*}{(K+j-1)\alpha^* + N} \right)^a \cdot \prod_{i=0}^{A-1} \frac{(K+j-1)\alpha^* + i + a}{(K+j-1)\alpha^* + i + a + N} \quad (33)$$

$$q_j = \frac{b_j}{b_{j-1}} \frac{\ln(K+j)}{\ln(K+j-1)} \leq \frac{K+j}{j} \frac{\ln(K+j)}{\ln(K+j-1)} \left(\frac{(K+j-1)\alpha^*}{(K+j-1)\alpha^* + N} \right)^a \quad (34)$$

The Raabe criterion [8] will state series of positive members $\sum_{n=0}^{\infty} a_n$ as convergent if exists $L = \lim_{n \rightarrow \infty} n \left(\frac{a_n}{a_{n+1}} - 1 \right)$ satisfying $L > 1$. Then we can calculate

$$L = \lim_{j \rightarrow \infty} j \left(\frac{b_{j-1}}{b_j} \frac{\ln(K+j-1)}{\ln(K+j)} - 1 \right) \geq \lim_{j \rightarrow \infty} j \left(\frac{j}{K+j} \frac{\ln(K+j-1)}{\ln(K+j)} \left(\frac{(K+j-1)\alpha^* + N}{(K+j-1)\alpha^*} \right)^a - 1 \right). \quad (35)$$

Substitution $x = K + j$ leads to

$$L = \lim_{x \rightarrow \infty} (x - K) \left(\frac{x - K}{x} \frac{\ln(x-1)}{\ln(x)} \left(\frac{(x-1)\alpha^* + N}{(x-1)\alpha^*} \right)^a - 1 \right), \quad (36)$$

and finally

$$L = -K + \lim_{x \rightarrow \infty} \left(x \left(1 + \frac{N - \alpha^*}{x\alpha^*} \right)^a - x \left(1 - \frac{1}{x} \right)^a \right). \quad (37)$$

Substituting $h = x^{-1} \rightarrow 0^+$ and applying l'Hospital rule, we obtain

$$L = -K + \lim_{h \rightarrow 0^+} \frac{\left(1 + \frac{N - \alpha^*}{\alpha^*} h \right)^a - (1 - h)^a}{h} = N - K. \quad (38)$$

Thus the series $\sum_{j=0}^{\infty} b_j \ln(K+j)$ converges absolutely for $K \leq N - 2$ because $L = N - K > 1$. According to majority rule, the series $\sum_{j=0}^{\infty} b_j = \sum_{n=K}^{\infty} \hat{p}(K | n, N)$ converges as well.

A.3 Matlab library

A.3.1 Bayesian estimate $H_{0, \text{Bayes}}$

```
function H0=HARTLEYBAYES(Nvec, alpha)
    N = sum(Nvec); K = sum(Nvec>0);
    bay=1; bay0=log(K); H0=bay0/bay; b=1;
    for j = 1 : length(Nvec)-K
        H0old=H0;
        b=b/j*(K+j) * prod(ones(1, N)-alpha./((K+j)*alpha:N-1+(K+j)*alpha));
        bay=bay*b;
        bay0=bay0+b*log(K+j);
        H0=bay0/bay;
        if bay>1e200
            bay0=bay0/bay;
            b=b/bay;
            bay=1;
        end
        if abs(H0-H0old) < 1e-8
            break
        end
    end
end
```

A.3.2 Bayesian estimate $H_{1, \text{Bayes}}$

```
function H1=SHANNONBAYES(Nvec, alpha)
    N=sum(Nvec); K=sum(Nvec>0);
    Nvec=Nvec(Nvec>0);
```

```

bay=1; bay1=SHANNONFIXED(Nvec); H1=bay1/bay; b=1;
for j=1: length(Nvec)-K
    H1old=H1;
    b=b/j*(K+j) * prod(ones(1, N)-alpha./((K+j)*alpha:N-1+(K+j)*alpha));
    bay=bay+b;
    Nvec=[Nvec, 0];
    bay1=bay1+b*SHANNONFIXED(Nvec);
    H1=bay1/bay;
    if bay>1e200;
        bay1=bay1/bay;
        b=b/bay;
        bay=1;
    end
    if abs(H1-H1old)< 1e-8
        break
    end
end
end
end

```

A.3.3 Bayesian estimate $H_{1,n}$

```

function H1=SHANNONFIXED(Nvec)
    N=sum(Nvec);
    n=length(Nvec);
    H1=((Nvec+1)/(N+n))'*(psi(N+n+1)-psi(Nvec+2));
end

```