REVISITED BOX COUNTING TECHNIQUE IN BAYESIAN SENSE

Václav Hubata–Vacek[1], Jaromír Kukal[1]

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

Abstract:

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

1 Introduction

Fractals are objects, which exceed its topological dimensions and its Hausdorff dimension is not integer. Box Counting method [6] is used for estimating fractal dimension in the form

$$\ln C(a) = A - D_0 \ln a, \tag{1}$$

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

2 Multinomial Distribution and Naive Entropy Estimates

Multinomial distribution [7] model plays main role in investigation of point set structures. Let $n \in \mathbb{N}$ be number of distinguish events. Let $p_j > 0$ be probability of j^{th} event for $j = 1, ..., n$ satisfying $\sum_{j=1}^{n} p_j = 1$. Then random variable j has multinomial distribution $\text{Mul}(p_1, ..., p_n)$. 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 number of j^{th} event occurrences for $j = 1, ..., n$ satisfying $\sum_{j=1}^{n} N_j = N$. Therefore, we define number of various events in sample as $K = \sum_{N_j>0} 1 \leq \min(n, N)$. Remembering Hartley [3] and Shannon [3] entropy definition as

$$H_0 = \ln n, \tag{2}$$

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

we can perform direct but naive estimation of them as

$$\hat{H}_{0,\text{naive}} = \ln K, \tag{4}$$

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

The main disadvantage of naive estimates is their biasness. Random variable $K = \{1, ..., n\}$ is upper constrained by n, then $\mathrm{E}\hat{H}_{0,\text{naive}} = \mathrm{E} \ln K < \mathrm{E} \ln n = \ln n = H_0$. Therefore, naive estimate of Hartley entropy $\hat{H}_{0,\text{naive}}$ is negative biased. On the other hand, traditional Box Counting Technique is based on this estimate because we plot logarithm of covering element number $C(a) \in \mathbb{N}$ against logarithm of covering element size $a > 0$ and then estimate their dependency in linear form $\ln C(a) = A_0 - \hat{D}_{0,\text{naive}} \ln a$. Recognizing equivalence $C(a) = K$, we obtain $\ln C(a) = \ln K = \hat{H}_{0,\text{naive}}$ and then $\hat{H}_{0,\text{naive}} = A_0 - \hat{D}_{0,\text{naive}} \ln a$. Defining $\hat{D}_{0,\text{naive}}$ as estimate of capacity dimension and recognizing the occurrence of $\hat{H}_{0,\text{naive}}$ in Box Counting procedure [6], we are not surprised to be victims of the bias of Hartley entropy estimate.

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 Shannon entropy H_1. Miller [1] modified naive estimate $\hat{H}_{1,\text{naive}}$ using first order Taylor expansion, which produces

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

Lately, Harris [1] improved the formula to

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

Finally, we can estimate capacity and information dimension according to relation

$$\hat{H}_d = A_d - \hat{D}_d \ln a \tag{8}$$

where $\hat{H}_d$ is any estimate of H_d. Therefore, we can also estimate Hausdorff dimension D_H using inequalities $D_1 \le D_\mathrm{H} \le D_0$ and then also supposing $\hat{D}_1 \le D_\mathrm{H} \le \hat{D}_0$ for any "good" estimates $\hat{D}_0$, $\hat{D}_1$ of capacity and information dimensions. Next section is oriented to Bayesian estimation of H_0, H_1 for $\hat{D}_0$ and $\hat{D}_1$ evaluations.

3 Bayesian Estimation of Hartley Entropy

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

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

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

$$S_{K,N} = \sum_{n=K}^{\infty} \hat{\mathrm{p}}\,(K \mid n, N). \tag{10}$$

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

$$S^*_{K,N} = \sum_{n=K}^{n_{\max}} \hat{\mathrm{p}}\,(K \mid n, N). \tag{11}$$

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

$$\hat{\mathrm{p}}\,(n \mid K, N) = \frac{\hat{\mathrm{p}}\,(K \mid n, N)}{S_{K,N}}, \tag{12}$$

for $n \ge K$. Therefore, Bayesian estimate of Hartley entropy is

$$\hat{H}_{0,\text{Bayes}} = \mathrm{E}H_0 = \sum_{n=K}^{\infty}\hat{\mathrm{p}}\,(n \mid K, N)\ln n = \sum_{n=K}^{\infty}\frac{\hat{\mathrm{p}}\,(K \mid n, N)\ln n}{S_{K,N}} = \frac{\sum_{n=K}^{\infty}\hat{\mathrm{p}}\,(K \mid n, N)\ln n}{\sum_{n=K}^{\infty}\hat{\mathrm{p}}\,(K \mid n, N)} > \ln K \tag{13}$$

which is also a convergent sum. Substituting $n = K + j$ we gain equivalent formula

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

where

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

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

$$\begin{aligned} b_0 &= 1 \\ b_j &= \frac{(K+j)}{j} \frac{\Gamma((K+j)\alpha^*)}{\Gamma((K+j)\alpha^* - \alpha^*)} \frac{\Gamma(N+(K+j)\alpha^* - \alpha^*)}{\Gamma(N+(K+j)\alpha^*)} b_{j-1}. \end{aligned} \tag{16}$$

Asymptotic properties of Bayesian estimate can be investigated for $N \to +\infty$ via limits

$$\begin{aligned} &\lim_{N\to+\infty} \hat{H}_{0,\text{Bayes}} = \ln K, \\ &\lim_{N\to+\infty} (\hat{H}_{0,\text{Bayes}} - \ln K) N = K(K+1)\ln(1+1/K), \\ &\lim_{N\to+\infty} \left(\hat{H}_{0,\text{Bayes}} - \ln K - \frac{K(K+1)\ln(1+1/K)}{N} \right) N^2 = \\ &\quad \frac{1}{2} \left(K(K+2)(K+1) \left(\ln(K+2) - \ln(K) - 2K\ln(K+1) + K\ln(K+2) + K\ln(K) \right) \right), \end{aligned} \tag{17}$$

Therefore

$$\begin{aligned} \hat{H}_{0,\text{Bayes}} \approx &\ln K + \frac{K(K+1)\ln(1+1/K)}{N} + \\ &\frac{\left(K(K+2)(K+1) \left(\ln(K+2) - \ln(K) - 2K\ln(K+1) + K\ln(K+2) + K\ln(K) \right) \right)}{2N^2} \end{aligned} \tag{18}$$

When K is also large, we can roughly approximate Hartley entropy as

$$\hat{H}_{0,\text{Bayes}} \approx \ln K + \frac{K+1}{N} \tag{19}$$

which is very similar to Miller correction (6) in the case of Shannon entropy estimation. Meanwhile formula (13) represents Bayesian estimate of H_0, formulas (4), (19), and (18) are approximations of zero, first, and second order.

Formula (14) can be also expanded to the form

$$\hat{H}_{0,\text{Bayes}} = \ln K + \sum_{j=1}^{\infty} \frac{1}{j} \varphi(K) \left(\frac{K}{N} \right)^j \tag{20}$$

where $\varphi(K) > 1$ for all $K \in \mathbb{N}$ and $\lim_{K\to\infty} \varphi(K) = 1$. Therefore, we obtain lower estimate

$$\hat{H}_{0,\text{Bayes}} > \ln K + \sum_{j=1}^{\infty} \frac{1}{j} \left(\frac{K}{N} \right)^j = \ln K - \ln(1 - \frac{K}{N}) = H_{0,\text{low}} \tag{21}$$

which exists for $K < N$.

4 Bayesian Estimation of Shannon Entropy

In the case when the number of events n is known, we can perform Bayesian estimation of Shannon entropy as

$$\hat{H}_{1,n} = \mathrm{E}H_1(K=n) = -\sum_{j=1}^{n} \left(\frac{N_j+1}{N+n} \left(\psi(N_j+2) - \psi(N+n+1) \right) \right) \tag{22}$$

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

$$\hat{H}_{1,\text{Bayes}} = \sum_{n=K}^{\infty} \mathrm{p}\left(n \mid K, N \right) \hat{H}_{1,n} \tag{23}$$

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

Substituting $n = K + j$ we obtain adequate formula.

$$\hat{H}_{1,\text{Bayes}} = \frac{\sum_{j=0}^{\infty} b_j H_{1,K+j}}{\sum_{j=0}^{\infty} b_j} \tag{24}$$

Unfortunately asymptotic expansion of (24) depends on individual frequencies N_j. But $\hat{H}_{1,n} \leq \ln n$ and therefore $\hat{H}_{1,K+j} \leq \ln K + j$ which implies the convergence of $\sum_{j=0}^{\infty} b_j H_{1,K+j}$ according to majority rule and (14).

5 Revisited Box Counting Method

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

$$\hat{H}_{0,\mathrm{Bayes}} = A - D_0 \ln a \tag{25}$$

via method of least squares is called Revisited Box Counting.

Revisited Box Counting can be modified by:

- using $\hat{H}_{1,\mathrm{Bayes}}$ instead of $\hat{H}_{0,\mathrm{Bayes}}$ comes to estimation of information dimension [5] according to model

$$\hat{H}_{1,\mathrm{Bayes}} = A - D_1 \ln a \tag{26}$$

- using non-trivial approximations of $\hat{H}_{0,\mathrm{Bayes}}$, namely: $\hat{H}_{0,1}, \hat{H}_{0,2}, \hat{H}_{0,\mathrm{low}}$ instead of $\hat{H}_{0,\mathrm{Bayes}}$

Remark:
Using of $\hat{H}_{0,0} \equiv \hat{H}_{0,\mathrm{naive}}$ instead of $\hat{H}_{0,\mathrm{Bayes}}$ comes back to traditional Box Counting.

6 Experimental Part

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 number of nonzero elements (units), $v > 1$ is matrix dimension, and $v < u < v^2$.

Recursive expansion of $\mathbb{G}_{u,v}$ generates binary matrix which represents fractal set $\mathbb{F}_{u,v}$ of similarity dimension $D_\mathrm{S} = D_\mathrm{H} = D_0 = D_1 = \frac{\log u}{\log v}$. Depth h of recursion depends on v and computer memory size.

Four testing sets: $\mathbb{F}_{3,2}$, $\mathbb{F}_{4,3}$, $\mathbb{F}_{5,3}$, $\mathbb{F}_{8,3}$ were generated by matrices:

- $\mathbb{G}_{3,2} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ for $h = 11$
- $\mathbb{G}_{4,3} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ for $h = 7$
- $\mathbb{G}_{5,3} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ for $h = 7$
- $\mathbb{G}_{8,3} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ for $h = 7$

Sets $\mathbb{G}_{3,2}, \mathbb{G}_{8,3}$ corresponds to Sierpinski gasket and carpet.

Dimensions of $\mathbb{F}_{3,2}$, $\mathbb{F}_{4,3}$, $\mathbb{F}_{5,3}$, $\mathbb{F}_{8,3}$ are

$$\begin{aligned} \dim(\mathbb{F}_{3,2}) &= \frac{\log 3}{\log 2} = 1.5850, \\ \dim(\mathbb{F}_{4,3}) &= \frac{\log 4}{\log 3} = 1.2619, \\ \dim(\mathbb{F}_{5,3}) &= \frac{\log 5}{\log 3} = 1.4650, \\ \dim(\mathbb{F}_{8,3}) &= \frac{\log 8}{\log 3} = 1.8928. \end{aligned} \tag{27}$$

Adequate points sets with given depth h were generated, first. Then, they were randomly rotated around origin and finally they were randomly shifted. After these operations the grid of size a were put on the data points and entropy estimates were calculated. Due to physical interpretation of entropy, the estimates were averaged over 10 realizations and mean values of entropy were calculated.

Various estimates of Hartley entropy for grid size $a = 2, 3, 4, ..., 30$ are depicted on Fig **??**. Hartley entropy estimates are similar each other except under-biased naive estimate. Estimates $H_{0,\mathrm{BAYES}}$ and $H_{0,\mathrm{low}}$ are very similar in these four cases. Corresponding estimates of Shannon entropy are depicted on Fig. **??** in the same range.

As seen in the Fig. **??** and **??**, too small grid size $a \leq 20$ comes to underestimation of $H_{0,\mathrm{naive}}, H_{1,\mathrm{naive}}$, but the other estimates are unfortunately overestimated. Therefore, revisited Box Counting was applied in the range $30 \leq a \leq 100$.

Using least square method we obtained various estimates of D_0 and D_1 and compared them with similarity dimension. Results of estimation are collected in Tabs. **??** - **??**, where ED is point estimate of given dimension, s_D is its standard deviation, and p_{value} is probability from t-test of hypothesis

$$\mathrm{H}_0 : \mathrm{E}D = D_{\mathrm{S}} \tag{28}$$

7 Conclusion

Acknowledgement: The paper was created with the support of 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 $\hat{\mathrm{p}}(K|n,N)$ in (9)

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

$$\mathrm{p}(K \mid n, N) = \mathrm{prob}\left(\sum_{N_j>0} 1 = K \,\middle|\, n, \sum_{j=1}^{n} N_j = N\right). \tag{29}$$

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

$$\mathrm{p}(K \mid n, N) = \binom{n}{K}\mathrm{prob}\left(\forall j = 1, ..., n : N_j > 0 \Leftrightarrow j \leq K \,\middle|\, n, \sum_{j=1}^{K} N_j = N\right). \tag{30}$$

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

$$\hat{\mathrm{p}}(K \mid n, N) = \binom{n}{K}\mathrm{E}\left(\sum_{\vec{N}\in\mathbb{D}_{K,N}} \binom{N}{N_1, ..., N_K} \prod_{j=1}^{K} p_j^{N_j} \prod_{j=k+1}^{n} p_j^0\right) = \binom{n}{K} \sum_{\vec{N}\in\mathbb{D}_{K,N}} \binom{N}{N_1, ..., N_K}\mathrm{E}\left(\prod_{j=1}^{K} p_j^{N_j}\right). \tag{31}$$

Using the generalized Beta function

$$B(\vec{x}) = \int_{\vec{p}\in\mathbb{Q}_m} \prod_{j=1}^{m} p_j^{x_j-1}\mathrm{d}\vec{p} = \frac{\prod_{j=1}^m \Gamma(x_j)}{\Gamma(\sum_{j=1}^m x_j)}, \tag{32}$$

we can calculate

$$\mathrm{E}\left(\prod_{j=1}^{K} p_j^{N_j}\right) = \frac{\int_{\vec{p}\in\mathbb{Q}_n} B(\vec{\alpha})^{-1} \prod_{j=1}^K p_j^{N_j+\alpha_j-1} \prod_{j=K+1}^n p_j^{\alpha_j-1}\mathrm{d}\vec{p}}{\int_{\vec{p}\in\mathbb{Q}_n} B(\vec{\alpha})^{-1} \prod_{j=1}^n p_j^{\alpha_j-1}\mathrm{d}\vec{p}} = \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)} \prod_{j=1}^{K} \frac{\Gamma(N_j+\alpha_j)}{\Gamma(\alpha_j)}, \tag{33}$$

where α is the sum of all α_j. Therefore

$$\hat{\mathrm{p}}(K|n, N) = \binom{n}{K} \sum_{\vec{N}\in\mathbb{D}_{K,N}} \frac{N!}{\prod_{j=1}^K N_j!} \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)} \frac{\prod_{j=1}^K \Gamma(N_j+\alpha_j)}{\prod_{j=1}^K \Gamma(\alpha_j)} = \binom{n}{K} \frac{\Gamma(N+1)\Gamma(\alpha)}{\Gamma(N+\alpha)} \sum_{\vec{N}\in\mathbb{D}_{K,N}} \prod_{j=1}^{K} \frac{\Gamma(N_j+\alpha_j)}{\Gamma(N_j+1)\Gamma(\alpha_j)} \tag{34}$$

In this particular paper, we assume $\alpha_j = \alpha^*, \forall j = 1, ..., n$ which results in easier form of the equation 34:

$$\hat{\mathrm{p}}(K \mid n, N) = \binom{n}{K} \frac{\Gamma(N+1)\Gamma(n\alpha^*)}{\Gamma(N+n\alpha^*)} \sum_{\vec{N}\in\mathbb{D}_{K,N}} \prod_{j=1}^{K} \frac{\Gamma(N_j+\alpha^*)}{\Gamma(N_j+1)\Gamma(\alpha^*)} \tag{35}$$

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:

$$q_j = \frac{b_j}{b_{j-1}} \frac{\ln(K+j)}{\ln(K+j-1)} = \frac{(K+j)}{j} \frac{\ln(K+j)}{\ln(K+j-1)} \frac{\Gamma((K+j)\alpha^*)}{\Gamma((K+j-1)\alpha^*)} \frac{\Gamma(N+(K+j-1)\alpha^*)}{\Gamma(N+(K+j)\alpha^*)}. \tag{36}$$

Starting with inequality proved by Wendel [2]:

$$\forall\Delta \in \langle 0; 1\rangle, \forall x > 0 : \frac{\Gamma(x+\Delta)}{\Gamma(x)} \leq x^{\Delta}, \tag{37}$$

we should realize that $\alpha^* \in \langle 0; 1\rangle$ and than see the similarity leading to

$$q_j = \frac{b_j}{b_{j-1}} \frac{\ln(K+j)}{\ln(K+j-1)} \leq \frac{K+j}{j} \frac{\ln(K+j)}{\ln(K+j-1)} \left(\frac{(K+j-1)\alpha^*}{(K+j-1)\alpha^*+N}\right)^{\alpha^*} \tag{38}$$

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

$$L = \lim_{j\to\infty} j\left(\frac{b_{j-1}}{b_j}\frac{\ln(K+j-1)}{\ln(K+j)} - 1\right) \geq \lim_{j\to\infty} j\left(\frac{j}{K+j}\frac{\ln(K+j-1)}{\ln(K+j)}\left(\frac{(K+j-1)\alpha^* + N}{(K+j-1)\alpha^*}\right)^{\alpha^*} - 1\right). \quad (39)$$

Substitution $x = K + j$ leads to

$$L = \lim_{x\to\infty}(x-K)\left(\frac{x-K}{x}\frac{\ln(x-1)}{\ln(x)}\left(\frac{(x-1)\alpha^* + N}{(x-1)\alpha^*}\right)^{\alpha^*} - 1\right) \quad (40)$$

and finally

$$L = -K + \lim_{x\to\infty}\left(x\left(1 + \frac{N-\alpha^*}{x\alpha^*}\right)^{\alpha^*} - x\left(1 - \frac{1}{x}\right)^{\alpha^*}\right) \quad (41)$$

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

$$L = -K + \lim_{h\to 0^+}\frac{\left(1 + \frac{N-\alpha^*}{\alpha^*}h\right)^{\alpha^*} - (1-h)^{\alpha^*}}{h} = N - K \quad (42)$$

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

A.3 Matlab library

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

```
function H0=HARTLEYBAYES(N, alpha)
    tol=1e200;
    nmax=length(N);
    K=sum(N>0);
    Ntotal=sum(N);

    bay=1; bay0=log(K); H0=bay0/bay; b=1;
    for j=1:nmax-K
       H0old=H0;
       b=b/j*(K+j)*gamma((K+j)*alpha)/gamma((K+j)*alpha-alpha)*
                gamma(Ntotal+(K+j)*alpha-alpha)/gamma(Ntotal+(K+j)*alpha);
       bay=bay+b;
        bay0=bay0+b*log(K+j);
        H0=bay0/bay;
        if bay>tol
            bay0=bay0/bay;
            b=b/bay;
            bay=1;
        end
        if abs(H0-H0old) < 1e-8
            break
        end
    end
end
```

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

```
function H1=SHANNONBAYES(N)
    nstar=1e7;
    tol=1e200;
    Ntotal=sum(N);
```

```
    nmax=length(N);
    k=sum(N>0);
    N=N(N>0);
    if k>Ntotal-2 && nmax==nstar
        H1=NaN;
        return
    end
    bay=1;
    bay1=SHANNONFIXED(N);
    b=1;
    H1=bay1/bay;
    for j=1:nmax-k
        H1old=H1;
        b=b/j*(k+j)*(k+j-1)/(k+Ntotal+j-1);
        bay=bay+b;
        N=[N,0];
        bay1=bay1+b*SHANNONFIXED(N);
        H1=bay1/bay;
        if bay>tol
            bay1=bay1/bay;
            b=b/bay;
            bay=1;
        end
        if abs(H1-H1old)< 1e-8
            break
        end
    end
end
```

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

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