ISSN 2279-9362

Collegio Carlo Alberto

Good-Toulmin type estimators for the number of unseen species

Federico Camerlenghi
Emanuele Dolera
Stefano Favaro

No. 536
December 2017

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

Good-Toulmin type estimators for the number of unseen species

Federico Camerlenghi[1,*], Emanuele Dolera[2] and Stefano Favaro[3,†]

[1] Department of Economics, Management and Statistics, University of Milano–Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy.
[2] Department of Mathematics, University of Pavia, via Ferrata 5, 27100 Pavia, Italy.
[3] Department of Economics and Statistics, University of Torino, Corso Unione Sovietica 218/bis, 10134, Torino, Italy.

Abstract

Many scientific disciplines use n samples to extrapolate information on the species composition of additional unobservable samples. Given a population with unknown species proportions $(p_i)_{i\geq 1}$, we consider the problem of estimating: i) the number $U_{\lambda n,r}$ of unseen species with frequency $r \geq 1$ in an additional sample of size λn; ii) the probability $V_{\lambda n,r}$ of discovering at the $(\lambda n + n + 1)$-th draw a species with frequency $r \geq 0$ in the enlarged sample of size $n + \lambda n$. The former is related to the problem of controlling how many unseen species are rare, whereas the letter is related to the problem of evaluating the cost-effectiveness of further sampling. We introduce nonparametric empirical Bayes estimators of $U_{\lambda n,r}$ and $V_{\lambda n,r}$, we show that they estimate $U_{\lambda n,r}$ and $V_{\lambda n,r}$ all of the way up $\lambda \propto \log_2(n)/2r\log_2(\log_2(n))$, and that this range is the best possible. The proposed estimators are distribution-free, namely no assumptions are imposed on $(p_i)_{i\geq 1}$. We then consider tuning our estimators for heavy-tailed $(p_i)_{i\geq 1}$. To do that we impose regular variation as a distribution-specific assumption for the p_i's. Interestingly, the resulting regularly varying estimators are asymptotic equivalent, for large n, to their Bayesian nonparametric counterparts under a Poisson-Dirichlet prior for $(p_i)_{i\geq 1}$.

Keywords: Euler's smoothing; Good-Toulmin type estimators; nonparametric empirical Bayes; rare species; regular variation; species estimation; two parameter Poisson-Dirichlet prior.

1 Introduction

Species estimation is an important problem in many scientific fields. It first appeared in ecology, and its importance has grown considerably in the recent years, driven by chal-

*Also affiliated to BIDSA, Bocconi University, Milano and to the Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Torino, Italy.

†Also affiliated to the Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Torino, Italy.

lenging applications in the broader area of biosciences. See, e.g., Bunge and Fitzpatrick (1993), Chao (2005), Bunge et al. (2014) and Chao and Chiu (2016) for comprehensive review papers. Among the numerous approaches to species estimation, the "extrapolation approach" of Fisher et al (1943) is arguably the most popular. Consider a population of individuals $(X_i)_{i\geq 1}$ of various species with unknown proportions $(p_i)_{i\geq 1}$, and let $(X_1, \ldots, X_n)$ be a sample from such a population. If $N_{n,i}$ denotes the frequency of the i-th species in the sample, then the "extrapolation approach" calls for estimating

$$U_{\lambda n} := \sum_{i\geq 1} \mathbb{1}_{\{N_{n,i}=0\}} \mathbb{1}_{\{N_{\lambda n,i}>0\}}, \tag{1}$$

the number of hitherto unseen species that would be observed if λn additional samples $(X_{n+1}, \ldots, X_{\lambda n-n})$ were collected from the same population, for $\lambda > 0$. Besides biosciences, the estimation of $U_{\lambda n}$ has found applications in linguistics to asses vocabulary size (Efron and Thisted (1976)), in network tomography for the analysis of node degrees with source-destination data (Vardi (1996) and Zhang (2005)), in database attribute variation (Haas et al. (1995)), in multi-armed bandits for security analysis of engineering systems (Bubeck et al. (2013)) and in password innovation (Florencio and Herley (2007)).

An intriguing solution to the problem of estimating $U_{\lambda n}$ was proposed in the seminal works of Good (1953) and Good and Toulmin (1956), and further developed by Efron and Thisted (1976). Let $M_{n,r}$ denote the number of species with frequency r in $(X_1, \ldots, X_n)$, for any $1 \leq r \leq n$, and denote by $m_{n,r}$ the corresponding value in the observed sample; the $M_{n,r}$'s are typically referred to as frequencies of frequencies, that is $M_{n,r} := \sum_{i\geq 1} \mathbb{1}_{\{N_{n,i}=r\}}$ such that $K_n := \sum_{1\leq r\leq n} M_{n,r}$ is the number of species in the sample and $n = \sum_{1\leq r\leq n} rM_{n,r}$. Good and Toulmin (1956) estimated $U_{\lambda n}$ by the simple formula

$$\hat{U}_{\lambda n} := -\sum_{j\geq 1} (-\lambda)^j m_{n,j}. \tag{2}$$

Unfortunately (2) is useless for $\lambda \geq 1$. This is because of its high variance due to the exponential growth of the coefficients $(-\lambda)^j$. To overcome this drawback Good and Toulmin (1956) and Efron and Thisted (1976) suggested to smooth $\hat{U}_{\lambda n}$ by truncating its Euler's transformation, which forces convergence of the series. However, no theoretical guarantees for the resulting estimator were given. Remarkable progress in this direction have been recently made by Orlitsky et al. (2017). They proposed to estimate $U_{\lambda n}$ by

$$\hat{U}^L_{\lambda n} := -\sum_{j\geq 1} (-\lambda)^j \mathbb{P}[L \geq j] m_{n,j}, \tag{3}$$

where L is a random variable whose tail probability compensates for the growth of $(-\lambda)^j$. If L is the Binomial random variable with parameter $(k, (1+\lambda)^{-1})$ then (3) coincides with the Euler-smoothed estimator of Efron and Thisted (1976), with k being the truncation level of (3). As Euler's smoothing, the approach of Orlitsky et al. (2017) is distribution-free, in the sense that no conditions are imposed on the p_i's. For some choices of L, including the Binomial with parameter $(k, (1+\lambda)^{-1})$, Orlitsky et al. (2017) proved that $\hat{U}^L_{\lambda n}$ estimates $U_{\lambda n}$ all of the way up to $\lambda \propto \log(n)$, and that this range is the best possible.

In Figure 1 we illustrate the performance of $\hat{U}^L_{\lambda n}$ for a population of 10^6 species whose proportions p_i's are the masses of the ubiquitous Zipf distribution, i.e., $p_i \propto i^{-s}$ for some $s > 0$. The parameter s controls how the probability mass is spread among the species in the population: the lager s, the smaller is the number of species with high mass. In other terms the larger s, the heavier is the tail of the species proportions $(p_i)_{i\geq 1}$. A sample of $n = 5 \cdot 10^5$ individuals is taken from the population, and $\hat{U}^L_{\lambda n}$ is applied for any $\lambda \leq 20n$. Each panel of Figure 1 corresponds to a different choice of s, i.e. from left to right and top to bottom $s = 0.6, 0.8, 1.0, 1.2, 1.4, 1.6$. All experiments are averaged over 100 iterations. The true value is shown in black, and estimated values are colored according to the three choices of the distribution of L considered in Table 1 of Orlitsky et al. (2017): i) a Poisson distribution with parameter $(2\lambda)^{-1}\log_e(n(\lambda+1)^2/(\lambda-1))$; ii) a Binomial distribution with parameter $(2^{-1}\log_2(n\lambda^2/(\lambda-1)), (\lambda+1)^{-1})$; iii) a Binomial distribution with parameter $(2^{-1}\log_3(n\lambda^2/(\lambda-1)), 2(\lambda+2)^{-1})$. Shaded bands correspond to one standard deviation. Figure 1 highlights how the tail behaviour of the p_i's affects the experimental performance of the estimator $\hat{U}^L_{\lambda n}$: the heavier the tail of $(p_i)_{i\geq 1}$, or rather the lower the species discovery rate, the worse the performance of $\hat{U}^L_{\lambda n}$. This noticeable underestimation phenomenon thus suggests that the sole Euler's smoothing, and in general distribution-free smoothing, may not be useful for heavy-tailed p_i's.

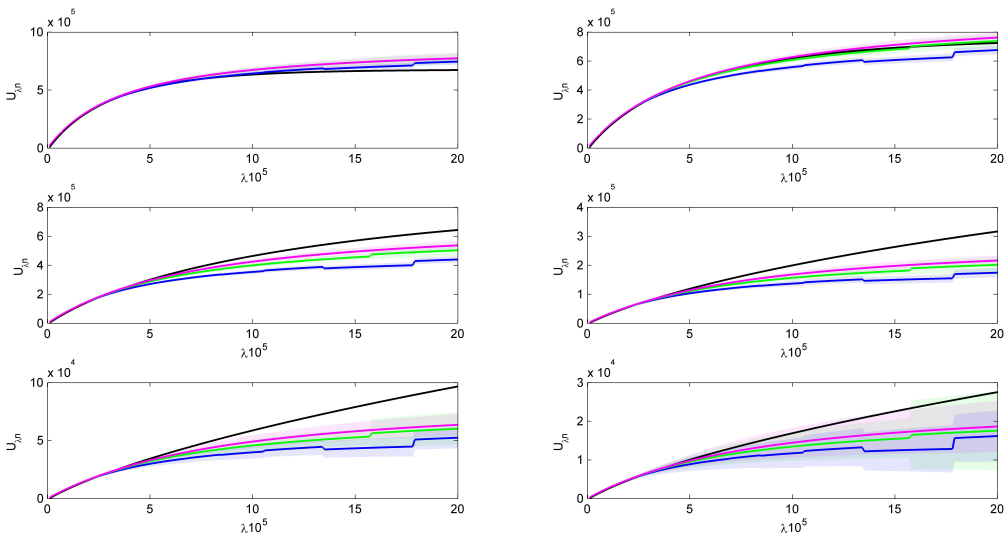

Figure 1: Estimator of $U_{\lambda n}$ in six Zipf scenarios. The true value is drawn in black, the estimated values are colored in blue (L being the Poisson distribution), green (L being the binomial distribution with success probability $1/(\lambda+1)$) and magenta (L being the binomial distribution with success probability $2/(\lambda+2)$). The shaded bands correspond to one standard deviation.

Besides estimating the total number of new species in additional samples, there is often a practical interest in estimating how many of these new species have a specified abundance. Such a refinement is crucial for understanding the composition of new species. In ecology and biology, for instance, conservation of biodiversity requires to control the

number of species with frequency less than a specific abundance threshold, the so-called rare species. See, e.g., Magurran (2003) and Thompson (2004). In genomics rare species is a fundamental issue, the reason being that species that appears only once or twice are often associated with deleterious diseases. See Laird and Lange (2010) and references therein. Within the "extrapolation approach", these problems calls for estimating

$$U_{\lambda n,r} := \sum_{i\geq 1} \mathbb{1}_{\{N_{n,i}=0\}} \mathbb{1}_{\{N_{\lambda n,i}=r\}}, \tag{4}$$

the number of hitherto unseen species that would be observed with frequency r in $(X_{n+1}, \ldots, X_{n+\lambda n})$, for $\lambda > 0$ and $1 \leq r \leq \lambda n$. One then set $U_{\lambda n,\overline{r}} := \sum_{1\leq r\leq \tau} U_{\lambda n,r}$ to be the number of new species with frequency less than or equal to τ. Clearly $U_{\lambda n} = U_{\lambda n,\overline{\lambda n}}$, thus showing that (4) is a refinement of (1). A quantity closely related to $U_{\lambda n,r}$ is

$$V_{\lambda n,r} := \sum_{i\geq 1} p_i \mathbb{1}_{\{N_{\lambda n+n,i}=r\}}, \tag{5}$$

the probability of discovering at the $(\lambda n+n+1)$-th draw a species with frequency r in the enlarged sample $(X_1, \ldots, X_n, X_{n+1}, X_{n+\lambda n})$, for $\lambda > 0$ and $0 \leq r \leq n + \lambda n$. In particular $V_{\lambda n,0}$ is the probability of discovering a new species. Note that (5) is well defined also for $\lambda = 0$, which is the classical discovery probability studied in Good (1953). For $\lambda > 0$ the estimation of $V_{\lambda n,r}$ is directly related to the problem of evaluating the cost-effectiveness of further sampling in problems where the sampling procedure is expensive. This is typically faced by setting a threshold ρ, which is related to the cost of sampling, and then estimating the sample size λn for which an estimate of $V_{\lambda n,r}$ falls below ρ.

While there exists a substantial amount of literature on the Good-Toulmin approach for estimating $U_{\lambda n}$, including the recent stimulating work of Orlitsky et al. (2017), we are not aware of any study investigating Good-Toulmin type estimators for $U_{\lambda n,r}$. In this paper we cover this gap. Our study relies upon the observation that the Good-Toulmin estimator $\hat{U}_{\lambda n}$ is nonparametric empirical Bayes in the sense Robbins (1956) originally attached to this term. See, e.g., Efron and Thisted (1976). Using Robbins' approach, we propose a nonparametric estimator $\hat{U}_{\lambda n,r}$ of $U_{\lambda n,r}$. The estimator of $U_{\lambda n,\overline{r}}$ is then $\hat{U}_{\lambda n,\overline{r}} := \sum_{1\leq r\leq \tau} \hat{U}_{\lambda n,r}$, whereas the Good-Toulmin estimator coincides with $\hat{U}_{\lambda n,\overline{\lambda n}}$. Not surprisingly $\hat{U}_{\lambda n,r}$ results to be useless for $\lambda \geq 1$, due to the same high variance issue affecting the Good-Toulmin estimator. We then show that one may still exploit Euler's smoothing in order to overcome this drawback, thus extending the range of predictability of $\hat{U}_{\lambda n,r}$ to $\lambda \geq 1$. In particular we introduce a truncation of a Euler's transformation of $\hat{U}_{\lambda n,r}$, we show that it estimates $U_{\lambda n,r}$ all of the way up $\lambda \propto \log_2(n)/2r \log_2(\log_2(n))$, and that this range is the best possible. An analogous study is presented with respect to the problem of estimating the discovery probability $V_{\lambda n,r}$, for any $\lambda > 0$.

The application of our estimators to the above Zipf's scenarios confirms the inadequacy of Euler's smoothing as the parameter $s \geq 1$ increases. We then consider tuning Euler-smoothed estimators by means of distribution-specific conditions on the tail behaviour of the species proportions $(p_i)_{i\geq 1}$. In particular we rely on Karlin (1967) theory of regular variation to give a general description of heavy-tailed p_i's. See, e.g., Gnedin et al. (2007), Ohannessian and Dahleh (2012) and Ben-Hamou et al. (2017) for recent contributions on

species sampling problems under regular variation. By combining Robbins' approach with Euler's smoothing and regular variation, we introduce a class of Good-Toulmin type estimators which allow to reduce the underestimation phenomenon displayed in Figure 1 for $s \geq 1$. Due to the Bayesian nature of Robbins' approach, the assumption of regular variation takes on the natural interpretation of a prior assumption on the species proportions $(p_i)_{i\geq 1}$. Interestingly, we show that our regularly varying estimators are asymptotically equivalent, for large n, to Bayesian nonparametric estimators of $U_{\lambda n,r}$ and $V_{\lambda n,r}$ recently proposed in Favaro et al. (2009) and Favaro et al. (2012) under the two parameter Poisson-Dirichlet prior for $(p_i)_{i\geq 1}$. See Pitman (2006) and references therein for a comprehensive account on such a nonparametric prior.

The paper is structured as follows. In Section 2 we introduce nonparametric empirical Bayes estimators of $U_{\lambda n,r}$ and $V_{\lambda n,r}$, we derive their Euler-smoothed versions, and we investigate properties of smoothed estimators to determine their best possible range of predictability. In Section 3 we apply Karlin's theory in order to tune our estimators under heavy-tailed species proportions $(p_i)_{i\geq 1}$, and we discuss the interplay between the resulting regularly varying estimators and their Bayesian nonparametric counterparts. In Section 4 we discuss the problem of deriving confidence intervals for the proposed classes of estimators. Proofs of our results are provided in the appendix.

2 Good-Toulmin type estimators

Among several species sampling models to which the "extrapolation approach" has been applied, the most common are the Multinomial and the Poisson abundance models. These models are closely related to each other because they both assume that the initial observable samples are independent and identically distributed according to the species proportions $(p_i)_{i\geq 1}$. However, while in the Multinomial abundance model the size of the initial sample is a fixed positive number n, in the Poisson abundance model such a sample size is assumed to be a Poisson random variable with parameter n. That is, under the Poisson abundance model the X_i's are independent and identically distributed according to $(p_i)_{i\geq 1}$, and result in K_N species with frequencies $(N_{N,1}, \ldots, N_{N,K_N})$ such that the sample size $\sum_{1\leq i\leq K_N} N_{N,i} = N$ is distributed as a Poisson distribution with parameter n. In particular $N_{N,i}$ given $N = n$ is Poisson distributed with parameter (np_i), and $N_{n,i}$ is independent of $N_{n,j}$ for any $i \neq j$. Following Efron and Thisted (1976), we present our methodology under the Poisson model. Similar results, but of a somewhat more tedious derivation, holds under the Multinomial model.

Under the Poisson abundance model, the Good-Toulmin estimator $\hat{U}_{\lambda n}$ is a nonparametric empirical Bayes estimator of $\mathbb{E}[U_{\lambda n}]$ in the sense of Robbins (1956). That is, $\hat{U}_{\lambda n}$ is the empirical posterior expectation of $\mathbb{E}[U_{\lambda n}]$ with respect to a prior on the unknown $(p_i)_{i\geq 1}$ which consists of the sole assumption that p_i, for each i, is independent and identically distributed from an unknown distribution function. We refer to Maritz and Lwin (1989) for a comprehensive account of empirical Bayes procedures. For any $\lambda < 1$, the use of $\hat{U}_{\lambda n}$ as an estimator of $U_{\lambda n}$ is then legitimated by the fact that $\mathbb{E}[U_{\lambda n} - \hat{U}_{\lambda n}] = 0$ and $\mathbb{E}[(U_{\lambda n} - \hat{U}_{\lambda n})^2] \lesssim n\lambda^2$. Here we apply Robbins' approach to estimate $U_{\lambda n,r}$ and $V_{\lambda n,r}$ for

$\lambda > 0$. This approach leads to estimate $\mathbb{E}[U_{\lambda n,r}]$ and $\mathbb{E}[V_{\lambda n,r}]$ by the formulae

$$\hat{U}_{\lambda n,r} := \lambda^r \sum_{j\geq 0} (-\lambda)^j \binom{r+j}{j} m_{n,j+r} \tag{6}$$

and

$$\hat{V}_{\lambda n,r} := \frac{1}{n}(r+1)(\lambda+1)^r \sum_{j\geq 0} (-\lambda)^j \binom{r+j+1}{j} m_{n,j+r+1}, \tag{7}$$

respectively. See A.1 for details on the derivation of (6) and (7). The Good-Toulmin estimator coincides with $\hat{U}_{\lambda n,\overline{\lambda n}}$. Note that formula (7) is well-defined also for $\lambda = 0$. In particular $\hat{V}_{0,r} = n^{-1}(r+1)m_{n,r+1}$ is the celebrated Good-Turing estimator for the probability of discovering at the $(n+1)$-th draw a species with frequency r in the observed sample. See the early works by Good (1953), Harris, B. (1968) and Robbins (1968), and more recently McAllester and Schapire (2000), Orlitsky et al. (2003), Ohannessian and Dahleh (2012) Mossel and Ohannessian (2015) and Ben-Hamou et al. (2017).

In the next theorem we show that $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$ are unbiased estimators of $U_{\lambda n,r}$ and $V_{\lambda n,r}$, respectively, besides we provide corresponding variance bounds under the assumption $\lambda < 1$. Unbiasedness of these estimators is not surprising, and in general it is a peculiar feature of the application of Robbins' approach under the Poisson abundance model. See, e.g., Maritz and Lwin (1989) for a thorough discussion. Variance bounds for $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$ can be easily derived by exploiting the independence of the frequency counts $N_{n,i}$'s. The resulting bounds are expressed as suitable functions of the expected value of $M_{n,r}$, i.e. $\mathbb{E}[M_{n,r}] = \sum_{i\geq 1} \mathrm{e}^{-p_i n}(p_i n)^r/r!$, and of the expected value of the cumulate counts $M_{n,\bar{r}} := \sum_{i\geq r} M_{n,i}$, for some choices of r and n. The derivation of variance bounds for $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$ represents a crucial step of our analysis. Indeed these bounds reveal that the assumption $\lambda < 1$ is necessary in order to obtain finite estimates of the variance. This is the first evidence that the estimators $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$ become useless for $\lambda \geq 1$ and, as discussed below, they require an adjustment via suitable smoothing techniques.

Theorem 1 *For any positive real numbers x and y let $\lfloor x \rfloor$ denote the integer part of x and let $x \vee y$ denote the maximum between x and y. Under the assumption that $\lambda < 1$,*

i) for any $r \geq 1$

$$\mathbb{E}[\hat{U}_{\lambda n,r}] = \mathbb{E}[U_{\lambda n,r}] = \frac{(\lambda n)^r}{r!} \sum_{i\geq 1} p_i^r e^{-np_i(\lambda+1)}$$

and

$$Var[U_{\lambda n,r} - \hat{U}_{\lambda n,r}] = \mathbb{E}[(U_{\lambda n,r} - \hat{U}_{\lambda n,r})^2] \leq \Psi^2_{\lambda,r}\mathbb{E}[M_{n,\bar{r}}] + \frac{\mathbb{E}[M_{\lambda n+n,r}]}{\left(\frac{\lambda+1}{\lambda}\right)^r}, \tag{8}$$

where $\Psi_{\lambda,r} = \binom{r+j_1^*}{r}\lambda^{j_1^*+r}$ and $j_1^* := \lfloor r\lambda/(1-\lambda) - 1 \rfloor \vee 0$;

ii) for any $r \geq 0$

$$\mathbb{E}[\hat{V}_{\lambda n,r}] = \mathbb{E}[V_{\lambda n,r}] = \frac{(n(\lambda+1))^r}{r!} \sum_{i \geq 1} p_i^{r+1} e^{-np_i(\lambda+1)}$$

and

$$Var[V_{\lambda n,r} - \hat{V}_{\lambda n,r}] = \mathbb{E}[(V_{\lambda n,r} - \hat{V}_{\lambda n,r})^2] \leq \frac{\Phi^2_{\lambda,r}}{n^2} \mathbb{E}[M_{n,\overline{r+1}}] + \frac{\mathbb{E}[M_{\lambda n+n,r+2}]}{\frac{(\lambda n+n)^2}{(r+1)(r+2)}}, \tag{9}$$

where $\Phi_{\lambda,r} := (r+1)(\lambda+1)^r(-\lambda)^{j_2^*}\binom{r+j_2^*+1}{r+1}$ *and* $j_2^* := \lfloor(\lambda(r+2)-1)/(1-\lambda)\rfloor \vee 0$.

See A.2 for the proof of Theorem 1. For any $\lambda < 1$ Theorem 1 legitimates the use of $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$ as estimators $U_{\lambda n,r}$ and $V_{\lambda n,r}$, respectively. Indeed, according to the variance bounds displayed in (8) and (9) one has $\mathbb{E}[(U_{\lambda n,r} - \hat{U}_{\lambda n,r})^2] \lesssim n$ and $\mathbb{E}[(V_{\lambda n,r} - \hat{V}_{\lambda n,r})^2] \lesssim 1/n$ upon noticing that $\mathbb{E}[M_{n,\overline{r}}] \leq n/r$ and $\mathbb{E}[M_{n,r}] \leq n/r$. These bounds show that, in expectation, $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$ approximate $U_{\lambda n,r}$ and $V_{\lambda n,r}$ to within $n^{1/2}$ and $n^{-1/2}$ respectively. Similarly, $\hat{U}_{\lambda n,\overline{\tau}}$ is a legitimate estimator of $U_{\lambda n,\overline{\tau}}$. In particular, by a direct application of Theorem 1 and the discrete Hölder inequality we can write

$$\mathrm{Var}[U_{\lambda n,\overline{\tau}} - \hat{U}_{\lambda n,\overline{\tau}}] \leq \tau \sum_{r=1}^{\tau} \mathrm{Var}[U_{\lambda n,r} - \hat{U}_{\lambda n,r}] \leq \tau \sum_{r=1}^{\tau} \left[\Psi^2_{\lambda,r} \mathbb{E}[M_{n,\overline{r}}] + \frac{\mathbb{E}[M_{\lambda n+n,r}]}{\left(\frac{\lambda+1}{\lambda}\right)^r} \right].$$

Corresponding results for the statistic $V_{\lambda n,\overline{\tau}} := \sum_{1 \leq r \leq \tau} V_{\lambda n,r}$ may be derived in a similar fashion. In Section 4 we discuss how the variance bounds of Theorem 1 can be usefully applied, via standard Chernoff bounding, to obtain confidence intervals for $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$.

We now consider the problem of estimating $U_{\lambda n,r}$ and $V_{\lambda n,r}$ for $\lambda \geq 1$. For $\lambda \geq 1$ the estimator $\hat{U}_{\lambda n,r}$ is useless, due to its high variance. As for the Good-Toulmin estimator $\hat{U}_{\lambda n}$, this high variance issue is determined by the geometrically increasing magnitude of the coefficients $(-\lambda)^j\binom{r+j}{j}$. Indeed for any fixed r, as $\lambda \geq 1$ increases, $\hat{U}_{\lambda n,r}$ grows superlinearly as $(-\lambda)^j\binom{r+j}{j}$ for the largest j such that $m_{n,j+r} > 0$, thus eventually far exceeding $U_{\lambda n,r}$ that grows at most linearly. Clearly this undesirable phenomenon worsens as r increases. Similar reasons motivate the inaccuracy of $\hat{V}_{\lambda n,r}$ for $\lambda \geq 1$. In order to overcome this drawback we follow Efron and Thisted (1976) and truncate a Euler's transformation of the series (6) and (7). This leads to estimate $U_{\lambda n,r}$ and $V_{\lambda n,r}$ by

$$\hat{U}^L_{\lambda n,r} := \lambda^r \sum_{j \geq 0} (-\lambda)^j \binom{r+j}{j} \mathbb{P}[L \geq r+j] m_{n,j+r} \tag{10}$$

and

$$\hat{V}^I_{\lambda n,r} := \frac{1}{n}(r+1)(\lambda+1)^r \sum_{j \geq 0} (-\lambda)^j \binom{r+j+1}{j} \mathbb{P}[I \geq r+j+1] m_{n,j+r+1}, \tag{11}$$

respectively, where L and I are two independent Binomial random variables with corresponding parameters $(k_L + r, (\lambda+1)^{-1})$ and $(k_I + r + 1, (\lambda+1)^{-1})$. See A.3 for details

on the derivation of (10) and (11). For any fixed r, as the index j increases, the Binomial right-tail probability decay compensates the geometric grow of the coefficients $(-\lambda)^j\binom{r+j}{j}$. Note that the Binomial right-tail probability is also affected by the value of r in order to take into account the fact that geometric grow of $(-\lambda)^j\binom{r+j}{j}$ sharpen as r increases. In the next theorem we show that $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^I_{\lambda n,r}$ are biased estimators of $U_{\lambda n,r}$ and $V_{\lambda n,r}$, respectively. We then provide corresponding bounds for the variance and the bias.

Theorem 2 *Let L and I be two independent Binomial random variables with parameters $(k_L+r,(\lambda+1)^{-1})$ and $(k_I+r+1,(\lambda+1)^{-1})$, respectively. Under the assumption that $\lambda\geq 1$,*

i) for any $r\geq 1$,

$$\mathbb{E}[\hat{U}^L_{\lambda n,r}]=\mathbb{E}[U_{\lambda n,r}]-\lambda^r\sum_{i\geq 1}e^{-p_in}\frac{(p_in)^r}{r!}\sum_{j\geq 0}\frac{\mathbb{P}(L\leq j+r-1)}{j!}(-\lambda p_in)^j,$$

$$Var[\hat{U}^L_{\lambda n,r}-U_{\lambda n,r}]\leq 4^{k_L}\left[\frac{\binom{k_L+r}{k_L}}{\left(\frac{\lambda+1}{\lambda}\right)^{(k_L+r)}}\right]^2\mathbb{E}[M_{n,\overline{r}}]+\frac{\mathbb{E}[M_{\lambda n+n,r}]}{\left(\frac{\lambda+1}{\lambda}\right)^r} \tag{12}$$

and

$$\begin{aligned}&|\mathbb{E}[\hat{U}^L_{\lambda n,r}-U_{\lambda n,r}]| \\ &\leq\frac{2^{r+1}\left(\frac{\lambda}{\lambda+1}\right)^{r+k_L+1}\frac{\lambda+1}{1+2\lambda}}{\binom{k_L+r}{r}^{-1}}\mathbb{E}[M_{n/2,r}]+\frac{\mathbb{P}(L\leq r-1)\mathbb{E}[M_{\lambda n+n,r}]}{\left(\frac{\lambda+1}{\lambda}\right)^r}\end{aligned} \tag{13}$$

ii) for any $r\geq 0$

$$\mathbb{E}[\hat{V}^I_{\lambda n,r}]=\mathbb{E}[V_{\lambda n,r}]-\frac{(r+1)(\lambda+1)^r}{n}\sum_{i\geq 1}e^{-np_i}\frac{(np_i)^{r+1}}{(r+1)!}\sum_{j\geq 0}\mathbb{P}(L\leq j+r)\frac{(-\lambda np_i)^j}{j!},$$

$$Var[\hat{V}^I_{\lambda n,r}-V_{\lambda n,r}]\leq\left[\frac{\binom{k_I+r+1}{k_I}}{\left(\frac{\lambda+1}{\lambda}\right)^{k_I}\frac{\lambda+1}{(r+1)2^{k_I}}}\right]^2\mathbb{E}[M_{n,\overline{r+1}}]+\frac{\mathbb{E}[M_{\lambda n+n,r+2}]}{\frac{(\lambda+1)^2n^2}{(r+1)(r+2)}} \tag{14}$$

and

$$\begin{aligned}&|\mathbb{E}[\hat{V}^I_{\lambda n,r}-V_{\lambda n,r}]| \\ &\leq\frac{\frac{r+1}{n}\left(\frac{\lambda}{\lambda+1}\right)^{k_I+1}\frac{2^{r+2}(\lambda+1)}{2\lambda+1}}{\binom{k_I+r+1}{k_I}^{-1}}\mathbb{E}[M_{n/2,r+1}]+\frac{\mathbb{P}(I\leq r)\mathbb{E}[M_{\lambda n+n,r+1}]}{\frac{n(\lambda+1)}{r+1}}.\end{aligned} \tag{15}$$

The mean square error (MSE) bounds for $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^I_{\lambda n,r}$ can be obtained by combining (12) with (13) and (14) with (15), respectively. See A.4 and A.5 for details. Noticing that $\mathbb{E}[M_{n,\bar{r}}] \leq n/r$ and $\mathbb{E}[M_{n,r}] \leq n/r$, these MSE bounds provide with a fundamental guideline for tuning k_L and k_I in such a way to achieve the best estimates of $U_{\lambda n,r}$ and $V_{\lambda n,r}$. In order to chose the parameters k_L and k_I, we look for estimators $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^I_{\lambda n,r}$ with vanishing normalized mean square error (NMSE). A similar approach has been considered in Orlitsky et al. (2017) with respect to the estimation of $U_{\lambda n}$. More precisely, since $0 \leq U_{\lambda n,r} \leq \lambda n/r$, we define the NMSE for the estimator $\hat{U}^L_{\lambda n,r}$ as

$$\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) := \mathbb{E}\left[\left(\frac{U_{\lambda n,r} - \hat{U}^L_{\lambda n,r}}{\lambda n/r}\right)^2\right], \tag{16}$$

which turns out to be the MSE divided by the maximum value of the statistic $U_{\lambda n,r}$. Similarly, since $0 \leq V_{\lambda n,r} \leq 1$ the NMSE for $\hat{V}^I_{\lambda n,r}$ coincides with the MSE, and hence we define

$$\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) := \mathbb{E}[(V_{\lambda n,r} - \hat{V}^I_{\lambda n,r})^2]. \tag{17}$$

NMSE bounds for (16) and (17) follow from Theorem (2). We then choose k_L and k_I which minimize the NMSE bounds. Since the analytic minimization of these bounds is not possible, we propose to choose the parameters k_L and k_I in such a way to obtain the best rate of convergence to zero of the NMSE bounds. See A.6 for details. Among all the possible values of k_L and k_I, we consider $k_L = k_I = 2^{-1}\log_2(n)$, thus avoiding the dependence on r and λ. In the next theorem we use this choice of k_L and k_I in order to determine NMSE bounds for $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^I_{\lambda n,r}$ vanishing as $n \to +\infty$ at the best possible rate.

Theorem 3 *Let $\lambda > 1$, and let $A_r(\lambda)$ and $B_r(\lambda)$ be continuous functions of λ, bounded on $[1, +\infty)$. Then,*

i) for any $r \geq 1$

$$\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq A_r(\lambda)\frac{\log_2^{2r}(n)}{n^{\log_2(1+1/\lambda)}}; \tag{18}$$

ii) for any $r \geq 0$

$$\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) \leq B_r(\lambda)\frac{\log_2^{2r+2}(n)}{n^{\log_2(1+1/\lambda)}}. \tag{19}$$

Orlitsky et al. (2017) showed that $\hat{U}^L_{\lambda n}$ estimates $U_{\lambda n}$ all of the way up to $\lambda \propto \log(n)$, and that this range is the best possible. Theorem 3 allows us to obtain an analogous result for the estimators $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^L_{\lambda n,r}$. In particular we obtain explicit limits of predictability of $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^I_{\lambda n,r}$, that is for some $\delta > 0$ we obtain the maximum value of λ such that $\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) < \delta$ and the maximum value of λ such that $\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) < \delta$. This provides a

theoretical guarantee for our Euler's smoothed estimators. The next corollary shows that $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^I_{\lambda n,r}$ estimates $U_{\lambda n,r}$ and $V_{\lambda n,r}$, respectively, all of the way up to

$$\lambda \propto \frac{\log_2(n)}{2r\log_2(\log_2(n))}. \tag{20}$$

Formula (20) shows that the limit of predictability of $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^L_{\lambda n,r}$ reduces as r increases. This behaviour is not surprising since, as discussed before, the high variance phenomenon affecting $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$ worsens as r increases. In general, for any $r \geq 1$ the limit of predictability of $\hat{U}^L_{\lambda n,r}$ is less than the limit of predictability of $\hat{U}^L_{\lambda n,r}$ as $n > 3$. A simulation study at the end of this section illustrates the behaviour of limit of predictability.

Corollary 4 *Let $\lambda \geq 1$ and let $A_r := \max_{\lambda\geq 1} A_r(\lambda) < +\infty$ and $B_r := \max_{\lambda\geq 1} B_r(\lambda) < +\infty$. Then,*

i) for any $r \geq 1$ and $\delta > 0$

$$\lim_{n\to+\infty} \max\left\{\lambda:\ \mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq \delta\right\} \times \frac{2r\log_2(\log_2(n)) + \log_2(A_r) + \log_2(1/\delta)}{\log_2(n)} \geq \frac{1}{\log(2)}; \tag{21}$$

ii) for any $r \geq 0$ and $\delta > 0$

$$\lim_{n\to+\infty} \max\left\{\lambda:\ \mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) \leq \delta\right\} \times \frac{2(r+1)\log_2(\log_2(n)) + \log_2(B_r) + \log_2(1/\delta)}{\log_2(n)} \geq \frac{1}{\log(2)}. \tag{22}$$

Similarly to case $\lambda < 1$ discussed above, for $\lambda \geq 1$ one can handle the estimation of the cumulative statistics $U_{\lambda n,\overline{\tau}}$ and $V_{\lambda n,\overline{\tau}}$ through a suitable adaptation of the results in Theorem 2 and Theorem 3. For the sake of explanation we focus on the statistic $U_{\lambda n,\overline{\tau}}$. In particular, for $\lambda \geq 1$ the estimator for $U_{\lambda n,\overline{\tau}}$ is $\hat{U}^L_{\lambda n,\overline{\tau}} := \sum_{1\leq r\leq\tau} \hat{U}^L_{\lambda n,r}$, with $\hat{U}^L_{\lambda n,r}$ being in (10). Moreover, noticing that $\mathbb{E}[M_{n,\overline{r}}] \leq n/r$ and $\mathbb{E}[M_{n,r}] \leq n/r$, an application of Theorem 2 and the discrete Hölder inequality yields to the following MSE bound:

$$\begin{aligned}
&\mathrm{MSE}[U_{\lambda n,\overline{\tau}} - \hat{U}^L_{\lambda n,\overline{\tau}}] \\
&\leq \tau \sum_{r=1}^{\tau} \left[\left(\frac{\lambda}{\lambda+1}\right)^{r-1} \frac{\lambda n}{r} + 2^{2k_L} \left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \binom{k_L+r}{r}^2 \frac{n}{r} \right. \\
&\quad \left. + \frac{\lambda^2 n^2}{r^2} \left[\left(\frac{\lambda}{\lambda+1}\right)^{r-1} \mathbb{P}(L \leq r-1) + \frac{2^r}{1+2\lambda} \left(\frac{\lambda}{\lambda+1}\right)^{q+k_L} \binom{k_L+q}{q} \right]^2 \right].
\end{aligned}$$

This bound constitutes the basic building block to find the limit of predictability of $\hat{U}^L_{\lambda n,\bar{\tau}}$. In particular it can be easily verified that $\hat{U}^L_{\lambda n,\bar{\tau}}$ estimates $U_{\lambda n,\bar{\tau}}$ all of the way up to

$$\lambda \propto \frac{\log_2(n)}{2\tau \log_2(\log_2(n))},$$

which coincides with the limit of predictability of the largest frequency, i.e., τ. This is because, by means of Theorem 3, the bound on the normalized mean square error corresponds to $\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,\bar{\tau}}) \leq C_\tau(\lambda) \log_2^{2\tau}(n) n^{-\log_2(1+1/\lambda)}$, for an appropriate bounded function $C_\tau(\lambda)$.

We apply the estimator $\hat{U}^L_{\lambda n,r}$, for $r = 1$ and $r = 2$, to the Zipf's scenarios presented in the Introduction. Figure 2 illustrates the performance of $\hat{U}^L_{\lambda n,1}$, whereas Figure 3 illustrates the performance of $\hat{U}^L_{\lambda n,2}$. For both figures each panel corresponds to a different choice for the parameter s of the Zipf distribution, i.e. from left to right and top to bottom $s = 0.6, 0.8, 1.0, 1.2, 1.4, 1.6$. All experiments are averaged over 100 iterations. The true value is shown in black, the estimated value in red, and the shaded band corresponds to one standard deviation. From a comparative analysis of Figure 2 and Figure 3 emerges that: i) there is a set of values for λ for which the performance of $\hat{U}^L_{\lambda n,1}$ is worsen than $\hat{U}^L_{\lambda n}$; ii) there exists a set of values for λ for which the performance of $\hat{U}^L_{\lambda n,2}$ is worsen than the performance of $\hat{U}^L_{\lambda n,1}$. This confirms our result in Corollary 4, that is the limit of predictability of $\hat{U}^L_{\lambda n,r}$ reduces as r increases, and the limit of predictability of $\hat{U}^L_{\lambda n,1}$ is smaller than the limit of predictability of $\hat{U}^L_{\lambda n}$. Both Figure 2 and Figure 3 show how the tail behaviour of $(p_i)_{i\geq 1}$ affects the performance of $\hat{U}^L_{\lambda n,1}$ and $\hat{U}^L_{\lambda n,2}$. Similarly to what we observed in Figure 1 for the estimation of $U_{\lambda n}$, such a performance worsen as $s \geq 1$ increases. As we pointed out in the Introduction, this behaviour suggests that the sole Euler's smoothing may not be useful for heavy-tailed $(p_i)_{i\geq 1}$.

We conclude with an application of the estimator $\hat{V}^I_{\lambda n,0}$. This is the estimator of the probability $V^I_{\lambda n,0}$ of discovering at the $(\lambda n + n + 1)$-th draw a new species. As recalled in the Introduction, the estimation of $V^I_{\lambda n,0}$ is directly related to the problem of determining the optimal sample size in species sampling problems. See Chao et al. (2009) for details. This is motivated by concrete applied problems where the sampling procedure is expensive, and further draws can be only motivated by the possibility of recording a new unobserved species. Knowledge of the cost associated with sampling might suggest a threshold ρ below which the sampling procedure is no more convenient. Hence, one can fix ρ and then determine λ such that $\hat{V}^I_{\lambda n,0}$ becomes for the first time smaller than ρ. This procedure introduces a criterion for evaluating the cost-effectiveness of further sampling. Figure 4 illustrates the performance of $\hat{V}^I_{\lambda n,0}$ under the Zipf's scenarios presented in the Introduction. All experiments are averaged over 100 iterations. The true value is shown in black, the estimated value in red, and the shaded band corresponds to one standard deviation. Not surprisingly, the performance of $\hat{V}^I_{\lambda n,0}$ worsen as $s \geq 1$ increases.

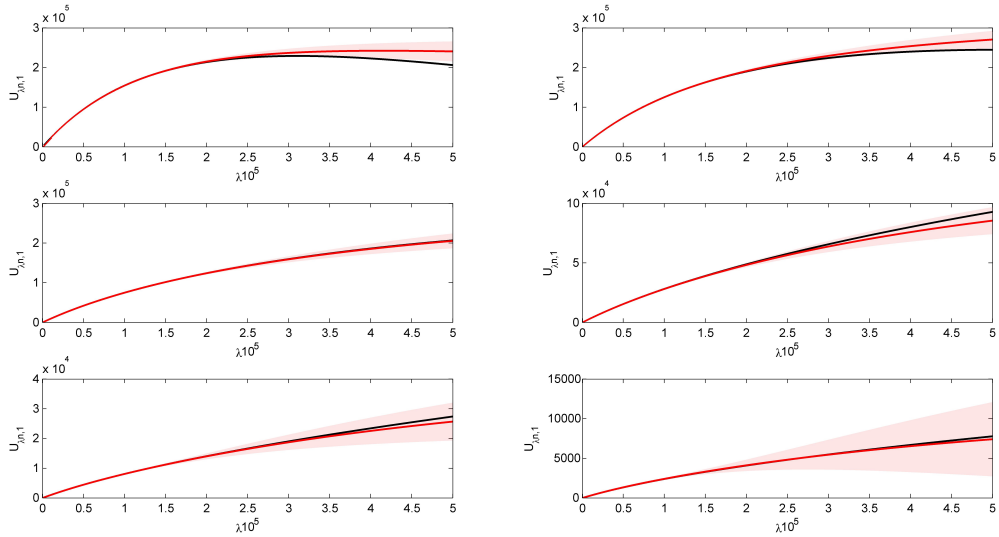

Figure 2: Estimator of $U_{\lambda n,1}$ in six Zipf scenarios. The true value is drawn in black, the estimated value in red. The shaded bands correspond to one standard deviation.

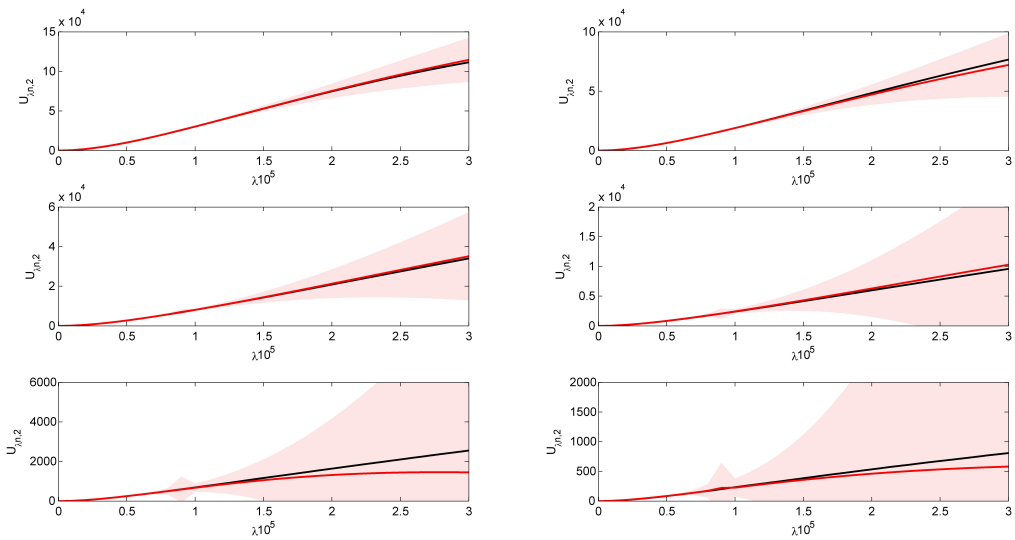

Figure 3: Estimator of $U_{\lambda n,2}$ in six Zipf scenarios. The true value is drawn in black, the estimated value in red. The shaded bands correspond to one standard deviation.

3 Regularly varying Good-Toulmin type estimators

The application of the estimators $\hat{U}^L_{\lambda n}$ and $\hat{U}^L_{\lambda n,r}$ to the various Zipf's scenarios suggests that the sole Euler's smoothing, and in general distribution-free smoothing, may not be

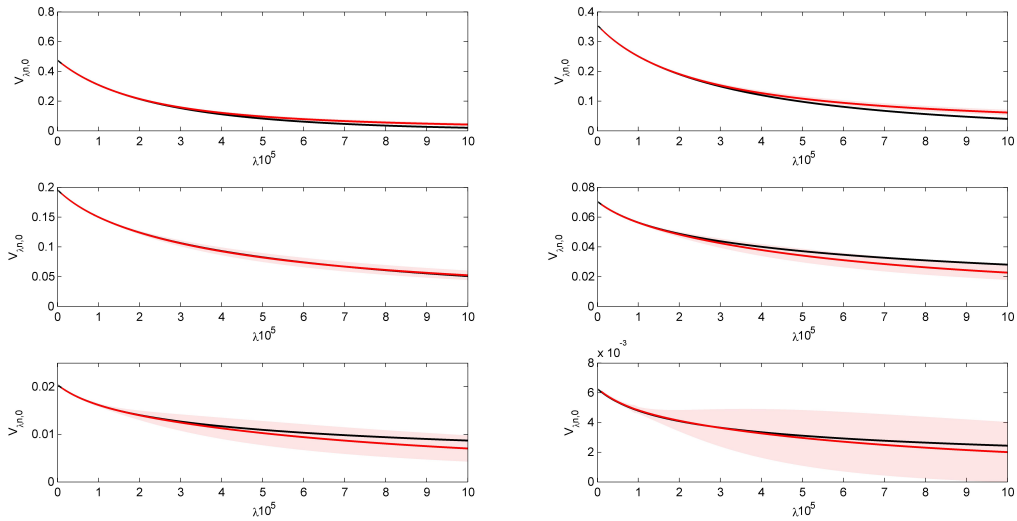

Figure 4: Estimator of $V_{\lambda n,0}$ in six Zipf scenarios. The true value is drawn in black, the estimated value in red. The shaded bands correspond to one standard deviation.

useful in the context of heavy-tailed $(p_i)_{i\geq 1}$. Indeed, as we pointed out in the Introduction, it turns up the following behaviour: the heavier the tail of $(p_i)_{i\geq 1}$, or rather the lower the rate of species discovery, the worse the underestimation of $\hat{U}^L_{\lambda n}$ and $\hat{U}^L_{\lambda n,r}$. Not surprisingly the same phenomenon holds for $\hat{V}^I_{\lambda n,r}$. This leads us to introduce an alternative class of Good-Toulmin type estimator under distribution-specific conditions for the tail behaviour of the p_i's. In particular we rely on Karlin (1967) theory of regular variation to give a general description of heavy-tailed $(p_i)_{i\geq 1}$. We use the limiting notation $f \sim g$ to mean $f/g \to 1$, and we use the superscript a.s. to indicate almost sure convergence with random quantities. Let $\nu(\mathrm{d}x) := \sum_{i\geq 1} \delta_{p_i}(\mathrm{d}x)$ and define the measure $\overline{\nu}(x) := \nu[x,1]$, which is the cumulative count of all species having no less than a certain probability mass. We say that $(p_i)_{i\geq 1}$ is regularly varying with regular variation index $\alpha \in (0,1)$ if $\overline{\nu}(x) \sim x^{-\alpha}\ell(1/x)$ as $x \downarrow 0$, where $\ell(t)$ is a slowly varying function, that is $\ell(ct)/\ell(t) \to 1$ as $t \to +\infty$ for all $c > 0$. The next theorem from Karlin (1967) provides with two conditions that are equivalent to regular variation.

Theorem 5 (Karlin (1967)) *Let $(p_i)_{i\geq 1}$ be regularly varying with $\alpha \in (0,1)$, and let $(X_1,\ldots,X_n)$ be a random sample from $(p_i)_{i\geq 1}$. If $\Gamma(\cdot)$ denotes the Gamma function, then as $n \to +\infty$*

i) $K_n \overset{a.s.}{\sim} \mathbb{E}[K_n] \sim \Gamma(1-\alpha)n^\alpha \ell(n)$,

ii) $M_{n,r} \overset{a.s.}{\sim} \mathbb{E}[M_{n,r}] \sim \frac{\alpha\Gamma(r-\alpha)}{r!} n^\alpha \ell(n)$

With additional care, it is possible to extend Theorem 5 to $\alpha = 0$ (slow variation, light tail) and $\alpha = 1$ (rapid variation, very heavy tail), as is done in Karlin (1967) and

Gnedin et al. (2007). For $\alpha \in (0,1)$, Theorem 5 shows how regular variation implies that $M_{n,r}$ is very well behaved, especially in terms of its strong laws, i.e., $M_{n,r}/\mathbb{E}[M_{n,r}] \to 1$ almost surely as $n \to +\infty$. Note that $M_{n,r}$ and $\mathbb{E}[M_{n,r}]$ are the main ingredients for the estimators derived in Section 2 and for their corresponding variance and MSE bounds. Theorem 5 also provides with a practical tool for estimating the regular variation index α. Indeed it turns out that the ratio of the number of species with frequency 1 to the total number of species defines a consistent estimator of α, namely

$$\hat{\alpha} := \frac{M_{n,1}}{K_n} \to \alpha$$

almost surely, as $n \to +\infty$. Note that this is not the only approach to estimating the regular variation index. Alternative estimators of α may be obtained by relying on techniques developed in extreme value theory, where regular variation is also pivotal. See, e.g., Beirlant et al. (2004), de Haan and Ferreira (2006) and references therein for details.

We now consider combining Robbins' approach with Euler's smoothing and regular variation. We focus on the estimation of $U_{\lambda n,r}$, although the same arguments can be applied to the estimation of $V_{\lambda n,r}$. Our approach is in two stage. We first apply Theorem 5 to tune the estimators $U_{\lambda n,r}$ and $\hat{U}^L_{\lambda n,r}$ under regularly varying heavy-tailed $(p_i)_{i\geq 1}$. We then choose an estimator $\hat{\alpha}$ for the regular variation index α. Let $(a)_n$ denote the falling factorial of order n of a, i.e., $(a)_n := a(a-1)\cdots(a-n+1)$ with the proviso $(a)_0 = 1$. The application of Theorem 5 leads to the following regularly varying estimators of $U_{\lambda n,r}$:

i) for $\lambda < 1$

$$\hat{U}_{\lambda n,r}(\alpha) := \frac{\alpha(r-1-\alpha)_{(r-1)}}{r!} k_n \left[r\lambda^r \sum_{j=0}^{n-r} \frac{\binom{j+r}{r}\binom{r+j-\alpha-1}{r+j-1}}{(r+j)\binom{r-1-\alpha}{r-1}} (-1)^j \lambda^j \right]; \tag{23}$$

ii) for $\lambda \geq 1$

$$\hat{U}^L_{\lambda n,r}(\alpha) := \frac{\alpha(r-1-\alpha)_{(r-1)}}{r!} k_n \left[\left(\frac{\lambda}{\lambda+1}\right)^r \sum_{z=0}^{k_L} \binom{z+\alpha-1}{z} \left(\frac{\lambda}{\lambda+1}\right)^z \right] \tag{24}$$

as $n \to +\infty$, respectively. See A.9 for details on the derivation of (23) and (24). By a direct application of (23) and (24) one obtains the following regularly varying estimators of $U_{\lambda n}$:

i) for $\lambda < 1$

$$\hat{U}_{\lambda n}(\alpha) := \sum_{r\geq 1} \hat{U}_{\lambda n,r}(\alpha) = -\alpha k_n \sum_{j=1}^{n} (-1)^j \frac{\binom{j-\alpha-1}{j-1}}{j} \lambda^j;$$

ii) for $\lambda \geq 1$

$$\hat{U}^L_{\lambda n}(\alpha) := \sum_{r\geq 1} \hat{U}^L_{\lambda n,r}(\alpha) = k_n \sum_{z=1}^{k_L} \binom{z+\alpha-1}{z} \left(\frac{\lambda}{\lambda+1}\right)^z.$$

We may consider regular variation as a non-distribution-free smoothing. Indeed it transforms $\hat{U}_{\lambda n,r}$ and $\hat{U}^L_{\lambda n,r}$ into new estimators by imposing an assumption on the tail behaviour of the p_i's. According to (23) and (24), the regular variation assumption results in a change of the information used for estimating $U_{\lambda n,r}$: while the estimators $\hat{U}_{\lambda n,r}$ and $\hat{U}^L_{\lambda n,r}$ have $(M_{n,r}, \ldots, M_{n,n})$ as sufficient statistic, the corresponding regularly varying estimators have K_n as sufficient statistic. It can be easily verified that $\hat{U}_{\lambda n,r}(\alpha)$ is a biased estimator of $U_{\lambda n,r}$.

We apply the estimators $\hat{U}^L_{\lambda n,1}(\alpha)$ and $\hat{U}^L_{\lambda n}(\alpha)$ to the Zipf's scenarios presented in the Introduction. We choose $\hat{\alpha} := M_{n,1}/K_n$ as an estimator for α, although better estimators of the regular variation index may be considered. Figure 5 illustrates the performance of $\hat{U}^L_{\lambda n}(\hat{\alpha})$, whereas Figure 6 illustrates the performance of $\hat{U}^L_{\lambda n,1}(\hat{\alpha})$. For both figures each panel corresponds to a different choice for the parameter s of the Zipf distribution, i.e. from left to right and top to bottom $s = 0.6, 0.8, 1.0, 1.2, 1.4, 1.6$. All experiments are averaged over 100 iterations. The true value is shown in black, the estimated value in red, and the shaded band corresponds to one standard deviation. A comparison between Figure 1 and Figure 5 show how the performance of the regularly varying estimator $\hat{U}^L_{\lambda n}(\hat{\alpha})$ behaves opposite to the performance of the Euler-smoothed estimator $\hat{U}^L_{\lambda n}$. That is, the heavier the tail of $(p_i)_{i\geq 1}$, or rather the lower the species discovery rate, the better the performance of $\hat{U}^L_{\lambda n}(\hat{\alpha})$. A comparison between Figure 2 and Figure 6 highlights the same opposite behaviour in the performances of $\hat{U}^L_{\lambda n,1}$ and $\hat{U}^L_{\lambda n,1}(\hat{\alpha})$.

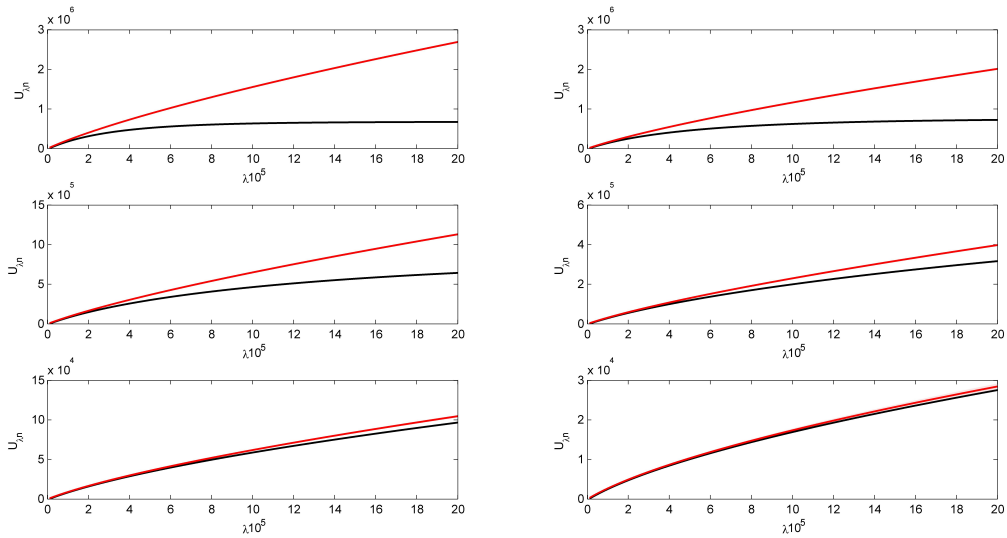

Figure 5: Estimator of $\hat{U}^L_{\lambda n}(\alpha)$ in six Zipf scenarios. The true value is drawn in black, the estimated value in red. The shaded bands correspond to one standard deviation.

We conclude this section with an intriguing link between the above regularly varying estimators and their Bayesian nonparametric counterparts introduced in Favaro et al. (2009) and Favaro et al. (2012). Differently from Robbin's approach, the Bayesian nonparametric

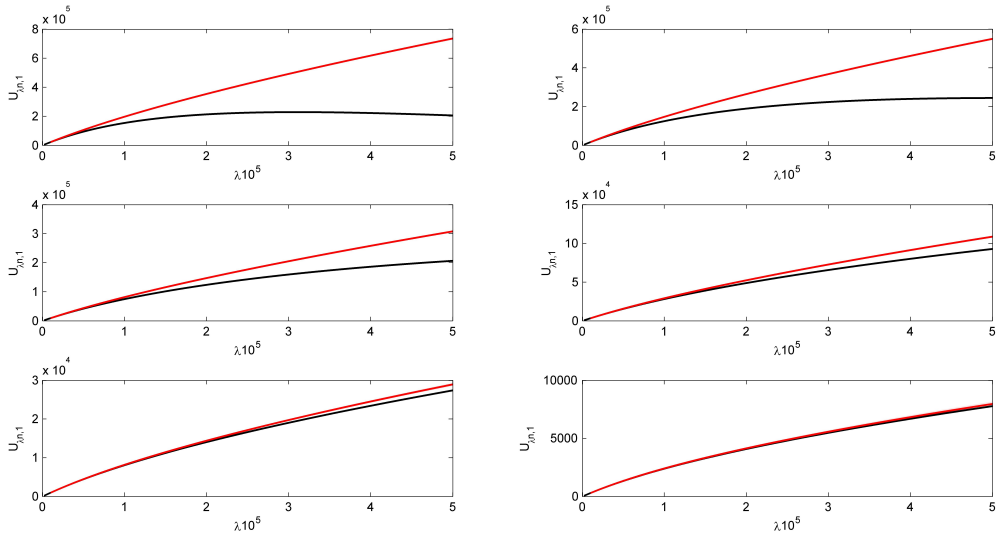

Figure 6: Estimator of $\hat{U}^L_{\lambda n,1}(\alpha)$ in six Zipf scenarios. The true value is drawn in black, the estimated value in red. The shaded bands correspond to one standard deviation.

approach for estimating $U_{\lambda n,r}$ relies on the specification of a species sampling prior for the unknown species proportions $(p_i)_{i\geq 1}$. This prior corresponds to the law of the discrete random probability measure $P = \sum_{i\geq 1} p_i \delta_{S_i}$, where the p_i's are nonnegative random weights such that $\sum_{i\geq 1} p_i = 1$ almost surely, whereas the S_i's are random labels independent of $(p_i)_{i\geq 1}$ and independent and identically distributed according to a nonatomic distribution. See Pitman (2006) for details. Then, under the Bayesian nonparametric approach, $(X_1, \ldots, X_n)$ is assumed to be a random sample from, i.e.,

$$\begin{aligned} X_i \,|\, P &\overset{\text{iid}}{\sim} P, \qquad i = 1, \ldots, n, \\ P &\sim \mathscr{P}, \end{aligned} \tag{25}$$

where $\mathscr{P}$ takes on the interpretation of the prior distribution on $(p_i)_{i\geq 1}$. A common choice for $\mathscr{P}$ is the law of the parameter Poisson-Dirichlet process, which was introduced in Perman et al. (1992) and further investigated in Pitman and Yor (1997). This is a species sampling prior whose random probabilities $(p_i)_{i\geq 1}$ are defined as follows: $p_1 = v_1$ and $p_i = v_i \prod_{1\leq j\leq i-1}(1 - v_j)$, for any $i \geq 2$, where the v_j's are independent random variables with v_j distributed according to a Beta distribution with parameter $(1 - \alpha, \theta + j\alpha)$, for $\alpha \in (0, 1)$ and $\theta > -\alpha$. We denote by $P_{\alpha,\theta}$ the two parameter Poisson-Dirichlet process.

Under the Bayesian nonparametric framework (25), with $\mathscr{P}$ being the law of $P_{\alpha,\theta}$, Favaro et al. (2009) and Favaro et al. (2012) characterized the posterior distributions of $U_{\lambda n,r}$ and $U_{\lambda n}$, given an initial observable sample $(X_1, \ldots, X_n)$. See also Lijoi et al. (2007) for details. Bayesian nonparametric estimators of $U_{\lambda n,r}$ and $U_{\lambda n}$, with respect to a squared loss function, are then given by the corresponding posterior expectations. Assuming $(X_1, \ldots, X_n)$ featuring $K_n = k_n$ species, for any $\lambda > 0$ these posterior expectations

are

$$\tilde{U}_{\lambda n,r}(\alpha,\theta) := \binom{\lambda n}{r}(1-\alpha)_{(r-1)}(\theta+k_n\alpha)\frac{(\theta+n+\alpha)_{\lambda n-r}}{(\theta+n)_{\lambda n}} \tag{26}$$

and

$$\tilde{U}_{\lambda n}(\alpha,\theta) := \left(k_n+\frac{\theta}{\alpha}\right)\left[\frac{(\theta+n+\alpha)_{\lambda n}}{(\theta+n)_{\lambda n}}-1\right], \tag{27}$$

respectively. Note that K_n is a sufficient statistic for both (26) and (27). We now consider the large n asymptotic behaviour of (26) and (27). For any $\alpha \in (0,1)$ and $\theta > -\alpha$, as $n \to +\infty$

$$\tilde{U}_{\lambda n,r}(\alpha,\theta) \sim \frac{\alpha(r-\alpha-1)_{r-1}}{r!}k_n\lambda^r(\lambda+1)^{r-\alpha} \sim \hat{U}^L_{\lambda n,r}(\alpha) \sim \hat{U}_{\lambda n,r}(\alpha) \tag{28}$$

and

$$\tilde{U}_{\lambda n}(\alpha,\theta) \sim k_n[(\lambda+1)^\alpha-1] \sim \hat{U}^L_{\lambda n}(\alpha) \sim \hat{U}_{\lambda n}(\alpha). \tag{29}$$

Formula (28) shows that the Bayesian nonparametric estimator $\tilde{U}_{\lambda n,r}(\alpha,\theta)$ is asymptotically equivalent, as $n \to +\infty$, to the regularly varying nonparametric empirical Bayes estimator $\hat{U}_{\lambda n,r}(\alpha)$. In particular it shows how the role of the parameter θ and of the Euler's smoothing in the estimation of $U_{\lambda n,r}$ become null as n becomes large. Formula (28) show that an analogous asymptotic equivalence holds between the estimators $\tilde{U}_{\lambda n}(\alpha,\theta)$ and $\hat{U}_{\lambda n}(\alpha)$.

Formulae (28) and (29) establish a link between two approaches for estimating $U_{\lambda n,r}$ and $U_{\lambda n}$. Both the approaches are Bayesian, but they rely on different prior specifications for $(p_i)_{i\geq 1}$: the Bayesian nonparametric approach relies on the two parameter Poisson-Dirichlet prior for the p_i's, whereas Robbins' approach relies on the sole assumption that p_i is independent and identically distributed from an unknown distribution function. According to (28) and (29), as $n \to +\infty$ these approaches are asymptotic equivalent if the prior specification in Robbins' approach is enriched by the assumption of regularly varying p_i's. The rationale of this asymptotic equivalence lies in the regularly varying behaviour of $P_{\alpha,\theta}$. More precisely, as noted in Gnedin et al. (2007), if $(p_i)_{i\geq 1}$ are the random probabilities of $P_{\alpha,0}$ then the counting measure $\overline{\nu}$ is such that

$$\overline{\nu}(x) \sim \frac{x^{-\alpha}}{\alpha B(1-\alpha,\alpha)} \tag{30}$$

as $x \downarrow 0$, where $B(\cdot,\cdot)$ is the Beta function. Equation (30) thus implies that $P_{\alpha,0}$ has regularly varying random probabilities. The random probabilities $(p_i)_{i\geq 1}$ of $P_{\alpha,\theta}$ are also regularly varying; this is because the law of $P_{\alpha,\theta}$ is absolutely continuous with respect to the law of $P_{\alpha,0}$. See Section 3 of Pitman and Yor (1997) for details. In general, we conjecture that the asymptotic equivalences (27) and (29) hold for any species sampling prior absolutely continuous with respect to the law of $P_{\alpha,0}$. These priors have been studied in Gnedin and Pitman (2006), and they are referred to as Gibbs-type priors.

4 Summary and concluding remarks

We presented a methodology for estimating the number $U_{\lambda n,r}$ of unseen species that would be observed with frequency $r \geq 1$ in an additional sample of size λn. This is an important quantity in biosciences, and especially in ecology and genetics, where there exists a practical interest in estimating not only the total number of unseen species, but also how many of these species are rare. We first introduced the estimator $\hat{U}^L_{\lambda n,r}$ of $U_{\lambda n,r}$, we showed that it estimates $U_{\lambda n,r}$ all of the way up $\lambda \propto \log_2(n)/2r\log_2(\log_2(n))$, and that this range is the best possible. The estimator $\hat{U}^L_{\lambda n,r}$ is obtained by combining Robbins' nonparametric empirical Bayes approach with Euler smoothing, which do not impose any distributional assumptions on the unknown species proportions $(p_i)_{i\geq 1}$. We then considered tuning $\hat{U}^L_{\lambda n,r}$ in order to improve its performance under heavy-tailed $(p_i)_{i\geq 1}$. In order to do that we imposed regular variation as a distribution-specific assumption of the tail behaviour of the p_i's. Interestingly, it turned out that the resulting regularly varying estimator is asymptotic equivalent, for large n, to the Bayesian nonparametric estimator of $U_{\lambda n,r}$ obtained under the assumption of a two parameter Poisson-Dirichlet prior for the unknown $(p_i)_{i\geq 1}$. Parallel to the methodology for estimating the number $U_{\lambda n,r}$, we presented a methodology for estimating the probability $V_{\lambda n,r}$ of discovering at the $(\lambda n + n + 1)$-th draw a species with frequency $r \geq 0$ in the enlarged sample of size $n + \lambda n$.

It remains an open problem to obtain confidence intervals for the proposed estimators of $U_{\lambda n,r}$ and $V_{\lambda n,r}$. In general, we are not aware of any work dealing with confidence intervals for Good-Toulmin type estimators, including the recent work by Orlitsky et al. (2017). For $\lambda < 1$ this is not a difficult task. Indeed from Theorem 1 we know that $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$ are unbiased estimators, and we have $\mathrm{Var}[U_{\lambda n,r} - \hat{U}_{\lambda n,r}] \lesssim n$ and $\mathrm{Var}[V_{\lambda n,r} - \hat{V}_{\lambda n,r}] \lesssim 1/n$. Then we may obtain confidence intervals via Bernstein inequalities with variance factors (8) and (9). See Boucheron et al. (2010) for details. Let $\Psi_{\lambda,r}$ and $\Phi_{\lambda,r}$ be as defined in Theorem 1. For any $s > 0$, Bernstein inequality leads to

$$\mathbb{P}\left[|U_{\lambda n,r} - \hat{U}_{\lambda n,r}| \geq \frac{2}{3}s\Psi_{\lambda,r} + \sqrt{2s\,\mathrm{Var}(U_{\lambda n,r} - \hat{U}_{\lambda n,r})}\right] \leq 2e^{-s} \tag{31}$$

and

$$\mathbb{P}\left[|V_{\lambda n,r} - \hat{V}_{\lambda n,r}| \geq \frac{2}{3}s\frac{\Phi_{\lambda,r}}{n} + \sqrt{2s\,\mathrm{Var}(V_{\lambda n,r} - \hat{V}_{\lambda n,r})}\right] \leq 2e^{-s}. \tag{32}$$

where $\mathrm{Var}[U_{\lambda n,r} - \hat{U}_{\lambda n,r}]$ and $\mathrm{Var}[V_{\lambda n,r} - \hat{V}_{\lambda n,r}]$ can be estimated by (8) and (9), respectively, upon noticing that $\mathbb{E}[M_{n,\bar{r}}] \leq n/r$ and $\mathbb{E}[M_{n,r}] \leq n/r$. See A.10 for details. Using a similar approach we may obtain a Bernstein-type confidence interval for the Good-Toulmin estimator $\hat{U}_{\lambda n}$. Note that (32) extends Proposition 5.4 and Proposition 5.5 of Ben-Hamou et al. (2017), where a Bernstein-type confidence interval for $V_{0,0}$ is obtained.

For $\lambda \geq 1$ the determination of confidence intervals for the estimators $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^I_{\lambda n,r}$ becomes a more challenging task, and we leave it as an open problem. A first issue in determining these confidence intervals arises by the fact that $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^I_{\lambda n,r}$ are no longer unbiased unbiased estimators of $U_{\lambda n,r}$ and $V_{\lambda n,r}$, respectively. Moreover, without any assumption on the species proportions p_i's, $\hat{U}^L_{\lambda n,r}$ and $\hat{V}^I_{\lambda n,r}$ are not even asymptotically unbiased. In addition, although the variance bounds (12) and (14) are useful in order to

tune the parameters k_L and k_I, they are not sharp enough to derive confidence intervals. This is because the binomial coefficients appearing in those bounds grow as $n \to +\infty$. The problem of the deriving confidence intervals is still open also for the regularly varying estimators proposed in Section 3. Here, due to the asymptotic nature of the regular variation assumption on the p_i's, one should look for asymptotic results to measure the uncertainty of the proposed estimators. Work on this is ongoing.

A Appendix

A.1 Details for the determination of (6) and (7)

The two estimators $\hat{U}_{\lambda n,r}$ and $\hat{V}_{\lambda n,r}$ are empirical Bayes estimators of $\mathbb{E}[U_{\lambda n,r}]$ and $\mathbb{E}[V_{\lambda n,r}]$ respectively in the sense of Robbins (1956), see also Maritz and Lwin (1989) for an account on the empirical Bayes approach. First of all we focus on the derivation of (6), note that the expectation of the statistic amounts to be

$$\mathbb{E}[U_{\lambda n,r}] = \mathbb{E}\left[\sum_{i=1}^{+\infty} \mathbb{1}_{\{N_{n,i}=0\}} \mathbb{1}_{\{N_{\lambda n,i}=r\}}\right] = \sum_{i=1}^{+\infty} e^{-(\lambda+1)np_i} \frac{(\lambda np_i)^r}{r!}. \tag{33}$$

Observe that $U_{\lambda n,r}$ depends on the observations only trough the counts $N_{n,i}$'s which are independent Poisson random variables, in particular, conditional on p_i they are Poisson with parameter np_i. We assume that $p_1, p_2, \ldots$ are distributed according to the empirical cumulative distribution function $G(p)$ of $p_{i_1}, \ldots, p_{i_k}$, corresponding to the k distinct species observed in the initial sample, namely $G(p) := k^{-1} \sum_{1 \le t \le k} p_{i_t}$ (see (Robbins, 1956, formula (12))). Under a square loss function, the nonparametric Bayes estimator for $\mathbb{E}[U_{\lambda n,r}]$ in (33) turns out to be

$$\sum_{x \ge 0} m_{n,x} \varphi_n(x), \quad \text{where } \varphi_n(x) := \frac{\displaystyle\int e^{-(\lambda+1)np} \frac{(\lambda np)^r}{r!} e^{-np} \frac{(np)^x}{x!} G(\mathrm{d}p)}{\displaystyle\int e^{-np} \frac{(np)^x}{x!} G(\mathrm{d}p)}, \tag{34}$$

as described in Robbins (1956). We focus on the determination of $\varphi_n(x)$

$$\begin{aligned}
\varphi_n(x) &= \frac{\int e^{-(\lambda+1)np} \frac{(\lambda np)^r}{r!} e^{-np} \frac{(np)^x}{x!} G(\mathrm{d}p)}{\int e^{-np} \frac{(np)^x}{x!} G(\mathrm{d}p)} \\
&= \frac{\frac{1}{x!} \int \sum_{j \ge 0} \frac{(-(\lambda+1)np)^j}{j!} \frac{(\lambda np)^r}{r!} e^{-np} (np)^x G(\mathrm{d}p)}{\int e^{-np} \frac{(np)^x}{x!} G(\mathrm{d}p)} \\
&= \frac{\frac{\lambda^r}{x! r!} \sum_{j \ge 0} \frac{(-(\lambda+1))^j}{j!} (j+r+x)! \int \frac{(np)^{j+r+x}}{(j+r+x)!} e^{-np} G(\mathrm{d}p)}{\int e^{-np} \frac{(np)^x}{x!} G(\mathrm{d}p)} \\
&= \frac{\frac{\lambda^r}{x! r!} \sum_{j \ge 0} \frac{(-(\lambda+1))^j}{j!} (j+r+x)! \mathbb{E}[M_{n,j+r+x}]}{\mathbb{E}[M_{n,x}]}.
\end{aligned}$$

Then the empirical Bayes estimator for $\mathbb{E}[U_{\lambda n,r}]$ may be obtained from (34) replacing the expectations appearing in the expression of $\varphi_n(x)$ with the observable quantities, more precisely

$$\begin{aligned}
\hat{U}_{\lambda n,r} &= \sum_{x\geq 0} m_{n,x} \frac{\frac{\lambda^r}{x!r!}\sum_{j\geq 0}\frac{(-(\lambda+1))^j}{j!}(j+r+x)!m_{n,j+r+x}}{m_{n,x}} \\
&= \sum_{x\geq 0}\frac{\lambda^r}{x!r!}\sum_{j\geq 0}\frac{(-(\lambda+1))^j}{j!}(j+r+x)!m_{n,j+r+x} \\
&= \frac{\lambda^r}{r!}\sum_{x\geq 0}\frac{1}{x!}\sum_{j\geq 0}\frac{(-(\lambda+1))^j}{j!}(j+r+x)!m_{n,j+r+x} \\
&= \frac{\lambda^r}{r!}\sum_{x\geq 0}\frac{1}{x!}\sum_{j\geq x}\frac{(-(\lambda+1))^{j-x}}{(j-x)!}(j+r)!m_{n,j+r} \\
&= \frac{\lambda^r}{r!}\sum_{j\geq 0}(-(\lambda+1))^j m_{n,j+r}(j+r)!\sum_{x=0}^{j}\frac{(-(\lambda+1))^{-x}}{x!(j-x)!} \\
&= \lambda^r\sum_{j\geq 0}(-(\lambda+1))^j m_{n,j+r}\frac{(j+r)!}{r!j!}\sum_{x=0}^{j}\binom{j}{x}(-(\lambda+1))^{-x} \\
&= \lambda^r\sum_{j\geq 0}(-(\lambda+1))^j m_{n,j+r}\frac{(j+r)!}{r!j!}\left(\frac{\lambda}{\lambda+1}\right)^j = \lambda^r\sum_{j\geq 0}(-\lambda)^j m_{n,j+r}\binom{j+r}{r}
\end{aligned}$$

and (6) has been proven.
The other estimator in (7) follows in a similar manner, in such a situation one has to observe that

$$\mathbb{E}[V_{\lambda n,r}] = \mathbb{E}\left[\sum_{i=1}^{+\infty} p_i \mathbb{1}_{\{N_{\lambda n+n,i}=r\}}\right] = \sum_{i=1}^{+\infty} e^{-(\lambda+1)np_i}\frac{(\lambda n+n)^r p_i^{r+1}}{r!}. \tag{35}$$

We want to show that $\hat{V}_{\lambda n,r}$ is the empirical Bayes estimator of (35). The Bayes estimator for $\mathbb{E}[V_{\lambda n,r}]$, under the square loss function, amounts to be

$$\sum_{x\geq 0} m_{n,x}\varphi_n(x), \quad \text{where } \varphi_n(x) := \frac{\displaystyle\int e^{-(\lambda+1)np}\frac{(\lambda+1)^r n^r p^{r+1}}{r!}e^{-np}\frac{(np)^x}{x!}G(\mathrm{d}p)}{\displaystyle\int e^{-np}\frac{(np)^x}{x!}G(\mathrm{d}p)},$$

as before it is easy to see that

$$\varphi_n(x) = \frac{1}{n}\sum_{j\geq 0}\frac{(-1)^j(\lambda+1)^{j+r}}{j!r!x!}(j+r+x+1)!\frac{\mathbb{E}[M_{n,j+r+x+1}]}{\mathbb{E}[M_{n,x}]}.$$

The empirical Bayes estimator may be derived replacing $\mathbb{E}[M_{n,j+r+x+1}]$ and $\mathbb{E}[M_{n,x}]$ with the observable quantities $m_{n,j+r+x+1}$ and $m_{n,x}$, respectively, thus obtaining

$$\begin{aligned}\hat{V}_{\lambda n,r} &= \sum_{x\geq 0} m_{n,x}\frac{1}{n}\sum_{j\geq 0}\frac{(-1)^j(\lambda+1)^{j+r}}{j!r!x!}(j+r+x+1)!\frac{m_{n,j+r+x+1}}{m_{n,x}}\\ &= \sum_{x\geq 0}\frac{1}{n}\sum_{j\geq 0}\frac{(-1)^j(\lambda+1)^{j+r}}{j!r!x!}(j+r+x+1)!m_{n,j+r+x+1}.\end{aligned}$$

Simple calculations, as before, show that the previous estimator coincides with the expression in (7).

A.2 Proof of Theorem 1

First of all we focus on the proof of (8). The independence of the random variables implies

$$\begin{aligned}&\mathrm{Var}(U_{\lambda n,r}-\hat{U}_{\lambda n,r})\\ &\quad=\sum_{i=1}^{\infty}\mathrm{Var}\left(\sum_{j=0}^{\infty}(-1)^j\binom{r+j}{r}\lambda^{j+r}\mathbb{1}_{\{N_{n,i}=r+j\}}-\mathbb{1}_{\{N_{n,i}=0\}}\mathbb{1}_{\{N_{\lambda n,i}=r\}}\right).\end{aligned}$$

Note that, for any $i\geq 1$, the random variable

$$\sum_{j=0}^{\infty}(-1)^j\binom{r+j}{r}\lambda^{j+r}\mathbb{1}_{\{N_{n,i}=r+j\}}-\mathbb{1}_{\{N_{n,i}=0\}}\mathbb{1}_{\{N_{\lambda n,i}=r\}}$$

has mean zero, indeed

$$\begin{aligned}\mathbb{E}\left[\lambda^r\sum_{j=0}^{\infty}(-1)^j\binom{r+j}{r}\lambda^j\mathbb{1}_{\{N_{n,i}=r+j\}}\right] &= \lambda^r\sum_{j\geq 0}(-1)^j\binom{r+j}{r}\lambda^j\mathbb{P}(N_{n,i}=r+j)\\ &= \lambda^r\sum_{j\geq 0}(-1)^j\binom{r+j}{r}\lambda^j e^{-np_i}\frac{(np_i)^{r+j}}{(r+j)!}\\ &= \frac{\lambda^r}{r!}(np_i)^r e^{-np_i}\sum_{j\geq 0}\frac{((-\lambda)np_i)^j}{j!}\\ &= \frac{(np_i)^r\lambda^r}{r!}e^{-(\lambda n+n)p_i},\end{aligned}$$

on the other side

$$\mathbb{E}\left[\mathbb{1}_{\{N_{n,i}=0\}}\mathbb{1}_{\{N_{\lambda n,i}=r\}}\right]=\mathbb{P}(N_{n,i}=0)\mathbb{P}(N_{\lambda n,i}=r)=\frac{(np_i)^r\lambda^r}{r!}e^{-(\lambda n+n)p_i}.$$

As a byproduct, the previous calculations show that $\hat{U}_{\lambda n,r}$ is an unbiased estimator of $U_{\lambda n,r}$. Then we can write

$$\mathrm{Var}(U_{\lambda n,r}-\hat{U}_{\lambda n,r})=\sum_{i=1}^{\infty}\mathbb{E}\left(\sum_{j=0}^{\infty}a_j\mathbb{1}_{\{N_{n,i}=r+j\}}-\mathbb{1}_{\{N_{n,i}=0\}}\mathbb{1}_{\{N_{\lambda n,i}=r\}}\right)^2, \tag{36}$$

where

$$a_j := (-1)^j \binom{r+j}{r} \lambda^{j+r}.$$

All the indicators in (36) are indicators of incompatible events, therefore we may write

$$\begin{aligned}
\mathrm{Var}(U_{\lambda n,r} - \hat{U}_{\lambda n,r}) &= \sum_{i=1}^{\infty} \mathbb{E}\left(\sum_{j=0}^{\infty} a_j^2 \mathbb{1}_{\{N_{n,i}=r+j\}} + \mathbb{1}_{\{N_{n,i}=0\}} \mathbb{1}_{\{N_{\lambda n,i}=r\}} \right) \\
&= \sum_{j\geq 0} a_j^2 \sum_{i=1}^{\infty} \mathbb{E}\left[\mathbb{1}_{\{N_{n,i}=r+j\}} \right] + \left(1 - \frac{1}{\lambda+1}\right)^r \mathbb{E}[M_{\lambda n+n,r}] \\
&= \sum_{j\geq 0} a_j^2 \mathbb{E}[M_{n,r+j}] + \left(\frac{\lambda}{\lambda+1}\right)^r \mathbb{E}[M_{\lambda n+n,r}]
\end{aligned}$$

Since $\lambda < 1$, one can easily verify that $a_{j+1} \leq a_j$ iff $j \geq j_1^*$, where $j_1^* = \lfloor r\lambda/(1-\lambda) - 1 \rfloor \vee 0$. Therefore we can estimate the a_j's with their maximum value to obtain

$$\mathrm{Var}(U_{\lambda n,r} - \hat{U}_{\lambda n,r}) \leq a_{j_1^*}^2 \mathbb{E}[M_{n,\bar{r}}] + \left(\frac{\lambda}{\lambda+1}\right)^r \mathbb{E}[M_{\lambda n+n,r}]$$

and this gives (8).
Now we can prove (9) with similar arguments. Straightforward calculations show that the random variables $\{(N_{\lambda n+n,i}, N_{n,i})\}_{i\geq 1}$ are independent couples, hence we may write

$$\begin{aligned}
&\mathrm{Var}(V_{\lambda n,r} - \hat{V}_{\lambda n,r}) \\
&\quad = \mathrm{Var}\left(\sum_{i=1}^{\infty} p_i \mathbb{1}_{\{N_{\lambda n+n,i}=r\}} - \frac{1}{n}(r+1)(\lambda+1)^r \sum_{j\geq 0} (-1)^j \binom{r+j+1}{r+1} \lambda^j \sum_{i=1}^{\infty} \mathbb{1}_{\{N_{n,i}=r+j+1\}} \right) \\
&\quad = \mathrm{Var}\left(\sum_{i=1}^{\infty} \left(p_i \mathbb{1}_{\{N_{\lambda n+n,i}=r\}} - \frac{1}{n}(r+1)(\lambda+1)^r \sum_{j\geq 0} (-1)^j \binom{r+j+1}{r+1} \lambda^j \mathbb{1}_{\{N_{n,i}=r+j+1\}} \right) \right) \\
&\quad = \sum_{i=1}^{\infty} \mathrm{Var}\left(p_i \mathbb{1}_{\{N_{\lambda n+n,i}=r\}} - \frac{1}{n}(r+1)(\lambda+1)^r \sum_{j\geq 0} (-1)^j \binom{r+j+1}{r+1} \lambda^j \mathbb{1}_{\{N_{n,i}=r+j+1\}} \right).
\end{aligned}$$

As before, for the sake of simplifying notations, we put

$$a_j := \frac{1}{n}(r+1)(\lambda+1)^r (-1)^j \binom{r+j+1}{r+1} \lambda^j,$$

then, the variance may be rewritten as

$$\mathrm{Var}(V_{\lambda n,r} - \hat{V}_{\lambda n,r}) = \sum_{i=1}^{\infty} \mathrm{Var}\left(p_i \mathbb{1}_{\{N_{\lambda n+n,i}=r\}} - \sum_{j\geq 0} a_j \mathbb{1}_{\{N_{n,i}=r+j+1\}} \right).$$

Noticing that, for any $i \geq 1$, the random variables inside the variance have zero mean, we obtain:

$$\begin{aligned}
\mathrm{Var}(V_{\lambda n,r} - \hat{V}_{\lambda n,r}) &= \sum_{i=1}^{\infty} \mathbb{E}\left[p_i \mathbb{1}_{\{N_{\lambda n+n,i}=r\}} - \sum_{j\geq 0} a_j \mathbb{1}_{\{N_{n,i}=r+j+1\}} \right]^2 \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left[p_i^2 \mathbb{1}_{\{N_{\lambda n+n,i}=r\}} + \sum_{j\geq 0} a_j^2 \mathbb{1}_{\{N_{n,i}=r+j+1\}} \right]
\end{aligned}$$

where we have used the incompatibility of the events. Simple calculations show that, for any $\lambda \in (0,1)$, the function a_j^2 has a unique maximum for $j = j_2^*$, where $j_2^* = \lfloor ((r+2)\lambda - 1)/(1-\lambda) \rfloor \vee 0$, then

$$\begin{aligned}
\mathrm{Var}(V_{\lambda n,r} - \hat{V}_{\lambda n,r}) &= \sum_{i=1}^{\infty} \mathbb{E}\left[p_i^2 \mathbb{1}_{\{N_{\lambda n+n,i}=r\}} + \sum_{j\geq 0} a_j^2 \mathbb{1}_{\{N_{n,i}=r+j+1\}} \right] \\
&= \sum_{i=1}^{\infty} p_i^2 e^{-p_i(\lambda n+n)} \frac{(p_i(n+\lambda n))^r}{r!} + \sum_{j\geq 0} a_j^2 \mathbb{E}[M_{n,r+j+1}] \\
&= \frac{(r+1)(r+2)}{(\lambda n+n)^2} \mathbb{E}[M_{\lambda n+n,r+2}] + \sum_{j\geq 0} a_j^2 \mathbb{E}[M_{n,r+j+1}] \\
&\leq \frac{(r+1)(r+2)}{(\lambda n+n)^2} \mathbb{E}[M_{\lambda n+n,r+2}] + a_{j_2^*}^2 \mathbb{E}[M_{n,\overline{r+1}}]
\end{aligned}$$

and (9) now follows. □

A.3 Details for the determination of (10) and (11)

First we focus on the derivation of $\hat{U}^L_{\lambda n,r}$, defined in (10). We employ the same technique described by Efron and Thisted (1976) to define the smoothed version of the estimator $\hat{U}_{\lambda n,r}$. Recall that

$$\hat{U}_{\lambda n,r} = \sum_{j\geq 0} (-1)^j \binom{r+j}{r} \lambda^{j+r} m_{n,r+j},$$

the Eulero transformation suggested in Efron and Thisted (1976) consists in a simple change of variable $\lambda = u/(2-u)$, which yields

$$\hat{U}_{\lambda n,r} = \sum_{j\geq 0} (-1)^j \eta_{j+r} \frac{u^{r+j}}{2^{r+j}(1-u/2)^{r+j}}$$

where

$$\eta_{j+r} = \binom{r+j}{r} m_{n,r+j}.$$

Since $u = 2\lambda/(\lambda+1) \leq 2$ for any $\lambda > 0$, we can use the Taylor series expansion to get

$$\begin{aligned}
\hat{U}_{\lambda n,r} &= \sum_{j\geq 0}(-1)^j \eta_{j+r}\frac{u^{r+j}}{2^{r+j}(1-u/2)^{r+j}} = \sum_{j\geq 0}\sum_{y\geq 0}\eta_{j+r}(-1)^{y+j}\binom{-(r+j)}{y}\left(\frac{u}{2}\right)^{r+j+y} \\
&= \sum_{j\geq 0}\sum_{z\geq j}\eta_{j+r}(-1)^{z}\binom{-(r+j)}{z-j}\left(\frac{u}{2}\right)^{r+z} = \sum_{j\geq 0}\sum_{z\geq j}\eta_{j+r}(-1)^{j}\binom{r+z-1}{r+j-1}\left(\frac{u}{2}\right)^{r+z} \\
&= \sum_{z\geq 0}\sum_{j=0}^{z}(-1)^j\eta_{r+j}\binom{r+z-1}{r+j-1}\frac{u^{z+r}}{2^{z+r}}.
\end{aligned}$$

Now define

$$\xi_z := \sum_{j=0}^{z}\frac{(-1)^j}{2^{z+r}}\eta_{r+j}\binom{r+z-1}{r+j-1}.$$

Let $k \in \mathbb{N}$, consider the partial sums

$$\Delta_k(u) := \sum_{z=0}^{k}\xi_z u^{z+r}, \quad \Delta_k(\lambda) := \sum_{z=0}^{k}(-1)^z\binom{r+z}{r}\lambda^{z+r}m_{n,r+z}$$

then the limit of both $\Delta_k(u)$ and $\Delta_k(\lambda)$ as $k \to \infty$ is the same and coincides with $\hat{U}_{\lambda n,r}$, whenever the limit exists. For $\eta_x > 0$, the partial sums $\Delta_k(u)$ usually converge more quickly to the common limit than the sums $\Delta_k(\lambda)$ (see Efron and Thisted (1976)). As a consequence, it is more convenient to use $\Delta_k(u)$ in order to define an estimator for $U_{\lambda n,r}$ whenever $\lambda \geq 1$. With this in mind, simple calculations show that

$$\begin{aligned}
\sum_{z=0}^{k}\xi_z u^{z+r} &= \sum_{j=0}^{k}(-1)^j\eta_{j+r}\sum_{z=j}^{k}\binom{r+z-1}{r+j-1}\left(\frac{u}{2}\right)^{z+r} \\
&= \sum_{j=r}^{k+r}(-1)^{j+r}\eta_j\lambda^j\sum_{z=j}^{k+r}\binom{z-1}{j-1}\frac{\lambda^{z-j}}{(\lambda+1)^z}
\end{aligned}$$

and we observe that

$$\sum_{z=j}^{k+r}\binom{z-1}{j-1}\frac{\lambda^{z-j}}{(\lambda+1)^z} = \mathbb{P}(L \geq j)$$

being $L \sim \text{Binom}(k+r, 1/(1+\lambda))$. Finally we can write $\Delta_k(u)$ as a function of this binomial random variable

$$\begin{aligned}
\sum_{z=0}^{k}\xi_z u^{z+r} &= \sum_{j=r}^{k+r}(-1)^{j+r}\eta_j\lambda^j\mathbb{P}(L \geq j) = \sum_{j=0}^{k}(-1)^j\eta_{j+r}\lambda^{j+r}\mathbb{P}(L \geq j+r) \\
&= \lambda^r\sum_{j=0}^{k}(-\lambda)^j\binom{r+j}{r}\mathbb{P}(L \geq j+r)m_{n,r+j}
\end{aligned}$$

$$= \lambda^r \sum_{j\geq 0} (-\lambda)^j \binom{r+j}{r} \mathbb{P}(L \geq j+r) m_{n,r+j},$$

having observed that $\mathbb{P}(L > k+r) = 0$. Hence

$$U^L_{\lambda n,r} := \lambda^r \sum_{j\geq 0} (-\lambda)^j \binom{r+j}{r} \mathbb{P}(L \geq j+r) m_{n,r+j}$$

where $L \sim \text{Binom}(k_L + r, 1/(1+\lambda))$ may be used as an estimator of $U_{\lambda n,r}$ when $\lambda \geq 1$ and (10) is justified.
A similar argument may be applied to determine $V^I_{\lambda n,r}$. □

A.4 Proof of Theorem 2: bounds (12) and (13)

We now focus on part i) of Theorem 2, providing the two estimates for the bias and the variance in the statement of the Theorem. Finally in Section A.4.3, thanks to these results, we are able to find a bound for the MSE, which will be useful in the rest of the paper.

A.4.1 Bound for the bias

First of all we provide an equality for the bias which holds true for a general random variable L, not necessarily a binomial one. To this end consider the bias of the estimator

$$\begin{aligned}
&\mathbb{E}[\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}] \\
&\quad = \mathbb{E}\left[\lambda^r \sum_{j\geq 0} (-1)^j \binom{r+j}{r} m_{n,r+j} \mathbb{P}(L \geq j+r) \lambda^j - \sum_{i\geq 1} \mathbb{1}_{\{N_{n,i}=0\}} \mathbb{1}_{\{N_{\lambda n,i}=r\}}\right]
\end{aligned}$$

noticing that $\mathbb{P}(L \geq j+r) = 1 - \mathbb{P}(L \leq j+r-1)$ and using the unbiasedness of the non–smoothed estimator $\hat{U}_{\lambda n,r}$ we have

$$\begin{aligned}
\mathbb{E}[\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}] &= -\mathbb{E}\left[\lambda^r \sum_{j\geq 0} (-1)^j \binom{r+j}{r} m_{n,r+j} \mathbb{P}(L \leq j+r-1) \lambda^j\right] \\
&= -\lambda^r \sum_{j\geq 0} (-1)^j \binom{r+j}{r} \mathbb{P}(L \leq j+r-1) \lambda^j \sum_{i\geq 1} \mathbb{P}(N_{n,i} = r+j) \\
&= -\lambda^r \sum_{j\geq 0} (-1)^j \binom{r+j}{r} \mathbb{P}(L \leq j+r-1) \lambda^j \sum_{i\geq 1} e^{-p_i n} \frac{(p_i n)^{r+j}}{(r+j)!} \\
&= -\lambda^r \sum_{i\geq 1} e^{-p_i n} \frac{(p_i n)^r}{r!} \sum_{j\geq 0} \frac{\mathbb{P}(L \leq j+r-1)}{j!} (-\lambda p_i n)^j. \qquad (37)
\end{aligned}$$

We focus on the evaluation of the sum over j in (37). For the sake of simplicity we put $y := \lambda p_i n$, hence we are dealing with the following sum

$$\sum_{j\geq 0} \mathbb{P}(L \leq j+r-1) \frac{(-y)^j}{j!} \quad = \quad \sum_{j=0}^{\infty} \sum_{i=0}^{j+r-1} \mathbb{P}(L = i) \frac{(-y)^j}{j!}$$

$$
\begin{aligned}
&= \sum_{i=0}^{r-1}\sum_{j\geq 0}\mathbb{P}(L=i)\frac{(-y)^j}{j!} + \sum_{i=r}^{\infty}\sum_{j=i-r-1}^{\infty}\mathbb{P}(L=i)\frac{(-y)^j}{j!} \\
&= \mathbb{P}(L\leq r-1)e^{-y} + \sum_{i\geq 0}\mathbb{P}(L=i+r)\sum_{j=i+1}^{\infty}\frac{(-y)^j}{j!} \\
&\overset{(a)}{=} \mathbb{P}(L\leq r-1)e^{-y} + \sum_{i\geq 0}\mathbb{P}(L=i+r)\frac{e^{-y}}{i!}\int_0^{-y}\tau^i e^{-\tau}\mathrm{d}\tau \\
&= \mathbb{P}(L\leq r-1)e^{-y} - e^{-y}\int_0^{y} e^s \sum_{i\geq 0}\mathbb{P}(L=i+r)\frac{(-s)^i}{i!}\mathrm{d}s.
\end{aligned}
$$

where (a) follows by the definition of the incomplete gamma function:

$$
\sum_{i=j+1}^{\infty}\frac{z^i}{i!} = \frac{e^z}{j!}\int_0^z \tau^j e^{-\tau}\mathrm{d}\tau.
$$

Summing up, the bias now equals

$$
\begin{aligned}
\mathbb{E}[\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}] = -\lambda^r \sum_{i\geq 1} e^{-p_i n}\frac{(p_i n)^r}{r!} \\
\times \left[\mathbb{P}(L\leq r-1)e^{-y} - e^{-y}\int_0^y e^s \sum_{i\geq 0}\mathbb{P}(L=i+r)\frac{(-s)^i}{i!}\mathrm{d}s\right],
\end{aligned} \tag{38}
$$

where $y = \lambda p_i n$. Recalling that $L \sim \mathrm{Binom}(k_L + r, q)$, being $q = 1/(\lambda+1)$, we concentrate on the integral in (38):

$$
\begin{aligned}
&\int_0^y e^s \sum_{i=0}^{\infty}\mathbb{P}(L=i+r)\frac{(-s)^i}{i!}\mathrm{d}s \\
&= q^r(1-q)^{k_L}\int_0^y e^s \sum_{i=0}^{k_L+r}\frac{1}{i!}\binom{k_L+r}{i+r}(-1)^i\left(\frac{sq}{1-q}\right)^i \mathrm{d}s \\
&= q^r(1-q)^{k_L}\int_0^y e^s \sum_{i=0}^{k_L}\frac{1}{i!}\binom{k_L+r}{k_L-i}(-1)^i\left(\frac{sq}{1-q}\right)^i \mathrm{d}s \\
&= q^r(1-q)^{k_L}\int_0^y e^s L^r_{k_L}\left(\frac{sq}{(1-q)}\right)\mathrm{d}s
\end{aligned}
$$

where $L^\alpha_n(z)$ denotes the Laguerre polynomial of degree n defined by

$$
L^\alpha_n(z) := \sum_{m=0}^{n}(-1)^m\binom{n+\alpha}{n-m}\frac{z^m}{m!}.
$$

Remembering that (see (Olver et al., 2010, pg. 450))

$$
|L^\alpha_n(z)| \leq \binom{n+\alpha}{n}e^{z/2}, \tag{39}
$$

the previous integral may be estimated as follows

$$\begin{aligned}
\Big|\int_0^y e^s \sum_{i=0}^{\infty} \mathbb{P}(L=i+r)\frac{(-s)^i}{i!}\mathrm{d}s\Big| &\leq q^r(1-q)^{k_L}\int_0^y e^s\Big|L^r_{k_L}\left(\frac{sq}{(1-q)}\right)\Big|\mathrm{d}s \\
&\overset{(39)}{\leq} q^r(1-q)^{k_L}\binom{r+k_L}{k_L}\int_0^y \exp\left\{s\left(1+\frac{q}{2(1-q)}\right)\right\}\mathrm{d}s \\
&= q^r(1-q)^{k_L}\binom{k_L+r}{k_L}\frac{2(1-q)}{(2-q)} \\
&\qquad \times\left(\exp\left\{y\left(1+\frac{q}{2(1-q)}\right)\right\}-1\right).
\end{aligned}$$

If we now use the resulting estimate for the integral to bound the bias in (38), recalling that $y=\lambda p_i n$, then we obtain

$$\begin{aligned}
&|\mathbb{E}[\hat{U}^L_{\lambda n,r}-U_{\lambda n,r}]| \\
&\quad\leq \lambda^r\sum_{i\geq 1} e^{-p_i n}\frac{(p_i n)^r}{r!}\Big[e^{-\lambda p_i n}\mathbb{P}(L\leq r-1) \\
&\qquad\qquad + q^r(1-q)^{k_L}\binom{k_L+r}{k_L}\frac{2(1-q)}{(2-q)}\left(\exp\left\{\frac{q\lambda p_i n}{2(1-q)}\right\}-e^{-\lambda p_i n}\right)\Big] \\
&\quad= \left(\frac{\lambda}{\lambda+1}\right)^r\mathbb{P}(L\leq r-1)\mathbb{E}[M_{\lambda n+n,r}] \\
&\qquad\qquad + 2(q\lambda)^r\frac{(1-q)^{k_L+1}}{(2-q)}\binom{k_L+r}{k_L}\sum_{i\geq 1}e^{-p_i n}\frac{(p_i n)^r}{r!}\left(\exp\left\{\frac{q\lambda p_i n}{2(1-q)}\right\}-e^{-\lambda p_i n}\right).
\end{aligned}$$

Remembering that $q=1/(1+\lambda)$, we have proved the following estimate for the bias of the estimator

$$\begin{aligned}
&|\mathbb{E}[\hat{U}^L_{\lambda n,r}-U_{\lambda n,r}]| \\
&\quad\leq \left(\frac{\lambda}{\lambda+1}\right)^r\mathbb{P}(L\leq r-1)\mathbb{E}[M_{\lambda n+n,r}] \\
&\qquad + 2\left(\frac{\lambda}{\lambda+1}\right)^{r+k_L+1}\frac{1+\lambda}{1+2\lambda}\binom{k_L+r}{k_L}\sum_{i\geq 1}e^{-p_i n/2}\frac{(p_i n)^r}{r!}\left(1-\exp\left\{-p_i n\left(\lambda+1/2\right)\right\}\right) \\
&\quad\leq \left(\frac{\lambda}{\lambda+1}\right)^r\mathbb{P}(L\leq r-1)\mathbb{E}[M_{\lambda n+n,r}] \\
&\qquad + 2^{r+1}\left(\frac{\lambda}{\lambda+1}\right)^{r+k_L+1}\frac{\lambda+1}{1+2\lambda}\binom{k_L+r}{r}\mathbb{E}[M_{n/2,r}], \qquad (40)
\end{aligned}$$

where we have used the inequality $(1-\exp\{-p_i n(\lambda+1/2)\})\leq 1$, which is legitimated by the fact that $n\to+\infty$.

A.4.2 Bound for the variance

First of all we prove a variance bound which holds true for a general random variable L, then we focus on the binomial case we are dealing with. Note that the variance may be expressed as follows

$$\begin{aligned}
&\mathrm{Var}(\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}) \\
&\quad = \sum_{i=1}^{\infty} \mathrm{Var}\left(\lambda^r \sum_{j=0}^{\infty} (-1)^j \mathbb{P}(L \geq j+r) \binom{r+j}{r} \lambda^j \mathbb{1}_{\{N_{n,i}=r+j\}} - \mathbb{1}_{\{N_{n,i}=0\}} \mathbb{1}_{\{N_{\lambda n,i}=r\}}\right) \\
&\quad \leq \sum_{i=1}^{\infty} \mathbb{E}\left[\lambda^r \sum_{j=0}^{\infty} (-1)^j \mathbb{P}(L \geq j+r) \binom{r+j}{r} \lambda^j \mathbb{1}_{\{N_{n,i}=r+j\}} - \mathbb{1}_{\{N_{n,i}=0\}} \mathbb{1}_{\{N_{\lambda n,i}=r\}}\right]^2 .
\end{aligned}$$

The events $\{N_{n,i} = r+j\}$ are incompatible for different values of j, hence we can write

$$\mathrm{Var}(\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}) \leq \sum_{i=1}^{\infty} \mathbb{E}\left[\sum_{j=0}^{\infty} a_j^2 \mathbb{1}_{\{N_{n,i}=r+j\}}\right] + \left(\frac{\lambda}{\lambda+1}\right)^r \mathbb{E}[M_{\lambda n+n,r}] \tag{41}$$

where we have put

$$a_j := \binom{r+j}{r} \lambda^{r+j} \mathbb{P}(L \geq j+r).$$

From (41) we obtain

$$\mathrm{Var}(\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}) \leq \max_{j\geq 0} a_j^2 \mathbb{E}[M_{n,\bar{r}}] + \left(\frac{\lambda}{\lambda+1}\right)^r \mathbb{E}[M_{\lambda n+n,r}]$$

and besides

$$\begin{aligned}
\max_{j\geq 0} |a_j| &= \max_{j\geq 0} \mathbb{P}(L \geq j+r) \binom{r+j}{r} \lambda^{r+j} = \max_{j\geq 0} \sum_{i=j}^{\infty} \mathbb{P}(L = i+r) \binom{r+j}{r} \lambda^{r+j} \\
&\overset{(a)}{\leq} \max_{j\geq 0} \sum_{i=j}^{\infty} \mathbb{P}(L = i+r) \binom{r+i}{r} \lambda^{r+i} \leq \sum_{i=0}^{\infty} \mathbb{P}(L = i+r) \binom{r+i}{r} \lambda^{r+i}
\end{aligned}$$

(a) is due to the fact that $\binom{r+j}{r}\lambda^{r+j}$ is increasing as a function of j, since $\lambda \geq 1$. Summing up, we have obtained the following variance bound for the difference between the estimator and the statistic

$$\begin{aligned}
&\mathrm{Var}(\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}) \\
&\quad \leq \left(\sum_{i=0}^{\infty} \mathbb{P}(L = i+r) \binom{r+i}{r} \lambda^{r+i}\right)^2 \mathbb{E}[M_{n,\bar{r}}] + \left(\frac{\lambda}{\lambda+1}\right)^r \mathbb{E}[M_{\lambda n+n,r}],
\end{aligned} \tag{42}$$

which holds true for a general random variable L. Now we remember that $L \sim \mathrm{Binom}(k_L + r, q)$, being $q = (\lambda+1)^{-1}$, hence the infinite sum in the variance bound (42) boils down to

$$\sum_{i=0}^{\infty} \mathbb{P}(L = i+r) \binom{r+i}{r} \lambda^{r+i} = \sum_{i=0}^{\infty} \binom{k_L+r}{i+r} q^{i+r} (1-q)^{k_L-i} \binom{r+i}{r} \lambda^{i+r}$$

$$
\begin{aligned}
&= (q\lambda)^r \binom{k_L + r}{k_L} \sum_{i=0}^{k_L} \binom{k_L}{i} (q\lambda)^i (1-q)^{k_L - i} \\
&= (q\lambda)^r \binom{k_L + r}{k_L} (1 + q(\lambda - 1))^{k_L},
\end{aligned}
$$

recalling that $q = 1/(1+\lambda)$ we obtain

$$
\sum_{i=0}^{\infty} \mathbb{P}(L = i + r) \binom{r+i}{r} \lambda^{r+i} = \left(\frac{\lambda}{\lambda+1}\right)^{k_L + r} 2^{k_L} \binom{k_L + r}{r}.
$$

Substituting the previous expression in (42), the bound on the variance becomes

$$
\begin{aligned}
&\mathrm{Var}(\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}) \leq \\
&\qquad 4^{k_L} \left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \binom{k_L + r}{k_L}^2 \mathbb{E}[M_{n,\bar{r}}] + \left(\frac{\lambda}{\lambda+1}\right)^r \mathbb{E}[M_{\lambda n + n, r}].
\end{aligned} \tag{43}
$$

A.4.3 Bound for the MSE

Remind that the mean square error boils down to the sum of the squared bias and the variance, in other terms

$$
\begin{aligned}
\mathrm{MSE}[U_{\lambda n,r} - \hat{U}^L_{\lambda n,r}] &= \mathbb{E}[(U_{\lambda n,r} - \hat{U}^L_{\lambda n,r})^2] \\
&= \mathrm{Bias}[\hat{U}^L_{\lambda n,r}]^2 + \mathrm{Var}[U_{\lambda n,r} - \hat{U}^L_{\lambda n,r}],
\end{aligned} \tag{44}
$$

where $\mathrm{Bias}[\hat{U}^L_{\lambda n,r}] = \mathbb{E}[\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}]$. In Sections A.4.1–A.4.2, we have proved (40) and (43):

$$
\begin{aligned}
|\,\mathrm{Bias}(\hat{U}^L_{\lambda n,r})| \leq &\left(\frac{\lambda}{\lambda+1}\right)^r \mathbb{P}(L \leq r-1) \mathbb{E}[M_{\lambda n + n, r}] \\
&+ 2^{r+1} \left(\frac{\lambda}{\lambda+1}\right)^{r+k_L+1} \frac{\lambda+1}{1+2\lambda} \binom{k_L + r}{r} \mathbb{E}[M_{n/2,r}]
\end{aligned}
$$

$$
\mathrm{Var}(\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}) \leq 4^{k_L} \left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \binom{k_L + r}{k_L}^2 \mathbb{E}[M_{n,\bar{r}}] + \left(\frac{\lambda}{\lambda+1}\right)^r \mathbb{E}[M_{\lambda n + n, r}].
$$

Estimating the expected values in the previous expressions with the maximum possible value that they can achieve, we obtain respectively that

$$
\begin{aligned}
|\,\mathrm{Bias}(\hat{U}^L_{\lambda n,r})| \leq &\left(\frac{\lambda}{\lambda+1}\right)^r \mathbb{P}(L \leq r-1) \frac{(\lambda+1)n}{r} \\
&+ 2^{r} \left(\frac{\lambda}{\lambda+1}\right)^{r+k_L+1} \frac{\lambda+1}{1+2\lambda} \binom{k_L + r}{r} \frac{n}{r}
\end{aligned}
$$

and

$$
\mathrm{Var}(\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}) \leq 2^{2k_L} \left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \binom{k_L + r}{k_L}^2 \frac{n}{r} + \left(\frac{\lambda}{\lambda+1}\right)^r \frac{(\lambda+1)n}{r},
$$

Replacing the previous expressions in (44) we obtain

$$\begin{aligned}\mathrm{MSE}[U_{\lambda n,r} - \hat{U}^L_{\lambda n,r}] &\leq 2^{2k_L}\left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)}\binom{k_L+r}{k_L}^2\frac{n}{r} + \left(\frac{\lambda}{\lambda+1}\right)^r\frac{(\lambda+1)n}{r} \\ &+ \frac{(\lambda n)^2}{r^2}\left[\mathbb{P}(L\leq r-1)\left(\frac{\lambda}{\lambda+1}\right)^{r-1} + 2^r\left(\frac{\lambda}{\lambda+1}\right)^{k_L+r}\frac{1}{1+2\lambda}\binom{k_L+r}{r}\right]^2\end{aligned} \tag{45}$$

□

A.5 Proof of Theorem 2: bound (15) and (14)

One can proceed exactly as we did in Section 13, for this reason we omit the details of calculations. Along the same lines of Section A.4.1, one can derive the following estimate for the bias

$$\begin{aligned}|\operatorname{Bias}(\hat{V}^I_{\lambda n,r})| &\leq \frac{r+1}{\lambda n+n}\mathbb{P}(I\leq r)\mathbb{E}[M_{\lambda n+n,r+1}] \\ &+ \frac{r+1}{n}\left(\frac{\lambda}{\lambda+1}\right)^{k_I+1}\frac{2^{r+2}(\lambda+1)}{2\lambda+1}\binom{k_I+r+1}{k_I}\mathbb{E}[M_{n/2,r+1}].\end{aligned} \tag{46}$$

Analogously, the bound for the variance of the difference $\hat{V}^I_{\lambda n,r} - V_{\lambda n,r}$, may be obtained through the techniques of Section A.4.2:

$$\begin{aligned}\mathrm{Var}(\hat{V}^I_{\lambda n,r} - V_{\lambda n,r}) &\leq \left(\frac{\lambda}{\lambda+1}\right)^{2k_I}\frac{(r+1)2^{2k_I}}{(\lambda+1)^2n^2}\binom{k_I+r+1}{k_I}^2\mathbb{E}[M_{n,\overline{r+1}}] \\ &+ \frac{(r+1)(r+2)}{(\lambda+1)^2n^2}\mathbb{E}[M_{\lambda n+n,r+2}].\end{aligned} \tag{47}$$

Estimating the expected values in (46) and (47) with their maximum possible values, we obtain respectively that

$$|\operatorname{Bias}(\hat{V}^I_{\lambda n,r} - V_{\lambda n,r})| \leq \mathbb{P}(I\leq r) + \left(\frac{\lambda}{\lambda+1}\right)^{k_I+1}\frac{2^{r+1}(\lambda+1)}{2\lambda+1}\binom{k_I+r+1}{k_I}$$

and

$$\mathrm{Var}(\hat{V}^I_{\lambda n,r} - V_{\lambda n,r}) \leq \left(\frac{\lambda}{\lambda+1}\right)^{2k_I}\frac{(r+1)2^{2k_I}}{(\lambda+1)^2n}\binom{k_I+r+1}{k_I}^2 + \frac{(r+1)}{(\lambda+1)n}.$$

The previous bounds are the basic building blocks to derive an estimate for the mean square error, indeed they yield

$$\begin{aligned}\mathrm{MSE}[V_{\lambda n,r} - \hat{V}^I_{\lambda n,r}] &= \mathbb{E}[(V_{\lambda n,r} - \hat{V}^I_{\lambda n,r})^2] = \mathrm{Bias}[\hat{V}^I_{\lambda n,r}]^2 + \mathrm{Var}[V_{\lambda n,r} - \hat{V}^I_{\lambda n,r}] \\ &\leq \left[\mathbb{P}(I\leq r) + \left(\frac{\lambda}{\lambda+1}\right)^{k_I+1}\frac{2^{r+1}(\lambda+1)}{2\lambda+1}\binom{k_I+r+1}{k_I}\right]^2 \\ &+ \left(\frac{\lambda}{\lambda+1}\right)^{2k_I}\frac{(r+1)2^{2k_I}}{(\lambda+1)^2n}\binom{k_I+r+1}{k_I}^2 + \frac{(r+1)}{(\lambda+1)n}.\end{aligned} \tag{48}$$

□

A.6 Details for the determination of k_L and k_I

First, we provide a guideline for choosing the parameter k_L. We recall here the definition of the NMSE given in (16)

$$\begin{aligned}\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) &= \frac{r^2}{\lambda^2 n^2}\mathbb{E}[(\hat{U}^L_{\lambda n,r} - U_{\lambda n,r})^2] \\ &= \frac{r^2}{\lambda^2 n^2}\left(\mathrm{Bias}(\hat{U}^L_{\lambda n,r})^2 + \mathrm{Var}(\hat{U}^L_{\lambda n,r} - U_{\lambda n,r})\right) = \frac{r^2}{\lambda^2 n^2}\mathrm{MSE}[\hat{U}^L_{\lambda n,r} - U_{\lambda n,r}]\end{aligned}$$

we can exploit the bound on the MSE (45) determined in Section A.4.3 to get

$$\begin{aligned}\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq\ & 2^{2k_L}\left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)}\binom{k_L+r}{k_L}^2\frac{r}{\lambda^2 n} + \left(\frac{\lambda}{\lambda+1}\right)^{r-1}\frac{r}{\lambda n} \\ &+ \left[\mathbb{P}(L \leq r-1)\left(\frac{\lambda}{\lambda+1}\right)^{r-1} + \frac{2^r}{2\lambda+1}\left(\frac{\lambda}{\lambda+1}\right)^{r+k_L}\binom{k_L+r}{r}\right]^2.\end{aligned} \tag{49}$$

Before proceeding with the bound of the NMSE, we now focus on the following term

$$\left(\frac{\lambda}{\lambda+1}\right)^{r-1}\mathbb{P}(L \leq r-1).$$

(i) assume that $k_L \leq r-1$, then

$$\begin{aligned}\left(\frac{\lambda}{\lambda+1}\right)^{r-1}\mathbb{P}(L \leq r-1) &\leq \left(\frac{\lambda}{\lambda+1}\right)^{r-1} = \left(\frac{\lambda}{\lambda+1}\right)^{r+k_L-1-k_L} \\ &= \left(\frac{\lambda}{\lambda+1}\right)^{r+k_L}\left(\frac{\lambda+1}{\lambda}\right)^{1+k_L} \leq \left(\frac{\lambda}{\lambda+1}\right)^{r+k_L} 2^{k_L+1} \\ &\leq 2^r\left(\frac{\lambda}{\lambda+1}\right)^{r+k_L}.\end{aligned}$$

(ii) assuming that $k_L \geq r$, first of all we observe that for any $k \leq r-1$ the following inequality holds true

$$\binom{k_L+r}{k} \leq \binom{k_L+r}{r-1}. \tag{50}$$

To prove (50), one has to show that the function

$$k \to \binom{k_L+r}{k} \tag{51}$$

attains its unique maximum value at $k^* = \lfloor (k_L+r-1)/2 \rfloor$. One can immediately realize that $r-1 \leq (k_L+r-1)/2$ iff $k_L \geq r-1$, thus the function (51) is increasing for any $k \leq r-1$, which implies the validity of (50).
As a consequence whenever $k_L \geq r$, the bound proceeds as follows:

$$\left(\frac{\lambda}{\lambda+1}\right)^{r-1}\mathbb{P}(L \leq r-1) = \left(\frac{\lambda}{\lambda+1}\right)^{r-1}\sum_{k=0}^{r-1}\binom{k_L+r}{k}\frac{1}{(\lambda+1)^k}\left(\frac{\lambda}{\lambda+1}\right)^{k_L+r-k}$$

$$
\begin{aligned}
&= \left(\frac{\lambda}{\lambda+1}\right)^{r-1+k_L+r} \sum_{k=0}^{r-1} \binom{k_L+r}{k} \frac{1}{\lambda^k} \\
&\leq \left(\frac{\lambda}{\lambda+1}\right)^{k_L+r} r \binom{k_L+r}{r-1}
\end{aligned}
$$

where we have observed that $1/\lambda \leq 1$, $\lambda/(\lambda+1) \leq 1$ and we have employed (50). From the previous inequality now follows

$$
\left(\frac{\lambda}{\lambda+1}\right)^{r-1} \mathbb{P}(L \leq r-1) \leq \left(\frac{\lambda}{\lambda+1}\right)^{k_L+r} r \binom{k_L+r}{r}.
$$

Thanks to (i) and (ii) we may conclude that

$$
\left(\frac{\lambda}{\lambda+1}\right)^{r-1} \mathbb{P}(L \leq r-1) \leq \left(\frac{\lambda}{\lambda+1}\right)^{k_L+r} 2^r r \binom{k_L+r}{r}. \tag{52}
$$

We now use the estimate (52) in the bound for the NMSE (49)

$$
\begin{aligned}
\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) &\leq 2^{2k_L} \left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \binom{k_L+r}{k_L}^2 \frac{r}{\lambda^2 n} + \left(\frac{\lambda}{\lambda+1}\right)^{r-1} \frac{r}{\lambda n} \\
&\quad + 2^{2r} \left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \binom{k_L+r}{r}^2 \left[r + \frac{1}{2\lambda+1}\right]^2 \\
&\leq \left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \binom{k_L+r}{k_L}^2 \left[\frac{2^{2k_L} r}{\lambda^2 n} + 2^{2r} \left(r + \frac{1}{2\lambda+1}\right)^2\right] \\
&\quad + \left(\frac{\lambda}{\lambda+1}\right)^{r-1} \frac{r}{\lambda n} \\
&\leq \left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \frac{(k_L+r)^{2r}}{(r!)^2} \left[\frac{2^{2k_L} r}{\lambda^2 n} + 2^{2r} \left(r + \frac{1}{3}\right)^2\right] \\
&\quad + \left(\frac{\lambda}{\lambda+1}\right)^{r-1} \frac{r}{\lambda n}
\end{aligned} \tag{53}
$$

At this point we should minimize the r.h.s. of (53) with respect to k_L, however this cannot be done analytically, as mentioned in Section 2. As a consequence we choose k_L as a function of n to obtain the best rate of convergence to zero of the bound (53). We note that

$$
\left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \frac{(k_L+r)^{2r}}{(r!)^2}
$$

goes to zero as $k_L \to +\infty$ and this is the best rate we can achieve, indeed

$$
2^{2r} \left(r + \frac{1}{3}\right)^2
$$

is a constant term which does not depend on n. Hence we suggest to choose k_L in such a way that the limit of $2^{2k_L} r/(\lambda^2 n)$ is a constant $C \in (0, +\infty)$ as $n \to +\infty$. A possible value could be

$$k_L := \frac{1}{2}\log_2(n).$$

In order to determine the best vale of k_I, we now have to deal with the MSE of $\hat{V}^I_{\lambda n,r}$, which coincides with the NMSE defined in (17):

$$\begin{aligned}\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) \leq & \left[\mathbb{P}(I \leq r) + \left(\frac{\lambda}{\lambda+1}\right)^{k_I+1} \frac{2^{r+1}(\lambda+1)}{2\lambda+1}\binom{k_I+r+1}{k_I}\right]^2 \\ & + \left(\frac{\lambda}{\lambda+1}\right)^{2k_I} \frac{(r+1)2^{2k_I}}{(\lambda+1)^2 n}\binom{k_I+r+1}{k_I}^2 + \frac{(r+1)}{(\lambda+1)n}\end{aligned}$$

where we have applied (48). Observe that $\mathbb{P}(I \leq r)$, being $I \sim \mathrm{Binom}(k_I+r+1, (\lambda+1)^{-1})$, may be estimated through equation (52) writing $r+1$ in place of r, hence

$$\begin{aligned}\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) \leq & \frac{(k_I+r+1)^{2(r+1)}}{((r+1)!)^2}\left(\frac{\lambda}{\lambda+1}\right)^{2(k_I+1)} \\ & \times \left[2^{2r+2}\left(r+1+\frac{\lambda+1}{2\lambda+1}\right)^2 + \frac{(r+1)2^{2k_I}}{\lambda^2 n}\right] + \frac{r+1}{n(\lambda+1)}.\end{aligned} \tag{54}$$

With the same argument as before we choose $k_I = \log_2(n)/2$. □

A.7 Proof of Theorem 3

The point of departure to prove (18) is the inequality (53), which amounts to be

$$\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq \left(\frac{\lambda}{\lambda+1}\right)^{2(k_L+r)} \frac{(k_L+r)^{2r}}{(r!)^2}\left[\frac{2^{2k_L} r}{\lambda^2 n} + 2^{2r}\left(r+\frac{1}{3}\right)^2\right] + \left(\frac{\lambda}{\lambda+1}\right)^{r-1}\frac{r}{\lambda n}$$

Choosing the value of k_L specified in Section A.5 the previous bound boils down to

$$\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq \left(\frac{1}{2}\log_2(n)+r\right)^{2r}\frac{1}{(r!)^2}\left(\frac{\lambda}{\lambda+1}\right)^{2r+\log_2(n)}\left(C_r + \frac{r}{\lambda^2}\right) + \left(\frac{\lambda}{\lambda+1}\right)^{r-1}\frac{r}{\lambda n}$$

having defined

$$C_r := 2^{2r}\left(r+\frac{1}{3}\right)^2.$$

An application of the Young's inequality yields

$$\begin{aligned}\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq & 2^{2r-1}\left(\frac{1}{2^{2r}}\log_2^{2r}(n) + r^{2r}\right)\frac{1}{(r!)^2}\left(\frac{\lambda}{\lambda+1}\right)^{2r+\log_2(n)}\left(C_r+\frac{r}{\lambda^2}\right) \\ & + \left(\frac{\lambda}{\lambda+1}\right)^{r-1}\frac{r}{\lambda n}\end{aligned}$$

$$= \frac{1}{2(r!)^2}\left(\log_2^{2r}(n) + (2r)^{2r}\right)\left(\frac{\lambda}{\lambda+1}\right)^{2r}\frac{1}{n^{\log_2(1+1/\lambda)}}\left(C_r + \frac{r}{\lambda^2}\right)$$
$$+ \left(\frac{\lambda}{\lambda+1}\right)^{r-1}\frac{r}{\lambda n}.$$

Now we put

$$A_r(\lambda) := \frac{1}{2(r!)^2}\left(1 + (2r)^{2r}\right)\left(\frac{\lambda}{\lambda+1}\right)^{2r}\left(C_r + \frac{r}{\lambda^2}\right) + \left(\frac{\lambda}{\lambda+1}\right)^{r-1}\frac{r}{\lambda}$$

thus we obtain

$$\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq A_r(\lambda)\frac{\log_2^{2r}(n)}{n^{\log_2(1+1/\lambda)}},$$

and (18) is now proved. Observe that the function $A_r(\lambda)$ is continuous in $[1, +\infty)$ and it is bounded as $\lambda \to +\infty$, indeed

$$\lim_{\lambda\to+\infty} A_r(\lambda) = \frac{(1 + (2r)^{2r})}{2(r!)^2}C_r.$$

The bound for $\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r})$ may be obtained in a similar fashion. In order to do this the point of departure is (54):

$$\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) \leq \frac{(k_I + r + 1)^{2(r+1)}}{((r+1)!)^2}\left(\frac{\lambda}{\lambda+1}\right)^{2(k_I+1)}$$
$$\times \left[2^{2r+2}\left(r + 1 + \frac{\lambda+1}{2\lambda+1}\right)^2 + \frac{(r+1)2^{2k_I}}{\lambda^2 n}\right] + \frac{r+1}{n(\lambda+1)},$$

choosing $k_I = 2^{-1}\log_2(n)$, the previous inequality can be rewritten as

$$\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) \leq \frac{1}{((r+1)!)^2}\left(\frac{1}{2}\log_2(n) + r + 1\right)^{2r+2}\left(\frac{\lambda}{\lambda+1}\right)^{2+\log_2(n)}$$
$$\times \left[2^{2r+2}\left(r + 1 + \frac{\lambda+1}{2\lambda+1}\right)^2 + \frac{r+1}{\lambda^2}\right] + \frac{r+1}{n(\lambda+1)}.$$

Observing that $(\lambda+1)/(2\lambda+1) \leq 2/3$ and putting

$$D_r := 2^{2r+2}\left(r + 1 + \frac{2}{3}\right)^2,$$

we get

$$\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) \leq \frac{1}{((r+1)!)^2}\left(\frac{1}{2}\log_2(n) + r + 1\right)^{2r+2}\left(\frac{\lambda}{\lambda+1}\right)^{2+\log_2(n)}$$
$$\times \left[D_r + \frac{r+1}{\lambda^2}\right] + \frac{r+1}{n(\lambda+1)}.$$

As before, one can apply the Young's inequality, and similar calculations finally yield

$$\mathscr{E}_{n,\lambda}(\hat{V}^I_{\lambda n,r}) \leq B_r(\lambda)\frac{\log_2^{2r+2}(n)}{n^{\log_2(1+1/\lambda)}},$$

where

$$B_r(\lambda) := \frac{1}{2((r+1)!)^2}(1+(2r+2)^{2r+2})\left(\frac{\lambda}{\lambda+1}\right)^2\left(D_r+\frac{r+1}{\lambda^2}\right)+\frac{r+1}{\lambda+1}.$$

The thesis now follows, having observed that the limit of $B_r(\lambda)$ is finite as $\lambda \to +\infty$.

A.8 Proof of Corollary 4

We focus on the proof of (21), the other bound (22) follows along the same lines. Thanks to (18) of Theorem 3 we know that

$$\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq A_r\frac{\log_2^{2r}(n)}{n^{\log_2(1+1/\lambda)}}$$

where the maximum A_r of the function $A_r(\lambda)$ exists and is finite (see Section A.7). For any $\delta > 0$ the inequality

$$A_r\frac{\log_2^{2r}(n)}{n^{\log_2(1+1/\lambda)}} \leq \delta$$

is satisfied if and only if

$$\lambda \leq \frac{1}{\exp\left\{\log(2)\left(2r\log_2(\log_2(n))+\log_2(A_r)+\log_2(1/\delta)\right)/\log_2(n)\right\}-1} =: \lambda^*$$

this tells us that the maximum value of λ such that $\mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq \delta$ is at least λ^*. Hence we have

$$\begin{aligned}\max\left\{\lambda:\ \mathscr{E}_{n,\lambda}(\hat{U}^L_{\lambda n,r}) \leq \delta\right\}&\frac{2r\log_2(\log_2(n))+\log_2(A_r)+\log_2(1/\delta)}{\log_2(n)}\\ &\geq \frac{(2r\log_2(\log_2(n))+\log_2(A_r)+\log_2(1/\delta))/\log_2(n)}{\exp\left\{\log(2)\left(2r\log_2(\log_2(n))+\log_2(A_r)+\log_2(1/\delta)\right)/\log_2(n)\right\}-1}\end{aligned}$$

and one can easily show the validity of (21) taking the limit of the previous inequality.

A.9 Details for the determination of (23) and (24)

Firstly we discuss the determination of (23). As $N \to +\infty$, and hence also $n \to +\infty$ since $N \overset{\text{a.s.}}{\sim} n$, the estimator $\hat{U}_{\lambda n,r}$ satisfies the asymptotic approximation For any $r \geq 1$, as $n \to +\infty$

$$\hat{U}_{\lambda n,r} = \lambda^r\sum_{j=0}^{N-r}\binom{j+r}{r}(-1)^j\lambda^j m_{n,r+j} \overset{\text{a.s.}}{\sim} \lambda^r\sum_{j=0}^{n-r}\binom{j+r}{r}(-1)^j\lambda^j\frac{\alpha\Gamma(r+j-\alpha)}{(r+j)!}n^\alpha\ell(n)$$

$$= \lambda^r \sum_{j=0}^{n-r} \frac{\binom{j+r}{r}\binom{r+j-\alpha-1}{r+j-1}}{r+j}(-1)^j \lambda^j \alpha\Gamma(1-\alpha)n^\alpha \ell(n)$$

$$\overset{\text{a.s.}}{\sim} \alpha k_n \lambda^r \sum_{j=0}^{n-r} \frac{\binom{j+r}{r}\binom{r+j-\alpha-1}{r+j-1}}{r+j}(-1)^j \lambda^j$$

where we have used Theorem 5 to say that

$$m_{n,r+j} \sim \frac{\alpha\Gamma(r+j-\alpha)}{(r+j)!} n^\alpha \ell(n), \quad k_n \sim \Gamma(1-\alpha)n^\alpha \ell(n)$$

as n goes to infinity.

We now discuss the determination of (24). As before , we may prove that

$$\hat{U}^L_{\lambda n,r} \overset{\text{a.s.}}{\sim} \frac{\alpha(r-\alpha-1)_{r-1}}{r!} k_n r \lambda^r \sum_{j=0}^{n-r} \frac{\binom{j+r}{r}\binom{r+j-\alpha-1}{r+j-1}}{r+j\binom{r-\alpha-1}{r-1}}(-1)^j \lambda^j \mathbb{P}(L \geq j+r) \tag{55}$$

as $n \to +\infty$. Now we focus on the sum over j in (55), recalling that L has a binomial distribution with parameters $(k_L + r, (\lambda+1)^{-1})$, we can write

$$\begin{aligned}
r\lambda^r &\sum_{j=0}^{n-r} \frac{\binom{j+r}{r}\binom{r+j-\alpha-1}{r+j-1}}{r+j\binom{r-\alpha-1}{r-1}}(-1)^j \lambda^j \mathbb{P}(L \geq j+r) \\
&= \sum_{j=r}^{n} \frac{(j-\alpha-1)_{j-1}}{(j-r)!(r-1-\alpha)_{r-1}}(-1)^{j+r}\lambda^j \mathbb{P}(L \geq j) \\
&= \sum_{j=r}^{n} \frac{(j-\alpha-1)_{j-1}}{(j-r)!(r-1-\alpha)_{r-1}}(-1)^{j+r} \sum_{z=j}^{k_L+r} \binom{z-1}{j-1} \frac{\lambda^z}{(\lambda+1)^z} \\
&= \sum_{j=0}^{n-r} \frac{(j+r-\alpha-1)_{j+r-1}}{j!(r-1-\alpha)_{r-1}}(-1)^{j} \sum_{z=j+r}^{k_L+r} \binom{z-1}{j+r-1} \left(\frac{\lambda}{\lambda+1}\right)^z \\
&= \sum_{j=0}^{n-r} \frac{(j+r-\alpha-1)_{j+r-1}}{j!(r-1-\alpha)_{r-1}}(-1)^{j} \sum_{z=j}^{k_L} \binom{z+r-1}{j+r-1} \left(\frac{\lambda}{\lambda+1}\right)^{z+r} \\
&= \frac{1}{(r-1-\alpha)_{r-1}} \sum_{z=0}^{k_L} \left(\frac{\lambda}{\lambda+1}\right)^{z+r} \sum_{j=0}^{z} (-1)^j \frac{(j+r-\alpha-1)_{j+r-1}}{j!} \binom{z+r-1}{j+r-1}.
\end{aligned}$$

The summation over j can be solved resorting to (Gradshtein and Ryzhik, 2007, identity 0.160.2), to obtain

$$\begin{aligned}
r\lambda^r &\sum_{j=0}^{n-r} \frac{\binom{j+r}{r}\binom{r+j-\alpha-1}{r+j-1}}{r+j\binom{r-\alpha-1}{r-1}}(-1)^j \lambda^j \mathbb{P}(L \geq j+r) \\
&= \frac{1}{(r-1-\alpha)_{r-1}} \sum_{z=0}^{k_L} \left(\frac{\lambda}{\lambda+1}\right)^{z+r} \frac{(z+r-1)!}{z!\Gamma(1-\alpha)} \frac{\Gamma(z+\alpha)\Gamma(r-\alpha)}{\Gamma(z+r)\Gamma(\alpha)}
\end{aligned}$$

$$= \frac{1}{(r-1-\alpha)_{r-1}} \sum_{z=0}^{k_L} \left(\frac{\lambda}{\lambda+1}\right)^{z+r} \frac{\Gamma(z+\alpha)\Gamma(r-\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)\Gamma(z+1)}$$
$$= \left(\frac{\lambda}{\lambda+1}\right)^{r} \frac{1}{\Gamma(\alpha)} \sum_{z=0}^{k_L} \frac{\Gamma(z+\alpha)}{\Gamma(z+1)} \left(\frac{\lambda}{\lambda+1}\right)^{z},$$

and now substituting the previous expression in (55), the (24) easily follows.

A.10 Details for the determination of (31) and (32)

We focus on the proof of (31), the confidence interval for $V_{\lambda n,r}$ follows along similar lines. First of all we bound the log–Laplace to obtain a concentration inequality for the difference $U_{\lambda n,r} - \hat{U}_{\lambda n,r}$. Indeed, exploiting the independence of the random variables $\{(N_{n,i}, N_{\lambda n,i})\}_{i\geq 1}$, for any $\mu \in \mathbb{R}$ we get

$$\log \mathbb{E}[e^{\mu(U_{\lambda n,r}-\hat{U}_{\lambda n,r})}] = \sum_{i=1}^{\infty} \log \mathbb{E}\left[\exp\left\{\sum_{j\geq 0} a_j \mathbb{1}_{\{N_{n,i}=r+j\}} - \mathbb{1}_{\{N_{n,i}=0\}}\mathbb{1}_{\{N_{\lambda n,i}=r\}}\right\}\right],$$

where

$$a_j := (-1)^j \binom{r+j}{r} \lambda^{j+r}.$$

We observe that the random variables appearing in the previous equation are bounded from above by the maximum value of the a_j's. Since the a_j's admit a unique maximum, as proved in Section A.2, we have

$$\left|\sum_{j\geq 0} a_j \mathbb{1}_{\{N_{n,i}=r+j\}} - \mathbb{1}_{\{N_{n,i}=0\}}\mathbb{1}_{\{N_{\lambda n,i}=r\}}\right| \leq |a_{j_1^*}| = \Psi_{\lambda,r}.$$

recalling that $\Psi_{\lambda,r}$ equals

$$\Psi_{\lambda,r} = \binom{r+j_1^*}{r} \lambda^{j_1^*+r}$$

and $j_1^* = \lfloor r\lambda/(1-\lambda) - 1\rfloor \vee 0$ for $r \geq 1$. Fixed $x > 0$, the Bennett inequality may be now applied to obtain

$$\log \mathbb{E}[e^{\mu(U_{\lambda n,r}-\hat{U}_{\lambda n,r})}] \leq \sum_{i=1}^{\infty} \mathrm{Var}\left(\sum_{j\geq 0} a_j \mathbb{1}_{\{N_{n,i}=r+j\}} - \mathbb{1}_{\{N_i(t)=0\}}\mathbb{1}_{\{N_{\lambda n,i}=r\}}\right) \frac{\phi(\Psi_{\lambda,r}\mu)}{\Psi_{\lambda,r}^2}$$
$$= \mathrm{Var}(U_{\lambda n,r} - \hat{U}_{\lambda n,r}) \frac{\phi(\Psi_{\lambda,r}\mu)}{\Psi_{\lambda,r}^2},$$

where $\phi(\lambda) := e^{\lambda} - 1 - \lambda$. Thanks to this bound on the log–Laplace of the difference $U_{\lambda n,r} - \hat{U}_{\lambda n,r}$, the following Bernstein inequality holds true

$$\mathbb{P}(|U_{\lambda n,r} - \hat{U}_{\lambda n,r}| \geq x) \leq 2\exp\left\{-\frac{x^2}{2(\mathrm{Var}(U_{\lambda n,r} - \hat{U}_{\lambda n,r}) + \Psi_{\lambda,r}x/3)}\right\},$$

and (31) follows by putting

$$\frac{x^2}{2(\mathrm{Var}(U_{\lambda n,r} - \hat{U}_{\lambda n,r}) + \Psi_{\lambda,r} x/3)} = s.$$

□

References

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of extremes: theory and applications*, Wiley.

Ben-Hamou, A., Boucheron, S. and Ohannessian, M.I. (2017). Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications. *Bernoulli*, **23**, 249–287.

Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration inequalities*, Oxford University Press.

Bubeck, S., Ernst, D. and Garivier, A. (2013). Optimal discovery with probabilistic expert advice: finite time analysis and macroscopic optimality. *J. Mach. Learn. Res.*, **14**, 601–623.

Bunge, J. and Fitzpatrick, M. (1993). Estimating the number of species: a review. *J. Amer. Statist. Assoc.*, **88**, 364–373.

Bunge, J., Willis, A. and Walsh, F. (2014). Estimating the number of species in microbial diversity studies. *Annu. Rev. Sta. Appl.*, **1**, 427–445.

Chao, A. (2005). Species estimation and applications. In *Encyclopedia of Statistical Sciences*, **12**, 7907–7916.

Chao, A., Colwell, R.K., Lin, C. and Gotelli, N.J. (2009). Sufficient sampling for asymptotic minimum species richness estimators. *Ecology*, **90**, 1125–1133.

Chao, A. and Chiu, C. (2005). Species richness: estimation and comparison. In *Wiley StatsRef: Statistics Reference Online*, to appear.

Efron, B. and Thisted, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika*, **63**, 435–447.

Favaro, S., Lijoi, A., Mena, R.H. and Prünster, I. (2009). Bayesian nonparametric inference for species variety with a two parameter Poisson-Dirichlet process prior. *J. Roy. Statist. Soc. Ser. B*, **71**, 993–1008.

Favaro, S., Lijoi, A. and Prünster, I. (2012). A new estimator of the discovery probability. *Biometrics*, **68**, 1188–1196.

Fisher, R.A., Corbet, A.S. and Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.*, **12**, 42–58.

Florencio, D. and Herley, C. (2007) A large-scale study of web password habits. *Proceedings of the 16th International Conference on World Wide Web*, 657–666.

Gnedin, A., Hansen, B. and Pitman, J. (2007). Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv.*, **4**, 146–171.

Gnedin, A. and Pitman, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci.*, **138**, 5674–5685.

Good, I.J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237–264.

Good, I.J. and Toulmin, G.H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika*, **43**, 45–63.

Gradshtein, I.S. and Ryzhik, I.M. (2007). *Tables of integrals, sums, series, and products*, Academic Press, seventh edition.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory.* Springer-Verlag, 2006.

Haas, P.J., Naughton, J.F., Seshadri, S. and Stokes, L. (1995) Sampling-based estimation of the number of distinct values of an attribute. *Proceedings of the 21st VLDB Conference*, 311–322.

Harris, B. (1968), Statistical inference in the classical occupancy problem, unbiased estimation of the number of classes. *J. Amer. Statist. Assoc.*, **63**, 837–847.

Karlin, S. (1967). Central limit theorems for certain infinite urn schemes. *J. Math. Mech.*, **17**, 373–401.

Laird, N.M. and Lange, C. (2003). *The fundamentals of modern statistical genetics*, Springer.

Lijoi, A., Mena, R.H. and Prünster, I. (2007). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika*, **94**, 769-786.

McAllester, D. and Schapire, R.E. (2000). On the convergence rate of Good-Turing estimator. In *13th Annual Conference on Computational Learning Theory*,

Magurran, A.E. (2003). *Measuring biological diversity*, Wiley, Blackwell.

Maritz, J.S. and Lwin, T. (1989). *Empirical Bayes methods with applications*, Chapman and Hall.

Mossel, E. and Ohannessian, M.I. (2015). On the impossibility of learning the missing mass. *Preprint: arXiv:1503.03613.*

Ohannessian, M.I. and Dahleh, M.A. (2012). Rare probability estimation under regularly varying heavy tails. *J. Mach. Learn. Res.*, **23**, 1–24.

Olver, F.W.J., Lozier, D.W., Boisvert, R.F. and Clark, C.W. (2010). *NIST Handbook of mathematical functions*, Cambridge University Press.

Orlitsky, A., Santhanam, N.P. and Zhang, J. (2003). Always Good Turing: asymptotically optimal probability estimation. *Science*, **302**, 427–431.

Orlitsky, A., Suresh, A.T. and Wu, Y. (2017). Optimal prediction of the number of unseen species. *Proc. Natl. Acad. Sci. USA*, to appear.

Perman, M., Pitman, J. and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields.* **92**, 21–39.

Pitman, J. and Yor, M. (1997). The two parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900.

Pitman, J. (2006). *Combinatorial Stochastic Processes.* Ecole d'Eté de Probabilités de Saint-Flour XXXII. Lecture notes in mathematics, Springer - New York.

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. 3rd Berkeley Symp.*, **1**, 157–163.

Robbins, H. (1968). Estimating the total probability of the unobserved outcomes of an experiment. *Ann. Statist.*, **39**, 256–257.

Thompson, W.K. (2004). *Sampling rare or elusive species*, Island Press.

Vardi, Y. (1996). Network tomography: Estimating source-destination traffic intensities from link data. *J. Amer. Statist. Assoc.*, **91**, 365–377.

Zhang, C.H. (2005). Estimation of sums of random variables: examples and information bounds. *Ann. Statist.*, **33**, 2022–2041.