Large Deviations Approach to Bayesian Nonparametric Consistency: the Case of Polya Urn Sampling

M. Grendar * G. G. Judge †

R. K. Niven ‡

*Institute of Measurement Sciences SAS, Bratislava, Slovakia

†University of California, Berkeley and Giannini Foundation

‡The University of New South Wales at ADFA

Large Deviations Approach to Bayesian Nonparametric Consistency: the Case of Polya Urn Sampling

Abstract

The Bayesian Sanov Theorem (BST) identifies, under both correct and incorrect specification of infinite dimensional model, the points of concentration of the posterior measure. Utilizing this insight in the context of Polya urn sampling, Bayesian nonparametric consistency is established. Polya BST is also used to provide an extension of Maximum Non-parametric Likelihood and Empirical Likelihood methods to the Polya case.

Large Deviations Approach to Bayesian Nonparametric Consistency: the Case of Pólya Urn Sampling

M. Grendár,* G. G. Judge,† and R. K. Niven‡

Sep 21, 2007

Abstract

The Bayesian Sanov Theorem (BST) identifies, under both correct and incorrect specification of infinite dimensional model, the points of concentration of the posterior measure. Utilizing this insight in the context of Pólya urn sampling, Bayesian nonparametric consistency is established. Pólya BST is also used to provide an extension of Maximum Non-parametric Likelihood and Empirical Likelihood methods to the Pólya case.

Keywords: Pólya L-divergence, Bayesian Maximum (A Posteriori) Probability method, Maximum Non-parametric Likelihood method, Empirical Likelihood method

AMS: 60F10, 60F15

1 Introduction

In Bayesian nonparametric (or infinite dimensional) statistics a strictly positive prior is put over a set Φ of probability distributions. In this context let r be the true data sampling distribution of a random sample $X^n \triangleq X_1, X_2, \ldots, X_n$. Provided that $r \in \Phi$, as the sample size grows to infinity, the posterior distribution $\pi(\cdot|X^n = x^n)$ over Φ is expected to concentrate in a neighborhood of the true sampling distribution r. Whether and under what conditions this indeed happens is a subject of Bayesian nonparametric consistency investigations. Surveys of the subject include [7], [9], [25], [23], [24], [26].

More formally, as in [7], consistency of a sequence of posteriors with respect to a metric d can be defined as follows: The sequence $\{\pi(\cdot|X^n), n \geq 1\}$ is said to be d-consistent at r, if there exists a $\Omega_0 \subset \mathbb{R}^\infty$ with $r(\Omega_0) = 1$ such that for $\omega \in \Omega_0$, for every neighborhood U of r, $\pi(U|X^n) \to 1$ as n goes to infinity. If a posterior is d-consistent for any $r \in \Phi$ then it is said to be d-consistent. There, two modes of convergence are usually considered: convergence in probability and almost sure convergence and d is usually either Hellinger distance or a metric which metricizes weak topology.

Freedman's [5] classic theorem on Bayesian nonparametric consistency for X taking on values from a finite set was in [1], [2], and independently in [6] proved by means of a Bayesian Sanov Theorem (BST). In [11] the consistency was via BST established for a countable set of densities. BST (a.k.a. Sanov Theorem for Sampling Distributions) is Bayesian counterpart of Sanov Theorem for Empirical Measures [21], [3]. The latter is a basic result of Large Deviations (LD) theory [4]. LD theory is a sub-field of probability theory where, informally, the typical concern is about the asymptotic behavior, on a logarithmic scale, of the probability of a given event. To promote and extend the Bayesian large deviations approach, we study Bayesian nonparametric consistency for a basic non-*iid* setting, where data are drawn according to multicolor Pólya urn scheme. We demonstrate that data sampling distributions from the set Φ asymptotically *a posteriori* concentrate on the Pólya L-projection(s) of the true sampling distribution r on Φ. The statement holds also under misspecifition (i.e., when $r \notin \Phi$).

In [13] BST [11] was used to provide a probabilistic interpretation and justification of the Empirical Likelihood (EL) method. Based on the Pólya BST pre-

*Department of Mathematics, FPV UMB, Tajovského 40, 974 01 Banská Bystrica, Slovakia; Institute of Mathematics and CS of the Slovak Academy of Sciences (SAS), Institute of Measurement Science of SAS. Email: marian.grendar@savba.sk. To whom correspondence should be addressed. Supported by VEGA 1/3016/06 grant.

†Graduate School, 207 Giannini Hall, University of California, Berkeley, CA, 94720. Email: judge@are.berkeley.edu.

‡The School of Aerospace, Civil and Mechanical Engineering, The University of New South Wales at ADFA, Canberra, ACT, 2600, Australia; and Niels Bohr Institute, Copenhagen University, 2100 Copenhagen Ø, Denmark; Email: r.niven@adfa.edu.au.

sented in this paper we extend EL and Maximum Non-parametric Likelihood methods to the Pólya sampling case.

The paper is organized as follows. First, Pólya urn sampling is briefly described. Next, the urn is embedded into a Bayesian setting. The Pólya L-divergence, which governs the exponential decay of posterior probability, is introduced next and then a Bayesian Sanov Theorem for Pólya Sampling is stated and proved. It directly implies Bayesian nonparametric consistency. Next, the consistency result is used to provide an extension of Maximum Non-parametric Likelihood and Empirical Likelihood methods to the Pólya sampling case. Finally, using the Bayesian large deviations approach, we provide a couple of insights into Bayesian consistency.

2 Multicolor Pólya Urn Sampling

Consider an urn containing $\alpha_i > 0$ balls of colors i, $i = 1, 2, \dots, m$; and let m be finite. There is a total number $N \triangleq \sum_{i=1}^{M} \alpha_i$ of balls in the urn and when it is necessary to stress it, the urn will be called an N-urn. We identify the set of possible colors with support $\mathscr{X}$ of a random variable X. A single ball is drawn from the urn, recorded and then returned together with $c \in \mathbb{Z}$ balls of the same color. Assuming $-nc \leq \min(\alpha_1, \alpha_2, \dots, \alpha_m)$, the drawing is repeated n times. This sampling is known as the multicolor Pólya Eggenberger (PE) urn scheme; c.f. [8], [22], [15]. Prominent special cases of the PE scheme are: *iid* sampling ($c = 0$), sampling without replacement ($c = -1$), and the case of $c = 1$.

Given the PE scheme, the probability $\pi(X^n = x^n \,|\, q^N; c)$ that a sequence x^n of n balls will be drawn from initial configuration q^N of N balls is (c.f. [22], [15]):

$$\pi(X^n = x^n \,|\, q^N; c) \triangleq \frac{\prod_{i=1}^{m} \alpha_i(\alpha_i + c)\cdots(\alpha_i + (n_i - 1)c)}{N(N + c)\cdots(N + (n - 1)c)}, \quad (1)$$

where vector q^N consists of $q_i^N \triangleq \frac{\alpha_i}{N}$, and n_i is the number of occurrences of the i-th outcome from $\mathscr{X}$ in the sample x^n, $i = 1, 2, \dots, m$.

3 Bayesian Embedding

Let $\mathscr{P}(\mathscr{X})$ be set of all probability mass functions with the support $\mathscr{X}$. Let $\Phi \subseteq \mathscr{P}(\mathscr{X})$ and let Φ_N denote intersection of Φ with the set of all possible configurations of the N-urn. Let Φ_N be the support of prior distribution $\pi(q^N)$ of initial configurations q^N of N-urn.

Let r^N be the true initial configuration of N-urn where r^N is not necessarily in Φ_N. From r^N a sequence x^n is drawn according to the Pólya sampling scheme, that we characterize by the parameter c. Consequently, under this framework the bayesian arrives at the posterior probability distribution $\pi(q^N \,|\, X^n = x^n; c)$ of initial configurations of the N-urn.

4 Pólya L-divergence

The Pólya L-divergence $L_\beta^c(q \,||\, p)$ of the probability mass function (pmf) $q \in \mathscr{P}(\mathscr{X})$ with respect to pmf $p \in \mathscr{P}(\mathscr{X})$ is

$$L_\beta^c(q \,||\, p) \triangleq -\sum_{i=1}^{m} p_i \log(q_i + \beta c p_i) + \frac{1}{\beta c} \sum_{i=1}^{m} q_i \log \frac{q_i}{q_i + \beta c p_i}.$$

By the continuity argument $L_\beta^0(q \,||\, p) \triangleq -\sum_{i=1}^{m} p_i \log q_i - 1$. More concisely, $L_\beta^c(q \,||\, p) = L(q + \beta c p \,||\, p) + \frac{1}{\beta c} I(q \,||\, q + \beta c p)$, where using standard conventions $I(\cdot \,||\, \cdot)$ is the I-divergence $I(a \,||\, b) \triangleq \sum a_i \log \frac{a_i}{b_i}$ [3] and $L(\cdot \,||\, \cdot)$ is the L-divergence $L(b \,||\, a) \triangleq -\sum a_i \log b_i$ [10], [11]. Though base of the logarithm is immaterial, the natural logarithm will be used.

The Pólya L_β^c-projection $\hat{q}$ of p on $\mathscr{A} \subseteq \mathscr{P}(\mathscr{X})$ is $\hat{q} \triangleq \arg\inf_{q \in \mathscr{A}} L_\beta^c(q \,||\, p)$. The value of L_β^c-divergence at an L_β^c-projection of p on $\mathscr{A}$ is denoted by $L_\beta^c(\mathscr{A} || p)$. Hereafter it is assumed finite.

5 Bayesian Sanov Theorem for Pólya Sampling

Sanov Theorem for Empirical Measures [21] is well-known; reader is directed to [4], [3]. Initiated by [18], the Sanov Theorem for Pólya Sampling was recently proven in [12]. The Bayesian counterpart of Sanov Theorem was to the best of our knowledge first studied in [1] and [2] and independently in [6]. Our proof of Bayesian Sanov Theorem (BST) for Pólya Sampling is based on [11], and also utilizes tools from [12].

Asymptotic investigations of posterior consistency will be carried on under the following assumptions: 1) n and N go to infinity in such a way that $\beta(n) \triangleq \frac{n}{N} \to \beta \in (0,1)$ as $n \to \infty$, 2), r^N converges in the total variation metrics to $r \in \mathscr{P}(\mathscr{X})$ as $n \to \infty$.

Topological qualifiers are meant in topology induced on the m-dimensional simplex by the usual topology on $\mathbb{R}^m$.

Bayesian Sanov Theorem for Pólya Sampling. *Let $\mathscr{A} \subset \Phi$ be an open set. Let $\beta(n) \to \beta \in (0,1)$, $r^N \to r$ as $n \to \infty$. Then, for $n \to \infty$,*

$$\frac{1}{n}\log \pi(q^N \in \mathscr{A}\,|\,x^n;c) = -\left\{L^c_\beta(\mathscr{A}\,||\,r) - L^c_\beta(\Phi\,||\,r)\right\}$$

with probability one.

Proof. The proof will be constructed separately for $c > 0$, $c < 0$, $c = 0$.

For $c \neq 0 \wedge \frac{Nq}{c} \notin (\mathbb{Z}^-)^m \wedge \frac{N}{c} \notin \mathbb{Z}^-$, formula (1) can equivalently be expressed as [15]:

$$\pi(x^n\,|\,q^N;c) = \frac{\Gamma\left(\frac{N}{c}\right)}{\Gamma\left(\frac{N}{c}+n\right)} \prod_{i=1}^m \frac{\Gamma\left(\frac{Nq_i}{c}+n_i\right)}{\Gamma\left(\frac{Nq_i}{c}\right)}, \qquad (2)$$

where $\Gamma(\cdot)$ is the Gamma function.

The following bounds [16] on the Gamma function $\Gamma(\cdot)$ are imposed:

$$(b-1)\log b - (a-1)\log a - (b-a) < \log\frac{\Gamma(b)}{\Gamma(a)} <$$
$$< \left(b-\frac{1}{2}\right)\log b - \left(a-\frac{1}{2}\right)\log a - (b-a), \qquad (3)$$

which is valid for $0 < a < b$.

Let $c > 0$ and note that the other restrictions under which (1) and (2) are equivalent are not active, since $-nc \leq \min(\alpha_1, \alpha_2, \dots, \alpha_m)$. We use the bounds in (3) to get the upper bound U_n of the probability $\pi(q^N \in \mathscr{A}\,|\,x^n;c)$:

$$U_n = \frac{\sum_{q^N\in\mathscr{A}} \pi(q^N) \prod_{i=1}^m e^{n\,l(q_i^N,\frac{1}{2n})}}{\sum_{q^N\in\Phi} \pi(q^N) \prod_{i=1}^m e^{n\,l(q_i^N,\frac{1}{n})}}$$

and the lower bound L_n similarly (to get L_n just replace $1/2n$ with $1/n$ in U_n). There,

$$l(q_i^N,\alpha) \triangleq -\left(\frac{q_i^N}{\beta(n)c} - \alpha\right)\log\left(\frac{q_i^N}{\beta(n)c}\right) +$$
$$+\left(\frac{q_i^N}{\beta(n)c} + \nu_i^n - \alpha\right)\log\left(\frac{q_i^N}{\beta(n)c} + \nu_i^n\right),$$

$\alpha \in \left\{\frac{1}{n}, \frac{1}{2n}\right\}$ and ν^n is the empirical measure induced by the sample x^n.

Next, we use the simple bounds of [11] to develop the upper upper bound U_n by $\bar{U}_n$ and lower bound L_n by $\underline{L}_n$, as follows:

$$\bar{U}_n = \frac{\prod_{i=1}^m e^{n\,l\left(\hat{q}_i^N(\mathscr{A},\frac{1}{2n}),\frac{1}{2n}\right)}}{\pi(\tilde{q}^N)\prod_{i=1}^m e^{n\,l\left(\tilde{q}_i^N(\Phi,\frac{1}{n}),\frac{1}{n}\right)}},$$
$$\underline{L}_n = \frac{\prod_{i=1}^m e^{n\,l\left(\hat{q}_i^N(\mathscr{A},\frac{1}{n}),\frac{1}{n}\right)}}{\pi(\tilde{q}^N)\prod_{i=1}^m e^{n\,l\left(\tilde{q}_i^N(\Phi,\frac{1}{2n}),\frac{1}{2n}\right)}},$$

where

$$\hat{q}^N(\mathscr{S},\alpha) \triangleq \arg\sup_{q^N\in\mathscr{S}} \sum_{i=1}^m l(q_i^N,\alpha),$$
$$\tilde{q}^N(\mathscr{S},\alpha) \triangleq \arg\sup_{q^N\in\mathscr{S}} \left(\sum_{i=1}^m l(q_i^N,\alpha) - \frac{\log\pi(q^N)}{n}\right).$$

By the Strong Law of Large Numbers for Pólya Sampling (which follows from [12], Thm. 2 and Borel Cantelli Lemma), $\nu^n \to r$, almost surely, as $n \to \infty$. The Pólya L-divergence is continuous in q and $\mathscr{A}$ is open, by assumption. Thus, $\frac{1}{n}\log\bar{U}_n$ converges, with probability one, to $-\left\{L^c_\beta(\mathscr{A}\,||\,r) - L^c_\beta(\Phi\,||\,r)\right\}$, as $n \to \infty$. This is the same as the 'point' of almost sure convergence of $\frac{1}{n}\log\underline{L}_n$ and the Theorem for $c > 0$ is thus proven.

For $c \neq 0 \wedge (1-\frac{Nq}{c}) \notin (\mathbb{Z}^-)^m \wedge (1-\frac{N}{c}) \notin \mathbb{Z}^-$, the formula (1) can equivalently be expressed as (cf. [12]):

$$\pi(x^n\,|\,q^N;c) = \frac{\Gamma\left(1-\frac{N}{c}-n\right)}{\Gamma\left(1-\frac{N}{c}\right)} \prod_{i=1}^m \frac{\Gamma\left(1-\frac{Nq_i}{c}\right)}{\Gamma\left(1-\frac{Nq_i}{c}-n_i\right)}. \qquad (4)$$

Let $c < 0$. Note that the other restrictions under which (1) and (4) are equivalent are not active, since $-nc \leq \min(\alpha_1, \alpha_2, \dots, \alpha_m)$. The proof then can be constructed along the same lines as for $c > 0$. At the final stage one arrives at an expression of the form $\sum_{i=1}^m \left\{\left(\frac{-q_i}{\beta c} - r_i\right)\log\left(\frac{-q_i}{\beta c} - r_i\right) - \left(\frac{-q_i}{\beta c}\right)\log\left(\frac{-q_i}{\beta c}\right)\right\}$ which, after little algebra, can be seen to be the Pólya L-divergence.

The case of $c = 0$ (i.e., *iid* urn) has already been studied in [11]. The exponential decay rate function is $L(\mathscr{A}\,||\,r) - L(\Phi\,||\,r)$, which is the same as $L^0_\beta(\mathscr{A}\,||\,r) - L^0_\beta(\Phi\,||\,r)$, implied by the continuity of the Pólya L-divergence $L^c_\beta(\cdot\,||\,\cdot)$. □

6 Bayesian Nonparametric Consistency for Pólya Sampling

Bayesian nonparametric consistency for Pólya sampling is just a corollary of the above Pólya BST.

To see this let for $\varepsilon > 0$, $\mathscr{A}^C_\varepsilon(\Phi) \triangleq \{q : L^c_\beta(q\,||\,r) - L^c_\beta(\Phi\,||\,r) > \varepsilon, q \in \Phi\}$.

Corollary. *Let there be a finite number of Pólya L-projections of r on* Φ. *As* $n \to \infty$, $\pi(q \in \mathscr{A}^C_\varepsilon(\Phi)\,|\,x^n; c) \to 0$, *with probability one.*

Standard Bayesian consistency (i.e., under correct specification; $r \in \Phi$) follows as a special case of the Corollary.

Informally, the posterior probability concentrates on the Pólya L-projections of r on Φ. Observe that the Pólya L-projection of r on Φ is an asymptotic instance of the sampling distribution with the supremal (over Φ) value of the posterior probability; hence it is asymptotic form of the Maximum A-posteriori Probable sampling distribution.

7 MNPL and EL in Pólya case

Consider the problem of selecting an initial Pólya urn composition q^N from a set Φ of such compositions, when there is a sample x^n (or empirical pmf ν^n which the sample induces) drawn from the 'true' urn r^N, according to PE sampling scheme with parameter c. Pólya BST dictates that we select in the asymptotic case ($n \to \infty$, $\nu^n \to r$, $\beta(n) \to \beta$) L^c_β-projection or r on Φ.

Most commonly, Φ is formed by moment constraints that define a linear family of distributions $\mathscr{L}(u) \triangleq \{q : \sum_{\mathscr{X}} q_i(u_j(x_i) - a_j) = 0, j = 0, 1, 2, \ldots, J\}$, where u_j is a real-valued function on $\mathscr{X}$, $u_0 = 1 \in \mathbb{R}^m$, $a \in \mathbb{R}^{J+1}$, $a_0 = 1$. The Pólya L-projection which Pólya BST selects in this case has the form

$$\hat{q}_i(\beta, c, \lambda) = \frac{\beta c r_i}{e^{\beta c \sum_{j=0}^J \lambda_j (u_j(x_i) - a_j)} - 1},$$

where $\lambda \in \mathbb{R}^{J+1}$ are Lagrange multipliers.

If the choice is to be made among all possible q^N for a fixed, sufficiently large N, then by Pólya BST, we should select the Pólya L-projection $\hat{q}(\beta, c, \lambda)$ of r on $\mathscr{P}(\mathscr{X})$, which in this case is just r, regardless of c and β.

For n, N not sufficiently large, there are two possibilities. It is possible either to select the initial configuration with highest value of the posterior probability[1] $\pi(q^N\,|\,\nu^n; c)$ or the Pólya L-projection of ν^n on Φ^N. As $n \to \infty$, the two methods select the same configuration(s). Observe that when $c = 0$, the latter method selects configuration(s) with the highest value of non-parametric likelihood $\sum_{i=1}^m \nu_i^n \log q_i$. Pólya BST thus extends Maximum Non-parametric Likelihood (MNPL) method into the Pólya sampling: Pólya MNPL selects the urn configuration(s) with the highest value of negative of $L^c_\beta(q\,||\,\nu^n)$.

In the case of *iid* sampling it was observed (cf. [13]) that the Bayesian Sanov Theorem [11] provides a probabilistic Bayesian interpretation and justification of Empirical Likelihood method [19], [17] in the parameter estimation context [20]. EL, viewed as estimation method, double-maximizes the non-parametric likelihood criterial function subject to parametrized constraints [20], [19], [17]. The above discussion thus directly shows how to extend EL into the Pólya sampling framework: the negative of Pólya L-divergence has to be double-maximized subject to parametric constraints.

8 Summary

The main advantage of Bayesian Sanov Theorem (BST) approach to Bayesian nonparametric consistency over the traditional one lays not that much on the technical side as on the conceptual one. BST identifies the rate function governing exponential decay of the posterior measure, and this in turn identifies 'points' of concentration of the posterior as those distributions which minimize the rate function. In the case of *i.i.d.* sampling the posterior concentration 'points' identified by BST are those distributions $\hat{q}$ which in the feasible set Φ maximize $\int r \log q$. If Φ is the set of all distributions (with the same support), then $\hat{q}$ is unique and equal r. Traditional approaches to Bayesian nonparametric consistency do not see the concentration point (i.e., r, under correct specification) as a solution of the optimization problem.

BST also shows that the 'points' of asymptotic concentration of posterior probability are asymptotic instances of *a posteriori* most probable (MAP) sampling distributions. This fact implies that the mean posterior sampling distribution (i.e., the predictive distribution)

[1] Hence the name of the method associated with this selection scheme: Bayesian Maximum Probability method; cf. [14]

is, in general, not the point of convergence under misspecification.

In this paper we have used under the Pólya sampling scheme the Bayesian Sanov Theorem (BST) to identify sampling distributions, on which the posterior probability asymptotically concentrates. This way, Bayesian nonparametric consistency for Pólya sampling was established both under correct specification of model as well as under misspecification.

In [13] it was pointed out that the non-parametric likelihood criterion, as well as methods that are based on its maximization (i.e., Maximum Non-parametric Likelihood (MNPL) and Empirical Likelihood (EL) methods) are limited to independent sampling. The point was made on the grounds of a Bayesian large deviations interpretation of the methods. On the same ground the Pólya extension of BST implies that under Pólya sampling it is the Pólya non-parametric likelihood function (i.e., negative of the Pólya L-divergence) that has to be maximized.

References

[1] Ben-Tal, A., Brown, D. E. and R. L. Smith. (1987). Posterior convergence under incomplete information. Tech. rep. 87-23. U. of Michigan, Ann Arbor.

[2] Ben-Tal, A., Brown, D. E. and R. L. Smith. (1988). Relative Entropy and the convergence of the posterior and empirical distributions under incomplete and conflicting information. Tech. rep. 88-12. U. of Michigan.

[3] Csiszár, I. and P. Shields. (2004). Information theory and statistics: a tutorial. *Foundations and Trends in Communications and Information Theory*, vol. 1, pp. 417-528.

[4] Dembo, A. and O. Zeitouni. (1998). *Large Deviations Techniques and Applications*. New York:Springer-Verlag.

[5] Freedman, D. A. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *Ann. Math. Statist.*, 34:1386-1403.

[6] Ganesh, A. and N. O'Connell. (1999). An inverse of Sanov's Theorem. *Stat. & Prob. Letters*, 42:201–206.

[7] Ghosal, A., Ghosh, J. K. and R. V. Ramamoorthi (1999). Consistency issues in Bayesian Nonanparametrics. *Asymptotics, Nonparametrics and Time Series: A Tribute to Madan Lal Puri*, 639–667, Dekker.

[8] Eggenberger, F. and G. Pólya. (1923). Über die Statistik verketter Vorgänge. *Z. Angew. Math. Mech.*, vol. 1, pp. 279-289.

[9] Ghosh , J. and R. V. Ramamoorthi. (1999). *Bayesian Nonparametrics*. New York: Springer Verlag.

[10] Grendár, M. (2005). Conditioning by rare sources. *Acta Univ. Belii, Math.* 12: 19-29. Online at http://actamath.savbb.sk.

[11] Grendár, M. (2006) L-divergence consistency for a discrete prior. *Jour. Stat. Res.* 40/1:73-76. Corrected at arXiv.

[12] Grendár, M and R. K. Niven. (2006). The Pólya urn: Limit Theorems, Pólya Divergence, Maximum Entropy and Maximum Probability. Available at arXiv:cond-mat/0612697; *submitted*.

[13] Grendár, M. and G. Judge. (2007). A Bayesian Large Deviations Probabilistic Interpretation and Justification of Empirical Likelihood. Preprint, ARE UC, Berkeley; *under revision*. Available at http://repositories.cdlib.org/are_ucb/1035. There is a typo in LST, inherited from [11]. The sets M^e and $\mathcal{N}$ should be replaced. Also, N_ε^C should be modified accordingly.

[14] Grendár, M. (2007). Maximum Probability and Maximum Entropy, Bayesian Maximum Probability and Maximum Non-parametric Likelihood. (Probabilistic Regularization of Inverse Problems), *unpublished manuscript*.

[15] Johnson, N. L., Kotz, S. and N. Balakrishnan. (1997). *Discrete Multivariate Distributions.* New York: Wiley.

[16] Kečkić, J. D. and P. M. Vasić. (1971). Some inequalities for the gamma function. *Publ. Inst. Math. (Beograd)(N.S.)*, vol. 11, pp. 107-114.

[17] Mittelhammer, R., Judge, G. and D. Miller (2000). *Econometric Foundations*. Cambridge:CUP.

[18] Niven, R. K. (2005). Combinatorial Information Theory: I. philosophical basis of cross-entropy and entropy. Available at arXiv:cond-mat/0512017; ver. 3 of July, 2006.

[19] Owen, A. (2001). *Empirical Likelihood*. New York:Chapman-Hall/CRC.

[20] Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* 22:300-325.

[21] Sanov, I. N. (1957). On the probability of large deviations of random variables. *Mat. Sbornik,* **42**, pp. 11–44. (In Russian).

[22] Steyn, H. S. (1951). On discrete multivariate probability functions. *Proceedings Koninklijke Nderlandse Akademie van Wetenschappen, Ser. A*, vol. 54, pp. 23-30.

[23] Walker, S. (2004). New appraoches to Bayesian consistency. *Ann. Statist.* 32:2028–2043.

[24] Walker, S. (2004). Modern Bayesian asymptotics. *Stat. Sci.* 19:111-117.

[25] Walker, S., Lijoi, A. and I. Prünster (2004). Contributions to the understanding of Bayesian consistency. *working paper*.

[26] Wasserman, L. (1998). Asymptotic properties of nonparametric Bayesian procedures. In *Practical Nonparametric and Semiparametric Bayesian Statistics*. New York: Springer Verlag.