



# Confidence regions for the multinomial parameter with small sample size

Djalil Chafai, Didier Concordet

## ► To cite this version:

Djalil Chafai, Didier Concordet. Confidence regions for the multinomial parameter with small sample size. 2008. hal-00278790v2

**HAL Id: hal-00278790**

**<https://hal.science/hal-00278790v2>**

Preprint submitted on 14 May 2008 (v2), last revised 11 Dec 2008 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Confidence regions for the multinomial parameter with small sample size

Djalil CHAFAÏ and Didier CONCORDET

Preprint – May 2008

## Abstract

Consider the observation of  $n$  iid realizations of an experiment with  $d \geq 2$  possible outcomes, which corresponds to a single observation of a multinomial distribution  $\mathcal{M}_d(n, p)$  where  $p$  is an unknown discrete distribution on  $\{1, \dots, d\}$ . In many applications in Biology, Medicine, Physics, and Engineering, the construction of a confidence region for  $p$  when  $n$  is small is crucial. This challenging concrete problem has a long history. It is well known that the confidence regions build from asymptotic statistics do not have good coverage for small  $n$ . In the binomial case ( $d = 2$ ), Clopper and Pearson provided a nice way to construct non-asymptotic confidence regions. We show how to generalize their approach to any  $d$ , by using the concept of covering collections. We also propose an attractive new alternative method which provides small confidence regions of controlled coverage. It corresponds to a special covering collection based on level sets of the multinomial distribution. We compare the performance of our new method to various other methods, including a method of Wald based on the Central Limit Theorem, a method based on concentration of measure and deviation probabilities, and a Bayesian method based on Dirichlet-Jeffrey priors.

**Keywords.** Confidence regions ; small samples ; multinomial distribution.

**MSC.** 62F25.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Various known methods</b>	<b>3</b>
2.1	Asymptotic regions based on the Central Limit Theorem . . . . .	3
2.2	Clopper and Pearson non-asymptotic regions for the binomial . . . . .	4
2.3	Non-asymptotic regions based on concentration of measure . . . . .	5
2.4	Bayesian regions based on Dirichlet-Jeffrey priors . . . . .	6
<b>3</b>	<b>Multivariate confidence regions via covering collections</b>	<b>8</b>
3.1	The Clopper and Pearson interval as a special binomial case . . . . .	9
3.2	Multinomial extension of the Clopper and Pearson interval . . . . .	10
<b>4</b>	<b>New alternative regions based on level sets</b>	<b>11</b>
4.1	Optimality . . . . .	12
<b>5</b>	<b>Simulation study</b>	<b>13</b>

## 1 Introduction

Consider the observation of  $n$  iid realizations  $Y_1, \dots, Y_n$  of an experiment with  $d \geq 2$  possible outcomes. In other words, the random variables  $Y_1, \dots, Y_n$  are iid with common discrete distribution  $p_1\delta_1 + \dots + p_d\delta_d$  on  $\{1, \dots, d\}$ . This corresponds to a single observation  $X = (X_1, \dots, X_d)$

of the multinomial distribution

$$\mathcal{M}_d(n, p) = \sum_{\substack{0 \leq k_1, \dots, k_d \leq n \\ k_1 + \dots + k_d = n}} p_1^{k_1} \dots p_d^{k_d} \frac{n!}{k_1! \dots k_d!} \delta_{(k_1, \dots, k_d)}.$$

where  $p = (p_1, \dots, p_d)$  and  $X_k = \text{Card}\{1 \leq i \leq n \text{ such that } Y_i = k\}$  for every  $1 \leq k \leq d$ . Here  $d$  is known,  $X$  is observed, and  $p$  is unknown. The present article deals with the problem of constructing a confidence region for  $p$  from the single observation  $X$  of  $\mathcal{M}_d(n, p)$ , in the *non-asymptotic* situation where  $n$  is *small*. The desired region must have *at least a coverage* of  $1 - \alpha$  for some prescribed level  $\alpha \in (0, 1)$ . More precisely, let

$$\Lambda_d = \{(u_1, \dots, u_d) \in [0, 1]^d \text{ such that } u_1 + \dots + u_d = 1\}$$

be the simplex of probability distributions on  $\{1, \dots, d\}$ . From the single observation  $X$  of  $\mathcal{M}_d(n, p)$  and for some prescribed level  $\alpha \in (0, 1)$ , we are interested in the construction of a random region  $R_\alpha(X) \subset \Lambda_d$  depending on  $X$  and  $\alpha$  such that

- the coverage probability has a prescribed lower bound:  $\mathbb{P}(p \in R_\alpha(X)) \geq 1 - \alpha$  ;
- the volume of  $R_\alpha(X)$  in  $\mathbb{R}^d$  is as small as possible.

In the literature, various other theoretical properties of confidence regions (e.g. equivariance and optimality) are considered, see for instance [8, 11]. However, the two properties mentioned above are the most important in practice. There is a rich literature on this old concrete statistical problem, and the binomial case ( $d = 2$ ) is much more understood than the general multinomial case ( $d \geq 2$ ). Most methods proposed in the literature for  $d > 2$  are asymptotic or Bayesian and do not ensure that the coverage probability is at least  $1 - \alpha$ , for a fixed value of  $n$ , see for instance [28, 27, 4, 15, 22]. The main problem here is that  $n$  is small. In particular, the confidence regions built from the asymptotic approaches based on the Central Limit Theorem (e.g. Wald methods) have a poor and uncontrolled coverage, even if  $n$  is large but finite. For the same reasons, it is also the case for the bootstrapped versions which only improve asymptotically the coverage probability (see [30, 22, 15, 17]). On the other hand, the discrete nature of the multinomial distribution produces a staircase effect which makes difficult the construction of non-asymptotic regions with coverage equal exactly to  $1 - \alpha$ . For a discussion of such aspects, we refer for instance to Agresti et al. [3, 2, 1]. In general, a reasonable expectation is to ask for a coverage of at least  $1 - \alpha$ , without being too conservative. Here the term conservative means that the coverage is greater than  $1 - \alpha$ .

We can summarize the situation as follows. Practically, there currently exist two kinds of methods for the construction of confidence region for  $p$ . On the first hand, methods that give confidence regions with small volume but that fail to control the prescribed coverage (e.g. Bayesian methods with Jeffrey prior, Wald Central Limit methods, Bootstrapped regions), and on the second hand, methods that control the prescribed coverage but have a too large volume to be useful (e.g. concentrations based methods, Clopper-Pearson type methods). We propose a new method which provides confidence regions with a strict control on the coverage while maintaining a volume comparable to the Wald Central Limit region. Namely, consider the discrete simplex

$$E_d = \{(x_1, \dots, x_d) \in \{0, \dots, n\}^d \text{ such that } x_1 + \dots + x_d = n\} \quad (1)$$

where lies the observation  $X \sim \mathcal{M}_d(n, p)$ . For all  $x \in E_d$  and  $p \in \Lambda_d$ , we define

$$\mu_p(x) = p_1^{x_1} \dots p_d^{x_d} \frac{n!}{x_1! \dots x_d!}.$$

For any prescribed  $\alpha \in (0, 1)$ , our confidence region  $R_\alpha(X) \subset \Lambda_d$  for  $p$ , based on level sets of the multinomial distribution, is defined by

$$R_\alpha(X) = \{p \in \Lambda_d \text{ such that } \mu_p(X) \geq u(p, \alpha)\} \quad (2)$$

where

$$u(p, \alpha) = \sup \left\{ u \in [0, 1] \text{ such that } \sum_{k \in A(p, u)} \mu_p(k) \geq 1 - \alpha \right\}$$

and  $A(p, u) = \{x \in E_d \text{ such that } \mu_p(x) \geq u\}$ . For this region, we have  $\mathbb{P}(p \in R_\alpha(X)) \geq 1 - \alpha$ , and computer simulations show that  $R_\alpha(X)$  has small volume and coverage close to  $1 - \alpha$ . Despite its apparent complex definition, such a region can be easily computed numerically. Surprisingly, this method turns out to be an excellent alternative to all the know methods so far.

## Outline of the rest of the article

In section 2 we recall various known methods used for the construction of confidence regions for  $p$  in the binomial ( $d = 2$ ) or general multinomial ( $d \geq 2$ ) case, including the well known Clopper and Pearson region for  $d = 2$ . In section 3, we provide an extension of the Clopper and Pearson method to the general multinomial case ( $d \geq 2$ ), by using the general concept of *covering collection*. In section 4, we propose a new alternative method (2) for the construction of confidence region in the general multinomial case, based on a special covering collection involving level sets of the multinomial distribution. Finally, in section 5, we present a numerical study where various confidence regions are compared. In particular, we compare in the “trinomial” case  $d = 3$ , the confidence regions produced with the Wald Central Limit method, with the uniform concentration of measure method, with the Dirichlet-Jeffrey prior method, with the extended Clopper and Pearson method, and with our new method (2) based on level sets.

## 2 Various known methods

For the special case of the binomial distribution ( $d = 2$ ), Newcombe [24] performed a comparison of several confidence intervals. These are either based on asymptotic expansions of the binomial distribution [6, 7, 10] or built using the exact distribution of  $X$  without assuming that  $n$  is large. Intervals built using asymptotic distributions have poor properties. In particular, Brown [9] showed that, even for large  $n$ , their coverage can be much smaller than the prescribed  $1 - \alpha$ .

### 2.1 Asymptotic regions based on the Central Limit Theorem

The most common frequentist way to construct confidence regions for the parameter  $p$  of a multinomial distribution is probably the Wald Central Limit Method. Let us recall briefly how works this asymptotic approach based on the maximum likelihood estimator  $\hat{p} = \frac{1}{n}X$  of  $p$ . By the multivariate Central Limit Theorem, when  $n$  is large, the random vector  $\hat{p}$  of  $\mathbb{R}^d$  is approximately distributed according to the multivariate Gauss distribution with mean  $p$  and covariance matrix  $\frac{1}{n}\Sigma(p) = \frac{1}{n}(\text{Diag}(p) - p \otimes p)$  given for every  $1 \leq i, j \leq d$  by

$$(\Sigma(p))_{i,j} = \frac{1}{n} \begin{cases} p_i(1 - p_i) & \text{if } i = j \\ -p_i p_j & \text{if } i \neq j. \end{cases}$$

Due to the constraint  $p_1 + \dots + p_d = 1$ , for every  $p \in \Lambda_d$ , the matrix  $\Sigma(p)$  is singular with rank smaller than or equal to  $d - 1$ . Also, one can consider the vector  $Z$  (respectively  $q$ ) which

contains the first  $(d-1)$  components of  $X$  (respectively  $p$ ). When  $\{q_1, \dots, q_{d-1}\} \cap \{0, 1\} = \emptyset$ , the  $(d-1) \times (d-1)$  variance matrix  $S(q)$  of  $Z$  obtained by taking the first  $(d-1)$  rows and columns of  $\Sigma(p)$  is a full rank matrix. Therefore, if  $S(q)^{-1/2}$  is the inverse of a matrix square root  $S(q)$  (e.g. via the Cholesky decomposition), then  $\sqrt{n}S(q)^{-1/2} \left(\frac{1}{n}Z - q\right)$  converges in distribution as  $n \rightarrow \infty$  to the Gaussian law  $\mathcal{N}(0, I_{d-1})$ . If,  $\chi_{d-1, \alpha}^2$  is the  $(1-\alpha)$  quantile of a  $\chi^2$  distribution with  $d-1$  degrees of freedom, then the set

$$\left\{ p \in \Lambda_d \text{ such that } \left\| \sqrt{n}S(p)^{-1/2} \left(\frac{1}{n}Z - p\right) \right\|^2 \leq \chi_{d-1, \alpha}^2 \right\} \quad (3)$$

is known as the Wald  $(1-\alpha)$  confidence region of  $q$  (or of  $p$  since  $p_d = 1 - q_1 - \dots - q_{d-1}$ ). Of course, if a component of  $q$  is equal to 0 or 1, then  $S(q)$  is a singular matrix and such a Wald type confidence region does not make any sense.

These Gaussian based confidence regions have poor coverage, even if  $n$  is large, due to the asymptotic nature of the Central Limit Theorem [9]. However, these confidence regions provide variance correction near the boundary of the simplex  $\Lambda_d$ .

Other approaches have been proposed. A natural one is to build confidence interval for each component of  $p$  and to deduce a confidence region as a product of these intervals. Proceeding that way allows to use all the existing methods for the binomial distribution. However, it is difficult to take into account the constraint that the components of  $p$  sum up to 1. Another difficulty in this approach is the control of the simultaneous coverage of the whole region by the coverage of each interval. Wang [29] gave a method to compute the confidence coefficient of five of these intervals. Hou [19] proposed power-divergence simultaneous confidence intervals that generalize the one proposed by Quesenberry and Hurst [25] and Goodman [16]. Lee & al. [21] reviewed twelve simultaneous confidence regions, that are not necessarily product of intervals, based on the likelihood ratio, the score statistic, the Pearson  $\chi^2$  statistic and various quadratic approximations to these. Unfortunately, at the practical level, all these approaches are questionable due to the non-controlled coverage of these regions, even when  $n$  is large [24, 9].

**Remark 2.1 (Sample size).** Suppose that  $d$  is well chosen, in the sense that  $0 < p_i < 1$  for every  $1 \leq i \leq d$ . It is quite natural to ask for the law of the random variable  $n_{crit}$  defined by

$$n_{crit} = \min\{n \geq 1 \text{ such that } \{Y_1, \dots, Y_n\} = \{1, \dots, d\}\}.$$

The random variable  $n_{crit}$  is the critical sample size for which we observe the whole  $d$  modalities, i.e. for which  $X$  belongs to the interior of the discrete simplex (1). This problem is well known in Probability Theory and Computer Science, and is often referred as the coupon collector problem. In the elementary case where  $p_1 = \dots = p_d = \frac{1}{d}$ , the distribution of  $n_{crit}$  can be computed explicitly, its mean is of order  $d \log(d)$  and its fluctuation around the mean follows a Gumbel type distribution. See for instance [14, 23, 18].

## 2.2 Clopper and Pearson non-asymptotic regions for the binomial

Consider the binomial case  $d = 2$ , for which  $p = (p_1, 1 - p_1)$ . The well known Clopper and Pearson interval for  $p_1$  relies on the exact distribution of  $X_1$  in the binomial case [13, 20, 12]. It was considered for a long time as outstanding. This interval  $[L, U]$  is given by

$$\begin{cases} L &= \inf \{ \theta \in [0, 1] \text{ such that } \sum_{i=x_1}^n \binom{n}{i} \theta^i (1-\theta)^{n-i} \geq \frac{1}{2} \alpha \} \\ U &= \sup \{ \theta \in [0, 1] \text{ such that } \sum_{i=0}^{x_1-1} \binom{n}{i} \theta^i (1-\theta)^{n-i} \geq \frac{1}{2} \alpha \}. \end{cases} \quad (4)$$

It has been shown that the Clopper and Pearson interval is often conservative. Also, some continuity corrections have been proposed, and give the so called “mid-p interval”, see for

instance [5] for a review. This trick reduces the staircase effect but the coverage probability can be less than  $1 - \alpha$ . The Beta-Binomial correspondence (see lemma 2.2 below) shows that the left and right limits  $L$  and  $R$  of the Clopper and Pearson confidence interval (4) are the  $\frac{1}{2}\alpha$  and  $(1 - \frac{1}{2}\alpha)$  quantiles of the Beta distribution  $\text{Beta}(X_1; n - X_1 + 1)$ .

**Lemma 2.2 (Beta-Binomial correspondence).** *If  $X \sim \text{Binom}(n, p_1)$  with  $p_1 \in [0, 1]$  and  $0 \leq k \leq n$  and  $B \sim \text{Beta}(k, n - k + 1)$  then following identity holds true.*

$$\mathbb{P}(X \geq k) = \mathbb{P}(B \leq p_1). \quad (5)$$

*Proof.* Let  $U_1, \dots, U_n$  be iid uniform random variables on  $[0, 1]$  and  $U_{(1)} \leq \dots \leq U_{(n)}$  be the reordered sequence. If we define  $V_{p_1} = \sum_{i=1}^n \mathbf{I}_{\{U_i \leq p_1\}}$  then  $V_{p_1} \sim \text{Bin}(n, p_1)$  and  $U_{(k)} \sim \text{Beta}(k, n - k + 1)$  and for every  $1 \leq k \leq n$ ,  $V_{p_1} \geq k$  if and only if  $U_{(k)} \leq p_1$ .  $\square$

To our knowledge, the Clopper and Pearson interval has no multinomial counterpart in the general multinomial case ( $d > 2$ ). We propose in section 3 an extension of the Clopper and Pearson method for the general multinomial case, by using the concept of covering collections.

### 2.3 Non-asymptotic regions based on concentration of measure

In the binomial case ( $d = 2$ ), we have  $(X_1, 1 - X_1) \sim \mathcal{M}_2(n, (p_1, 1 - p_1))$  and  $X_1 \sim \text{Binom}(n, p)$ . Many textbooks propose confidence intervals for  $p_1$  based on the Tchebychev inequality, which gives for any real  $r > 0$ , integer  $n > 0$ , and real  $p_1 \in [0, 1]$ ,

$$\mathbb{P}\left(\left|\frac{X_1}{n} - p_1\right| \leq r\right) \geq 1 - \frac{p_1(1 - p_1)}{nr^2} \geq 1 - \frac{1}{4nr^2}. \quad (6)$$

This gives the  $(1 - \alpha)$  confidence interval (for  $p_1$ )

$$\frac{X_1}{n} \pm \frac{1}{\sqrt{4n\alpha}}.$$

An alternative approach is to make use of the Hoeffding inequality which reads

$$\mathbb{P}\left(\left|\frac{X_1}{n} - p_1\right| \leq r\right) \geq 1 - 2e^{-2nr^2} \quad (7)$$

and which gives the  $(1 - \alpha)$  confidence interval (for  $p_1$ )

$$\frac{X_1}{n} \pm \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}.$$

The Tchebychev interval is based on the bound  $\sup_{p_1 \in [0, 1]} p_1(1 - p_1) \leq \frac{1}{4}$  on  $\text{Var}(X_1)$ , whereas the Hoeffding interval is based on the boundedness of the support. When  $n$  is small, these intervals are rather crude. A possible workaround is to use some sort of refined non-asymptotic concentration bounds for the binomial distribution, semi-Gaussian when  $p$  is near  $1/2$  and semi-Poissonian when  $p$  is close to 0 or 1 (this corresponds to two type of variance correction). Alternatively, an “optimal” concentration method consists in some sort of “uniform” bound on the deviation probability. Namely, let us define for every  $r > 0$ ,  $n \geq 1$ , and  $p_1 \in [0, 1]$ ,

$$C(n, r, p_1) = \mathbb{P}\left(\left|\frac{X_1}{n} - p_1\right| > r\right) \quad \text{and} \quad C(n, r) = \sup_{p_1 \in [0, 1]} C(n, p_1, r)$$

where  $X_1 \sim \text{Binom}(n, p_1)$ . If  $r_\alpha$  is such that  $\alpha = C(n, r_\alpha)$ , the interval

$$\frac{X_1}{n} \pm r_\alpha$$

has a coverage of at least  $(1 - \alpha)$  uniformly on  $p_1$ . The extension to the general multinomial case (i.e.  $d \geq 2$ ) is straightforward. Namely, we define for every  $r > 0$ ,  $n \geq 1$ , and  $p \in \Lambda_d$ ,

$$C(n, r, p) = \mathbb{P} \left( \left\| \frac{X}{n} - p \right\| > r \right) \quad \text{and} \quad C(n, r) = \sup_{p \in \Lambda_d} C(n, p, r).$$

where  $\|\cdot\|$  is a norm on  $\Lambda_d \subset \mathbb{R}^d$  (for instance an  $L^1$ ,  $L^2$ , or  $L^\infty$  norm), and where  $X \sim \mathcal{M}_d(n, p)$ . If  $r_\alpha$  is chosen in such a way that  $\alpha < C(n, r_\alpha)$ , the ball

$$\left\{ p \in \Lambda_d \text{ such that } \left\| \frac{X}{n} - p \right\| \leq r_\alpha \right\}$$

has a coverage of at least  $(1 - \alpha)$ . The confidence regions built by this way are centered on the maximum likelihood estimator, and their radius is deterministic and depends only on  $n$ ,  $d$ ,  $\alpha$ , and  $\|\cdot\|$ . It is thus possible to pre-compute the radius.

It is tempting to incorporate a (co)variance correction inside the deviation probability before the optimization as in the Wald Central Limit method. Namely, for  $d = 2$ , this corresponds to  $C(n, r) = \sup_{p_1 \in (0,1)} C(n, r, p_1)$  where  $C(n, r, p_1) = \mathbb{P}(|n^{-1}X_1 - p_1|(np_1(1 - p_1))^{-1/2} \geq r)$ . Unfortunately, when  $p$  tends to the boundary of the simplex  $\Lambda_2$ , the quantity  $C(n, r, p_1)$  tends to 1 and thus  $C(n, r)$  is equal to 1 except if  $X_1 = np_1$ . One may then replace  $(np_1(1 - p_1))^{1/2}$  by some function  $\sigma_{n,p_1}$  with a cutoff in a neighborhood of the boundary of  $[0, 1]$ . The corresponding confidence region for  $p_1$  is given by

$$\{p_1 \in (0, 1) \text{ such that } |n^{-1}X_1 - p_1|(\sigma_{n,p_1})^{-1} \leq r_\alpha\}$$

where  $r_\alpha$  is chosen for some prescribed  $\alpha \in (0, 1)$  in such a way that  $\alpha < C(n, r_\alpha)$ . These strange regions were not used in our simulation studies, due to the arbitrary choice of the variance correction  $\sigma_{n,p_1}$  and to the induced numerical difficulties.

## 2.4 Bayesian regions based on Dirichlet-Jeffrey priors

In contrast to the asymptotic method based on the Central Limit Theorem (e.g. Wald method), both the Clopper and Pearson method and the concentration of measure method produce confidence regions with a coverage of at least  $1 - \alpha$ , for arbitrary values of  $n$ . These methods must be compared to popular Bayesian methods. In a Bayesian framework, building a confidence set for the multinomial parameter by using a Dirichlet prior is a natural choice [26], and the Jeffrey approach based on the Fisher information matrix is a quite common way to choose the parameter of the Dirichlet prior. Following the Bayesian point of view, the parameter  $p$  is random and  $\mathcal{L}(X|p) = \mathcal{M}_d(n, p)$ . A straightforward computation of the Fisher information matrix of the multinomial shows that the Jeffrey prior on  $p$  is a Dirichlet distribution on  $\Lambda_d$  with parameter  $(\frac{1}{2}, \dots, \frac{1}{2})$ . The Dirichlet distribution is conjugate to the multinomial distribution in the sense that the posterior distribution  $\mathcal{L}(p|X)$  is a Dirichlet law with parameter  $(\frac{1}{2} + X_1, \dots, \frac{1}{2} + X_d)$ . One can obtain a confidence region for  $p$  with a coverage at least  $1 - \alpha$  by considering a set of mass  $1 - \alpha$  for the posterior distribution. However, this does not guarantee at all that the coverage of the confidence region is greater than or equal to  $1 - \alpha$ .

When  $d = 2$ , the Dirichlet distribution reduces to the Beta distribution. In this case, there is a clear link between the quantiles of the Beta distribution and the confidence region, as shown by

lemma 2.2. This shows also a clear link between the Dirichlet-Jeffrey approach and the Clopper and Pearson approach when  $d = 2$ . Actually, the Beta-Binomial identity (5) is a special case of a more general Dirichlet-Multinomial identity as shown by the following lemma.

**Lemma 2.3 (Dirichlet-Multinomial correspondence).** *Let  $p \in \Lambda_d$  and  $k_0, k_1, \dots, k_d$  such that  $k_0 = 0 \leq k_1 \leq \dots \leq k_{d-1} \leq n \leq k_d = n + 1$ . If*

$$X \sim \mathcal{M}_d(n, p) \quad \text{and} \quad D \sim \text{Dirichlet}_d(k_1 - k_0, k_2 - k_1, \dots, k_d - k_{d-1})$$

*then the following identity holds true.*

$$\begin{aligned} \mathbb{P}(X_1 \geq k_1, X_1 + X_2 \geq k_2, \dots, X_1 + \dots + X_{d-1} \geq k_{d-1}) \\ = \mathbb{P}(D_1 \leq p_1, D_1 + D_2 \leq p_2, \dots, D_1 + \dots + D_{d-1} \leq p_{d-1}). \end{aligned} \quad (8)$$

*Proof.* Let  $I_1, \dots, I_d$  be the sequence of adjacent sub-intervals of  $[0, 1]$  of respective lengths  $p_1, \dots, p_d$ ,  $U_1, \dots, U_n$  be iid uniform random variables on  $[0, 1]$  and  $U_{(1)} \leq \dots \leq U_{(n)}$  be the reordered sequence. For any  $1 \leq r \leq d$ , let us define

$$V_{p,r} = \sum_{i=1}^n \mathbf{I}_{\{U_i \in I_r\}} = \text{Card}\{1 \leq i \leq n \text{ such that } U_i \in I_r\}.$$

We have  $V_p = (V_{p,1}, \dots, V_{p,r}) \sim \mathcal{M}_d(n, p)$ . Now, for every  $0 \leq k_1 \leq \dots \leq k_{d-1} \leq n$ ,

$$V_{p,1} \geq k_1, \dots, V_{p,1} + \dots + V_{p,d-1} \geq k_{d-1} \quad \text{iff} \quad U_{(k_1)} \leq p_1, \dots, U_{(k_{d-1})} \leq p_1 + \dots + p_{d-1}.$$

But by using the notation  $U_{(0)} = 0$  and  $U_{(n+1)} = 1$ , we have

$$(U_{(1)} - U_{(0)}, \dots, U_{(n+1)} - U_{(n)}) \sim \text{Dirichlet}_{n+1}(1, \dots, 1).$$

and therefore, by the stability of Dirichlet laws by sum of blocs, with  $k_0 = 0$  and  $k_d = n + 1$ ,

$$(U_{(k_1)} - U_{(k_0)}, \dots, U_{(k_d)} - U_{(k_{d-1})}) \sim \text{Dirichlet}_d(k_1, k_2 - k_1, \dots, k_d - k_{d-1}).$$

□

This nice property allows to easily compute confidence regions that are products of intervals. Alternatively, a “level set” approach on the posterior distribution gives a confidence region with smoother boundary. The “density” of the Dirichlet distribution on  $\Lambda_d$  with parameter  $a = (a_1, \dots, a_d) \in \mathbb{R}_+^d$  is given for every  $u \in \Lambda_d$  by

$$f_a(u) = \frac{1}{B(a)} u_1^{a_1-1} \dots u_d^{a_d-1} \quad \text{where} \quad B(a) = \frac{\Gamma(a_1) \dots \Gamma(a_d)}{\Gamma(a_1 + \dots + a_d)}.$$

For every real  $\ell > 0$ , one may consider the level set

$$A(a, \ell) = \{p \in \Lambda_d \text{ such that } f_a(p) \geq \ell\}$$

and the conditional critical level

$$u_\alpha(X) = \sup \left\{ \ell > 0 \text{ such that } \mathbb{P} \left( p \in A(X + \frac{1}{2}, \ell) \middle| X \right) \geq 1 - \alpha \right\}.$$

One can then defined the  $1 - \alpha$  Dirichlet-Jeffrey confidence region as  $A(X + \frac{1}{2}, u_\alpha(X))$ . Conditional on  $X$ , this confidence region has a coverage of at least  $1 - \alpha$ . However, nothing guarantees that the unconditional confidence region keeps this property.



### 3 Multivariate confidence regions via covering collections

The aim of this section is to introduce the notion of *covering collection*, which will allow us in particular to extend the Clopper and Pearson method to the general multinomial case ( $d \geq 2$ ). Covering collections allow to build confidence regions in a general abstract space. Let us consider a random variable  $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{B}_E)$  having a distribution  $\mu_{\theta^*}$  where  $\theta^* \in \Theta$ . For some  $\alpha \in (0, 1)$ , we would like to construct a confidence region  $R_\alpha(X)$  for  $\theta^*$  with a coverage of at least  $(1 - \alpha)$ , from a single realization of  $X$ . In other words,

$$\mathbb{P}(\theta^* \in R_\alpha(X)) \geq 1 - \alpha. \quad (9)$$

**Definition 3.1 (Covering collection).** A covering collection of  $E$  is a collection of measurable events  $(A_k)_{k \in \mathcal{K}} \subset \mathcal{B}_E$  such that

- $\mathcal{K}$  is totally ordered and admits a minimal element and a maximal element;
- if  $k \leq k'$  then  $A_k \subset A_{k'}$  with equality if and only if  $k = k'$ ;
- $A_{\min(\mathcal{K})} = \emptyset$  and  $A_{\max(\mathcal{K})} = E$ .

For instance, for  $E = \{0, 1, \dots, n\}$ , the sequence of sets

$$\emptyset, \{\sigma(0)\}, \{\sigma(0), \sigma(1)\}, \dots, \{\sigma(0), \sigma(1), \dots, \sigma(n)\} = E$$

is a covering collection of  $E$  for any permutation  $\sigma$  of  $E$ . For  $E = \mathbb{R}$ , the collection  $(A_t)_{t \in \overline{\mathbb{R}}}$  where  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$  defined by  $A_{-\infty} = \emptyset$ ,  $A_t = (-\infty, t]$  for every  $t \in \mathbb{R}$ , and  $A_{+\infty} = \mathbb{R}$  is a covering collection of  $E$ . Many other choices are possible, like  $A_t = [-t, +t]$  or  $A_t = [t, +\infty)$ . We recognize the usual shapes of the confidence regions used in univariate Statistics.

**Theorem 3.2 (Confidence region associated with a covering collection).** Let  $(A_k)_{k \in \mathcal{K}}$  be a covering collection of  $E$ , and  $k_X$  be the smallest  $k \in \mathcal{K}$  such that  $X \in A_k$ . For every  $\alpha \in (0, 1)$ , the region  $R_\alpha(X)$  defined below satisfies to (9).

$$R_\alpha(X) = \{\theta \in \Theta \text{ such that } \mu_\theta(A_{k_X}) \geq \alpha\}. \quad (10)$$

*Proof.* For every  $\theta \in \Theta$ , let  $k_\alpha(\theta)$  be the largest  $k \in \mathcal{K}$  such that  $\mu_\theta(A_k) < \alpha$ . With this definition of  $k_\alpha(\cdot)$ , we have then

$$x \in A_{k_\alpha(\theta)} \text{ if and only if } \mu_\theta(A_{k_x}) < \alpha.$$

Thus we have

$$\begin{aligned} \mathbb{P}(\theta^* \in R_\alpha(X)) &= \mathbb{P}(\mu_{\theta^*}(A_{k_X}) \geq \alpha) \\ &= \mathbb{P}(X \notin A_{k_\alpha(\theta^*)}) \\ &= 1 - \mu_{\theta^*}(A_{k_\alpha(\theta^*)}) \\ &\geq 1 - \alpha. \end{aligned}$$

□

These confidence regions highly depend on the chosen covering collection  $(A_k)_{k \in \mathcal{K}}$ . Each choice of covering collection gives a particular region  $R_\alpha(X)$ . One can notice that a small value of  $k_X$  gives a small set  $A_{k_X}$  and thus leads to a confidence region with a small volume. For instance, assume that we have two realizations  $x_1$  and  $x_2$  of  $X$  with  $k_{x_1} < k_{x_2}$ . For a given

sequence  $(A_k)_{k \in \mathcal{K}}$ , we have  $A_{k_{x_1}} \subset A_{k_{x_2}}$  and thus  $R_\alpha(x_1) \subset R_\alpha(x_2)$ . One could be tempted to choose the covering collection  $(A_k)_{k \in \mathcal{K}}$  in such a way that  $k_X$  is as small as possible. In such a case, the covering collection  $(A_k)_{k \in \mathcal{K}}$  would become random and the coverage of the associated region could be less than the prescribed level  $1 - \alpha$ .

Notice that the set  $A_{k_X}$  can be empty, which means that a confidence region cannot be built with such a sequence  $(A_k)_{k \in \mathcal{K}}$ . In contrast, the case where  $A_{k_X} = E$  leads to the trivial region  $R_\alpha(X) = \Theta$ . In the case where  $A_{k_X} = \{X\}$ , we have  $\mu_\theta(A_{k_X}) = \mu_\theta(\{X\})$ , which is the likelihood of  $X$  at point  $\theta$ , and the region  $R_\alpha(X)$  corresponds to the complement of a level set of the likelihood.

**Remark 3.3 (Discrete case and staircase effect).** *Let  $(A_k)_{k \in \mathcal{K}}$  be a covering collection of a finite set  $E$ . Due to staircase effects, the coverage of the confidence regions constructed from this covering collection cannot take arbitrary values in  $(0, 1)$ . These staircase effects can be reduced by using a fully granular collection for which  $\text{Card}(\mathcal{K}) = \text{Card}(E)$ . The term fully granular means that the elements of the collection are obtained by adding the points of  $E$  one by one. It is impossible to remove completely the staircase effects when  $E$  is discrete, while maintaining a nominal lower bound on the coverage.*

**Remark 3.4 (Reverse regions).** *For the region  $R_\alpha(X) = \{\theta \in \Theta; \mu_\theta(A_{k_X}) \leq 1 - \alpha\}$  we have*

$$\mathbb{P}(R_\alpha) = \mathbb{P}(\mu_\theta(A_{k_X}) \leq 1 - \alpha) = \mathbb{P}(X \in A_{k_{1-\alpha}}) = \mu_\theta(A_{k_{1-\alpha}}) \leq 1 - \alpha.$$

**Remark 3.5 (Symmetrization).** *If  $R_1$  and  $R_2$  are two confidence regions with a coverage of at least  $1 - \frac{1}{2}\alpha$  such that  $R_1 \cup R_2 = E$  then  $R_1^c$  and  $R_2^c$  are disjoint and thus  $R_1 \cap R_2 = (R_1^c \cup R_2^c)^c$  is a confidence region with a coverage of at least  $1 - \alpha$ .*

Consider a covering collection  $(A_k)_{0 \leq k \leq \kappa}$  of  $E$ . Now, for any  $0 \leq k \leq \kappa$ , let us define  $A'_k = E \setminus A_{\kappa-k}$ . For any  $\theta \in \Theta$ , any  $X \sim \mu_\theta$ , and any  $\alpha \in (0, 1)$ , we construct

$$R_{\frac{1}{2}\alpha} = \left\{ \theta \in \Theta; \mu_\theta(A_{k_X}) > \frac{1}{2}\alpha \right\} \quad \text{and} \quad R'_{\frac{1}{2}\alpha} = \left\{ \theta \in \Theta; \mu_\theta(A'_{k'_X}) > \frac{1}{2}\alpha \right\}$$

where  $k'_X$  is built from  $(A'_k)_{0 \leq k \leq \kappa}$  as  $k_X$  from  $(A_k)_{0 \leq k \leq \kappa}$  and  $A'_{k'_X} = E \setminus A_{k_X-1}$ . Since

$$\mu_\theta(A_{k_X}) + \mu_\theta(A'_{k'_X}) = 1 + \mu_\theta(\{X\}) \geq 1,$$

the regions  $R_{\frac{1}{2}\alpha}$  and  $R'_{\frac{1}{2}\alpha}$  have disjoint complements. Therefore, by Remark 3.5, the set

$$R_{\frac{1}{2}\alpha} \cap R'_{\frac{1}{2}\alpha}$$

is a confidence region with coverage greater than or equal to  $1 - \alpha$ .

### 3.1 The Clopper and Pearson interval as a special binomial case

Let us show why the Clopper and Pearson confidence interval can be considered as a special case of the method based on covering collections. Recall that we are in the case where  $d = 2$  and  $X_1 \sim \text{Binom}(n, p_1)$  for some unknown  $p_1 \in [0, 1]$ . Equivalently, we can write

$$(X_1, n - X_1) \sim \mathcal{M}_2(n, (p_1, 1 - p_1)).$$

The unidimensional nature of  $E = \{0, \dots, n\}$  suggests the following two covering collections  $(A_k^1)_{k \in E}$  and  $(A_k^2)_{k \in E}$  defined by  $A_0^1 = \emptyset$  and  $A_0^2 = \emptyset$ , and for every  $0 \leq k \leq n$ ,

$$A_{k+1}^1 = \{0, \dots, k\} \quad \text{and} \quad A_{k+1}^2 = \{n - k, \dots, n\}.$$

Here  $\mathcal{K} = E$  for both the top-to-bottom and bottom-to-top sequences. The bottom-to-top sequence  $(A_k^1)_{k \in E}$  leads to the  $(1 - \alpha)$  one-sided confidence interval for  $p_1$  given by

$$R_\alpha^1(X_1) = \left\{ \theta \in [0, 1] \text{ such that } \sum_{i=0}^{X_1} \binom{n}{i} \theta^i (1 - \theta)^{n-i} \geq \alpha \right\} = [0, U_\alpha(X_1)] \quad (11)$$

where

$$U_\alpha(x) = \sup \left\{ \theta \in [0, 1] \text{ such that } \sum_{i=0}^x \binom{n}{i} \theta^i (1 - \theta)^{n-i} \geq \alpha \right\}.$$

On the other hand, the top-to-bottom covering collection  $(A_k^2)_{k \in E}$  leads to an  $(1 - \alpha)$  confidence interval of  $p_1$  given by

$$R_\alpha^2(X_1) = \left\{ \theta \in [0, 1] \text{ such that } \sum_{i=X_1}^n \binom{n}{i} \theta^i (1 - \theta)^{n-i} \geq \alpha \right\} = [L_\alpha(X_1); 1], \quad (12)$$

where

$$L_\alpha(x) = \sup \left\{ \theta \in [0, 1] \text{ such that } \sum_{i=x}^n \binom{n}{i} \theta^i (1 - \theta)^{n-i} \geq \alpha \right\}.$$

By virtue of Remark 3.5, we can combine these two confidence intervals in order to obtain a symmetrized  $(1 - \alpha)$  confidence interval of  $p_1$ , which is the two-sided interval

$$R_{\frac{1}{2}\alpha}^1(X_1) \cap R_{\frac{1}{2}\alpha}^2(X_1) = [L_{\frac{1}{2}\alpha}(X_1); U_{\frac{1}{2}\alpha}(X_1)].$$

We recognize the Clopper-Pearson interval (4). The discrete nature of  $E$  precludes the construction of a confidence interval of  $p_1$  with coverage exactly equal to  $1 - \alpha$ . Actually, the Clopper-Pearson two-sided interval is not exactly symmetric and there is no guaranty that

$$\mathbb{P}(p < L_{\frac{1}{2}\alpha}(X_1)) = \mathbb{P}(p > U_{\frac{1}{2}\alpha}(X_1)).$$

### 3.2 Multinomial extension of the Clopper and Pearson interval

Consider the multinomial case where  $X \sim \mathcal{M}_d(n, p)$  with  $p \in \Lambda_d$  and  $d \geq 2$ . The set  $E_d$  defined by (1) appears as a discrete simplex, and we have

$$\text{Card}(E_d) = \binom{n+d-1}{d-1} = \frac{(n+d-1)!}{(d-1)!n!}.$$

For the binomial case  $d = 2$ , the two-sided Clopper and Pearson confidence interval appears naturally. For the multinomial case with  $d > 2$ , one can propose a region built by mimicking the confidence interval of the binomial case. This leads naturally to the notion of equivariance.

The choice of the covering collection  $(A_k)_{k \in \mathcal{K}}$  is quite arbitrary. However, some additional constraints can help to reduce this choice. As advocated by Casella [11] for the binomial distribution, the proposed confidence region  $R_\alpha(X)$  should be *equivariant*, that is not sensitive to the order chosen to label the  $d$  categories of the multinomial distribution.

**Definition 3.6 (Equivariance).** A confidence region  $R_\alpha(X)$  is equivariant when

$$\mathbb{P}(\sigma(\theta^*) \in R_\alpha(\sigma(X))) = \mathbb{P}(\theta^* \in R_\alpha(X)) \quad (13)$$

for every permutation  $\sigma$  of  $\{1, \dots, d\}$ . In other words, if and only if

$$\sigma(R_\alpha(X)) = R_\alpha(\sigma(X)).$$

The following lemma gives a criterion of equivariance for covering collections.

**Theorem 3.7 (Equivariance criterion for covering collections).** *The confidence region  $R_\alpha(X)$  constructed from a covering collection  $(A_k)_{k \in \mathcal{K}}$  is equivariant if and only if  $A_k$  is invariant by permutation of coordinates for every  $k \in \mathcal{K}$ .*

*Proof.* Let  $\sigma$  be a permutation of  $\{1, \dots, d\}$ ,  $i = (i_1, \dots, i_d) \in E$ , and for every  $\theta \in \Theta$ ,

$$\sigma(\theta) = (\theta_{\sigma(1)}, \dots, \theta_{\sigma(d)}) \quad \text{and} \quad \sigma(i) = (i_{\sigma(1)}, \dots, i_{\sigma(d)}).$$

By invariance of  $A_k$  by permutation, we have  $X \in A_k \Leftrightarrow X \in \sigma(A_k)$  and thus  $k_X = k_{\sigma(X)}$ . If  $\theta \in \sigma(R_\alpha(X))$  then  $\mu_{\sigma^{-1}(\theta)}(A_{k_X}) \geq \alpha$ . But, for every  $i \in E$ ,

$$\mu_{\sigma^{-1}(\theta)}(\{i\}) = \mu_\theta(\{\sigma(i)\}).$$

If  $A_k$  is invariant permutations, then for every  $i \in A_k$ , we have  $\sigma(i) \in A_k$  and consequently

$$\mu_{\sigma^{-1}(\theta)}(A_k) = \mu_\theta(\sigma(A_k)) = \mu_\theta(A_k).$$

Thus,  $\theta \in \sigma(R_\alpha(X))$  if and only if  $\mu_\theta(A_{k_X}) = \mu_\theta(A_{k_{\sigma(X)}}) \geq \alpha$ , that is  $\theta \in R_\alpha(\sigma(X))$ .  $\square$

As already noticed, a large set  $A_{k_X}$  gives a large confidence region. Since confidence regions with small volume are desirable, it is interesting, when  $E$  is discrete, to consider a covering collection  $(A_k)_{k \in \mathcal{K}}$  which grows by adding the points of  $E$  one after the other. Unfortunately, this method of construction is not compatible with equivariance: the  $A_k$  cannot be invariant by permutations of coordinates. A lighter condition could be that there exists a subsequence  $(A_{k_l})_l$  that is invariant by permutation of coordinates. An example of such a sequence for  $d = 3$  is represented in Figure 1. Equivariance is a strong constraint on the covering collection.

## 4 New alternative regions based on level sets

In this section, we propose a new method that fully uses the concept of coverage collection. We chose to consider decreasing coverage collection, to ease its presentation. The corresponding confidence regions are not exactly build as in Theorem 3.2. Let us consider a random variable  $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{B}_E)$  with law  $\mu_{\theta^*}$  where  $\theta^* \in \Theta$ . For every  $u \geq 0$  and  $\theta \in \Theta$ , let us define

$$A(\theta, u) = \{x \in E \text{ such that } \mu_\theta(x) \geq u\}.$$

For every  $\theta \in \Theta$ , the collection  $(A(\theta, u))_{u \geq 0}$  is decreasing with  $A(\theta, 0) = E$  and there exists  $u_{\max}$  that can be equal to  $+\infty$  such that  $A(\theta, u_{\max}) = \emptyset$ . Also,  $(A(\theta, u_{\max} - u))_{u \in [0, u_{\max}]}$  is a covering collection of  $E$ . Next, define

$$u(\theta, \alpha) = \sup \{u \in [0, u_{\max}] \text{ such that } \mu_\theta(A(\theta, u)) \geq 1 - \alpha\}$$

and

$$K(\theta, \alpha) = A(\theta, u(\theta, \alpha)).$$

We would like to construct a confidence region for  $\theta^*$  from the observation of  $X \sim \mu_{\theta^*}$ . If

$$R_\alpha(X) = \{\theta \in \Theta \text{ such that } X \in K(\theta, \alpha)\} \tag{14}$$

then

$$\mathbb{P}(\theta^* \in R_\alpha(X)) = \mathbb{P}(X \in K(\theta^*, \alpha)) = \mu_{\theta^*}(K(\theta^*, \alpha)) \geq 1 - \alpha.$$

This shows that  $R_\alpha(X)$  is a confidence region for  $\theta^*$  with a coverage of at least  $1 - \alpha$ .

Let us make precise the expression of the confidence region for the general multinomial case where  $X \sim \mathcal{M}_d(n, p)$  with  $p \in \Lambda_d$  and  $d \geq 2$ . Here the value of  $p$  used for the observed data  $X$  plays the role of  $\theta^*$ . We have  $\Theta = \Lambda_d$ ,  $E = E_d$  as described by (1),  $\mu_\theta = \mathcal{M}_d(n, \theta)$ , and  $u_{\max} = 1$ . For every  $\alpha \in (0, 1)$ , the confidence region given by the level sets method writes as in (2) given in the introduction.

**Remark 4.1 (Binomial case).** *It turns out that the confidence interval obtained by the level sets method for the binomial case ( $d = 2$ ) does not coincide with the Clopper and Pearson confidence interval. This is due to the special covering collection that is used in the level sets method.*

## 4.1 Optimality

Let us focus on the case where  $E$  is a finite set. The confidence region constructed above is not optimal among all the  $1 - \alpha$  conservative sets and thus could be improved by a more detailed analysis. Let us first notice that by its very construction, for all  $\theta \in \Theta$ ,  $K(\theta, \alpha)$  is minimal with respect to its cardinality that is, there does not exist a set  $B(\theta, \alpha)$  so that  $\mu_\theta(B(\theta, \alpha)) \geq 1 - \alpha$  and  $\text{card}(B(\theta, \alpha)) < \text{card}(K(\theta, \alpha))$ . However, in some circumstances, it may exist sets  $L(\theta, \alpha)$  with the same cardinality as  $K(\theta, \alpha)$  so that  $\mu_\theta(K(\theta, \alpha)) \geq \mu_\theta(L(\theta, \alpha)) \geq 1 - \alpha$ . The following theorem gives a condition that allows to build conservative sets but with a coverage closer to  $1 - \alpha$  than the coverage of  $R_\alpha(X)$ . For all  $\alpha \in [0, 1]$  and  $\theta \in \Theta$ , let us denote  $\gamma(\theta, \alpha) = 1 - \mu_\theta(K(\theta, \alpha))$  and let us notice that  $\gamma(\theta, \alpha) \leq \alpha$ .

**Theorem 4.2.** *For each  $\theta \in \Theta$ , assume that it exist two subsets  $V(\theta, \alpha) \subset K(\theta, \alpha)$  and  $W(\theta, \alpha) \subset E \setminus K(\theta, \alpha)$  with the same cardinality so that*

$$\alpha - \gamma(\theta, \alpha) \geq \mu_\theta(V(\theta, \alpha)) - \mu_\theta(W(\theta, \alpha)) > 0.$$

*Then, there exists a set  $T_\alpha(X) \neq R_\alpha(X)$  so that*

$$1 - \alpha \leq \mathbb{P}(\theta^* \in T_\alpha(X)) < \mathbb{P}(\theta^* \in R_\alpha(X)).$$

*Proof.* Let us consider the set  $L(\theta, \alpha) = K(\theta, \alpha) \setminus V(\theta, \alpha) \cup W(\theta, \alpha)$  and notice that thanks to the conditions imposed the sets  $V$  and  $W$  we have for all  $\theta \in \Theta$ ,

$$1 - \alpha \leq \mu_\theta(L(\theta, \alpha)) < \mu_\theta(K(\theta, \alpha)).$$

Now, set

$$T_\alpha(X) = \{\theta \in \Theta; X \in L(\theta, \alpha)\}.$$

But,

$$\begin{aligned} \mathbb{P}(\theta^* \in T_\alpha(X)) &= \mathbb{P}(X \in L(\theta^*, \alpha)) \\ &= \mathbb{P}(X \in K(\theta^*, \alpha) \setminus V(\theta^*, \alpha) \cup W(\theta^*, \alpha)) \\ &= 1 - \gamma(\theta^*, \alpha) - \mu_{\theta^*}(V(\theta^*, \alpha)) + \mu_{\theta^*}(W(\theta^*, \alpha)) \\ &\leq 1 - \gamma(\theta^*, \alpha). \end{aligned}$$

On the other hand, we have already seen that for all  $\theta \in \Theta$ ,

$$1 - \alpha \leq \mu_\theta(L(\theta, \alpha)).$$

This last inequality holds true when  $\theta = \theta^*$  and thus

$$1 - \alpha \leq \mu_{\theta^*}(L(\theta^*, \alpha)) = \mathbb{P}(\theta^* \in T_\alpha(X)).$$

□

This theorem can be used to build less conservative confidence sets than  $R_\alpha(X)$ . A convenient way to proceed is to take  $V(\theta, \alpha) = \{y\}$  where  $y$  is such that

$$\mu_\theta(y) = \min_{z \in K(\theta, \alpha)} \mu_\theta(z)$$

and to iteratively try several sets  $W^k$  as follows. Set  $W^0(\theta, \alpha) = \emptyset$ , and at iteration  $k \geq 1$ , set  $W^k(\theta, \alpha) = \{w_k\}$  and  $L^k(\theta, \alpha) = K(\theta, \alpha) \setminus V(\theta, \alpha) \cup W^k(\theta, \alpha)$  where

$$w_k = \arg \max_{z \in L^{k-1}(\theta, \alpha)} \mu_\theta(z).$$

This process is iterated until the set  $L^k(\theta, \alpha)$  is such that  $\mu_\theta(L^k(\theta, \alpha)) - (1 - \alpha)$  is non-negative and minimum.

Since for  $\theta \in \Theta$  there may exist  $x \neq y$  with  $\mu_\theta(x) = \mu_\theta(y)$ , there also may exist several sets  $(L^i(\theta, \alpha))_i$  which have the same mass  $\mu_\theta(L^i(\theta, \alpha)) = 1 - \delta(\theta, \alpha)$ . Several confidence sets with the same coverage can thus be derived using these sets. A simple way to choose between these concurrent confidence sets is to adopt the one that optimizes a criterion such as having a minimum volume (for the Lebesgue measure).

## 5 Simulation study

We compared our method based on level sets to existing methods for the binomial distribution. For each method described in the previous section, we calculated the 95% (i.e.  $\alpha = 0.05$ ) confidence intervals of  $p$  for each observed value  $x$  of a Binom(10,  $p$ ) and Binom(20,  $p$ ). The obtained intervals are respectively given in Tables 1 and 2.

These tables show that the intervals given by the Tchebychev inequality, by the Hoeffding inequality, and by the uniform concentration bound are too wide to be useful when  $n$  is small. On the contrary, the intervals given by the Bayesian method with Jeffrey prior are narrow. The length of the intervals given by the uniform concentration, the Wald Central Limit and the proposed level-sets method is in-between. The main advantage of the Clopper and Pearson interval is that it is only based on the sample distribution of  $X$ . However, it is sometimes not used in practice because it is usually larger than asymptotic based intervals. As shown in [9] for the binomial distribution, there are “lucky” values of  $p$  and  $n$  for which the interval obtained with the Wald Central Limit method has a coverage close to its nominal value. The same author has shown that for “unlucky”  $n$  and  $p$ , both the Bayesian interval and the Wald interval have coverage that can be much smaller than the prescribed coverage (0.95). Figure 2 shows that the coverage of our level-sets method is, as expected, always greater than or equal to 0.95. When  $n$  increases, the set of  $p$  for which the actual coverage is close to  $1 - \alpha$  increases. On the whole, these simulations show that our level-sets method gives the narrowest interval among the methods that have controlled coverage. On the other hand, our interval is wider than the Wald Central Limit and Bayesian intervals, two intervals for which the coverage is not controlled.

Recall that our objective is the construction of confidence regions for any  $d \geq 2$ . In order to evaluate the performances of various available methods, we built the confidence regions in the case where  $n = 10$  and  $d = 3$ . The methods that we used are Wald Central Limit, uniform concentration, Bayesian Jeffrey, extended Clopper-Pearson, and our level-sets. Since Tchebychev and Hoeffding regions lead to rough intervals for the binomial case  $d = 2$  and are outperformed by the uniform concentration method, we decided to ignore them for  $d > 2$ . Figures 3 and 4 respectively show the confidence sets obtained for the observations  $x = (3, 2, 5)$  and  $x = (0, 2, 8)$ . In the case  $x = (0, 2, 8)$  which lies in the boundary of  $E_3$ , the Wald and uniform methods proposed solutions outside  $\Theta$  (these strange solutions were discarded). We can see that the

Clopper-Pearson method with the set  $A_k$  described by Figure 1 gives confidence sets that are not useful in practice because they are too large. The uniform method is very convenient to work with. With the sample volume and the desired confidence level, it is easy to build an  $\|\cdot\|_2$ -ball centered on  $\hat{p}$  with a constant radius. Unfortunately, this advantage is balanced by a large area. On the whole, the Wald Central Limit method and our level-sets method give comparable regions. Table 3 gives the area of the obtained 95% confidence regions. When  $x$  belongs to the boundary of  $E_3$ , the Bayesian Jeffrey region has a smaller area than the one obtained with the Wald and level-sets methods. However, it has a larger area when  $x$  lies in the interior of  $E$ .

The area of a confidence region is certainly an important property from a practical point of view, but having a small area is only useful if the actual coverage is close to the prescribed one  $(1-\alpha)$ . Thus, we computed the coverage of the regions given by each of the previous methods for  $n = 10$ . The results are shown in Figure 5. As expected, the uniform method is too conservative. For the Wald region, the same phenomena as the one observed for the binomial case occurs for the trinomial case: the actual coverage can be quite far from the nominal confidence level but there exist “lucky” values of  $p$  for which the coverage is close to the nominal value. The Jeffrey prior distribution puts some mass near the boundaries of  $\Theta$  and  $n = 10$  is probably not large enough to counterbalance this prior. This probably explains why the Jeffrey method has a low coverage when  $p$  is close to the center of  $\Lambda_3$ . We have to mention that in the present simulation study, we computed the Jeffrey confidence region by using a Monte-Carlo approach, which is computer intensive. Our level-sets method gives regions with coverage very close to the nominal level, while being always greater than or equal to the nominal level by construction. Surprisingly, we did not observed for  $d = 3$  the same amplitude of the staircase effect as we did for  $d = 2$ .

## References

- [1] A. Agresti. Dealing with discreteness: making ‘exact’ confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Stat. Methods Med. Res.*, 12(1):3–21, 2003.
- [2] A. Agresti and B. Caffo. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *Amer. Statist.*, 54(4):280–288, 2000.
- [3] A. Agresti and B. A. Coull. Approximate is better than “exact” for interval estimation of binomial proportions. *Amer. Statist.*, 52(2):119–126, 1998.
- [4] J. Albert. Pseudo-Bayes estimation of multinomial proportions. *Comm. Statist. A—Theory Methods*, 10(16):1587–1611, 1981.
- [5] G. Berry and P. Armitage. Mid-p confidence intervals: a brief review. *The Statistician*, 44(4):417–423, 1995.
- [6] C. R. Blyth. Approximate binomial confidence limits. *J. Amer. Statist. Assoc.*, 81(395):843–855, 1986.
- [7] C. R. Blyth and H. A. Still. Binomial confidence intervals. *J. Amer. Statist. Assoc.*, 78(381):108–116, 1983.
- [8] A. A. Borovkov. *Mathematical statistics*. Gordon and Breach Science Publishers, Amsterdam, 1998. Translated from the Russian by A. Moullagaliev and revised by the author.

- [9] L. D. Brown, T. T. Cai, and A. DasGupta. Interval estimation for a binomial proportion. *Statist. Sci.*, 16(2):101–133, 2001. With comments and a rejoinder by the authors.
- [10] L. D. Brown, T. T. Cai, and A. DasGupta. Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.*, 30(1):160–201, 2002.
- [11] G. Casella. Refining binomial confidence intervals. *Canad. J. Statist.*, 14(2):113–129, 1986.
- [12] X. Chen, K. Zhou, and J. L. Aravena. On the binomial confidence interval and probabilistic robust control. *Automatica J. IFAC*, 40(10):1787–1789 (2005), 2004.
- [13] C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26:404–413, 1934.
- [14] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968.
- [15] J. Glaz and C. P. Sison. Simultaneous confidence intervals for multinomial proportions. *J. Statist. Plann. Inference*, 82(1-2):251–262, 1999. Multiple comparisons (Tel Aviv, 1996).
- [16] L. A. Goodman. On simultaneous confidence intervals for multinomial proportions. *Technometrics*, 7:247–254, 1965.
- [17] P. Hall. *The bootstrap and Edgeworth expansion.* Springer Series in Statistics. Springer-Verlag, New York, 1992.
- [18] L. Holst. On birthday, collectors’, occupancy and other classical urn problems. *Internat. Statist. Rev.*, 54(1):15–27, 1986.
- [19] C. D. Hou, J. Chiang, and J. J. Tai. A family of simultaneous confidence intervals for multinomial proportions. *Comput. Statist. Data Anal.*, 43:29–45, 2003.
- [20] S. A. Julious. Two-sided confidence intervals for the single proportion: comparison of seven methods, letter to the editor. *Statist. Med.*, 24:3383–3384, 2005.
- [21] A. J. Lee, S. O. Nyangomab, and G. A. F. Seber. Confidence regions for multinomial parameters. *Comput. Statist. Data Anal.*, 39:329–342, 2002.
- [22] D. Morales, L. Pardo, and L. Santamaría. Bootstrap confidence regions in multinomial sampling. *Appl. Math. Comput.*, 155(2):295–315, 2004.
- [23] R. Motwani and P. Raghavan. *Randomized algorithms.* Cambridge University Press, Cambridge, 1995.
- [24] R. G. Newcombe. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statist. Med.*, 17:857–872, 1998.
- [25] C. P. Quesenberry and D. C. Hurst. Simultaneous confidence intervals for multinomial proportions. *Technometrics*, 6:191–195, 1964.
- [26] C. P. Robert. *The Bayesian choice.* Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2001. From decision-theoretic foundations to computational implementation, Translated and revised from the French original by the author.
- [27] C. P. Sison and J. Glaz. Simultaneous confidence intervals and sample size determination for multinomial proportions. *J. Amer. Statist. Assoc.*, 90(429):366–369, 1995.



- [28] A. Ullah, A. T. K. Wan, and A. Chaturvedi, editors. *Handbook of applied econometrics and statistical inference*, volume 165 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York, 2002.
- [29] H. Wang. Exact confidence coefficients of simultaneous confidence intervals for multinomial proportions. *J. Multivariate Anal.*, To appear, 2007.
- [30] X.-D. Zheng and W.-Y. Loh. Bootstrapping binomial confidence intervals. *J. Statist. Plann. Inference*, 43(3):355–380, 1995.

DJALIL CHAFAÏ, CORRESPONDING AUTHOR, [d.chafai@envt.fr](mailto:d.chafai@envt.fr)

UMR181 INRA, ENVT, ÉCOLE NATIONALE VÉTÉRINAIRE DE TOULOUSE  
23 CHEMIN DES CAPELLES, F-31076 CEDEX 3, TOULOUSE, FRANCE.

UMR 5219 CNRS, INSTITUT DE MATHÉMATIQUES, UNIVERSITÉ DE TOULOUSE  
118 ROUTE DE NARBONNE, F-31062 CEDEX 4, TOULOUSE, FRANCE.

$x_1$	Tchebychev	Hoeffding	Uniform	Level	Clopper	Wald	Jeffrey
0	[0, 0.707]	[0, 0.429]	[0, 0.308]	[0, 0.29]	[0, 0.308]	[−, −]	[0, 0.217]
1	[0, 0.807]	[0, 0.529]	[0, 0.408]	[0.006, 0.446]	[0.002, 0.445]	[0, 0.285]	[0.011, 0.381]
2	[0, 0.907]	[0, 0.629]	[0, 0.508]	[0.037, 0.553]	[0.025, 0.556]	[0, 0.447]	[0.044, 0.503]
3	[0, 1]	[0, 0.729]	[0, 0.608]	[0.088, 0.619]	[0.066, 0.652]	[0.015, 0.584]	[0.093, 0.606]
4	[0, 1]	[0, 0.829]	[0.092, 0.708]	[0.151, 0.709]	[0.121, 0.737]	[0.096, 0.703]	[0.153, 0.696]
5	[0, 1]	[0.07, 0.929]	[0.192, 0.808]	[0.223, 0.777]	[0.187, 0.812]	[0.19, 0.809]	[0.18, 0.688]

Table 1: This table shows the 95% ( $\alpha = 0.05$ ) confidence intervals for  $p$  obtained with different methods when  $X_1 \sim \text{Binom}(10, p)$  is observed. The intervals obtained with concentration inequalities are wide. On the whole the Bayesian intervals obtained with Jeffrey prior are the narrowest while the intervals obtained with the Wald and level-sets methods are close. When  $x_1 = 0$ , the Wald interval cannot be computed.

$x_1$	Tchebychev	Hoeffding	Uniform	Level	Clopper	Wald	Jeffrey
0	[0, 0.5]	[0, 0.303]	[0, 0.218]	[0, 0.166]	[0, 0.168]	[−, −]	[0, 0.117]
1	[0, 0.55]	[0, 0.353]	[0, 0.268]	[0.003, 0.244]	[0.001, 0.248]	[0, 0.145]	[0.005, 0.211]
2	[0, 0.6]	[0, 0.403]	[0, 0.318]	[0.019, 0.319]	[0.012, 0.316]	[0, 0.231]	[0.021, 0.284]
3	[0, 0.65]	[0, 0.453]	[0, 0.368]	[0.043, 0.372]	[0.032, 0.378]	[0, 0.306]	[0.044, 0.349]
4	[0, 0.7]	[0, 0.503]	[0, 0.418]	[0.072, 0.423]	[0.057, 0.436]	[0.024, 0.375]	[0.072, 0.408]
5	[0, 0.75]	[0, 0.553]	[0.032, 0.468]	[0.105, 0.474]	[0.086, 0.491]	[0.06, 0.439]	[0.102, 0.464]
6	[0, 0.5]	[0, 0.603]	[0.082, 0.518]	[0.14, 0.525]	[0.119, 0.542]	[0.099, 0.5]	[0.136, 0.517]
7	[0, 0.5]	[0, 0.653]	[0.132, 0.568]	[0.167, 0.576]	[0.154, 0.592]	[0.14, 0.559]	[0.172, 0.568]
8	[0, 0.55]	[0, 0.703]	[0.182, 0.618]	[0.209, 0.627]	[0.191, 0.639]	[0.185, 0.614]	[0.211, 0.616]
9	[0, 0.6]	[0, 0.753]	[0.232, 0.668]	[0.245, 0.68]	[0.23, 0.684]	[0.231, 0.668]	[0.251, 0.662]
10	[0, 0.65]	[0, 0.803]	[0.282, 0.718]	[0.293, 0.707]	[0.272, 0.728]	[0.28, 0.719]	[0.293, 0.707]

Table 2: This table shows the 95% ( $\alpha = 0.05$ ) confidence intervals for  $p$  obtained with the different methods when  $X_1 \sim \text{Binom}(20, p)$  is observed. As for  $n = 10$ , the intervals given by concentration inequalities are rough and the Bayesian intervals obtained with the Jeffrey prior are the narrowest. The level-sets intervals are narrower than the Wald intervals for medium  $x_1$ .

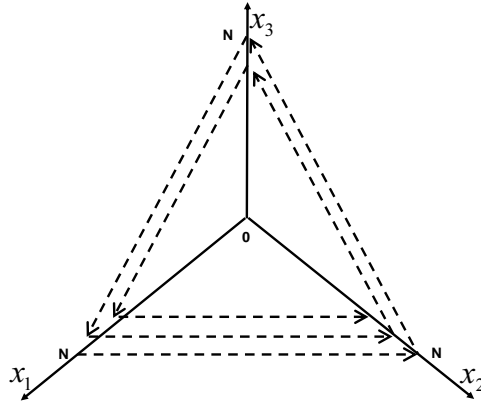


Figure 1: This figure represents the construction of  $A_k$ . Set  $A_0 = \emptyset$  and  $A_1 = \{(n, 0, 0)\}$ . The point in  $A_1$  is at the beginning of the starting arrow represented in dotted line. Each time the arrow meets a point in the simplex, this point is added to  $A_k$  to give  $A_{k+1}$ . The set obtained with the three first arrows is invariant by permutation of coordinates.

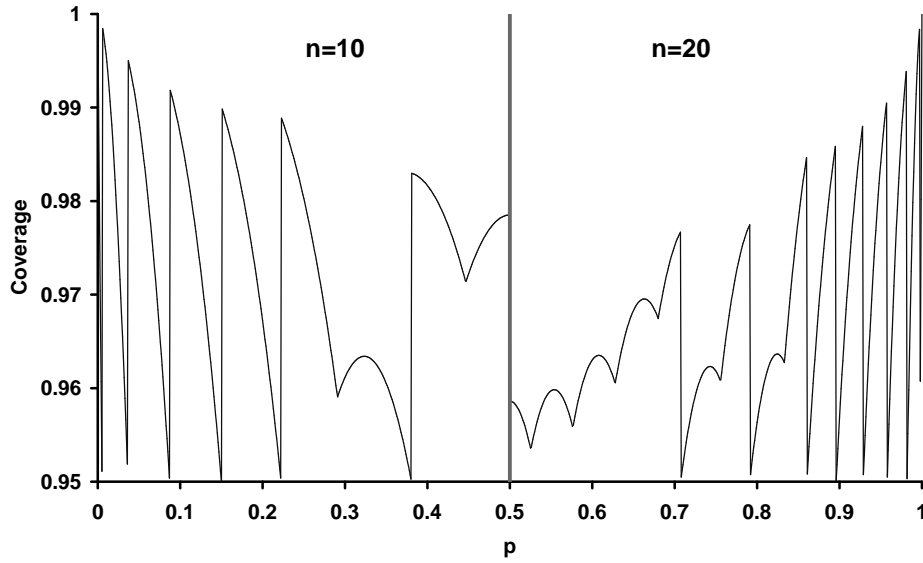


Figure 2: The left hand side curve represents for  $0 \leq p \leq \frac{1}{2}$  the coverage of our level-sets method obtained with  $n = 10$ . The right hand side curve represents for  $\frac{1}{2} < p \leq 1$  the coverage obtained with  $n = 20$ . Since the coverage for  $p$  is equal to one for  $1 - p$ , these two curves can be completed by symmetry with respect to the vertical line  $p = \frac{1}{2}$ . All the coverages are greater than or equal to the prescribed level 0.95 ( $\alpha = 0.05$ ). The staircase effect is quite clear here.

	Clopper-Pearson	Jeffrey	Level	Wald	Uniform
$x = (3, 2, 5)$	0.479	0.369	0.340	0.339	0.557
$x = (0, 2, 8)$	0.682	0.132	0.203	0.197	0.265

Table 3: When  $d = 3$ , the area of  $\Lambda_3$  is  $\sqrt{3}/2 \simeq 1.22$ . This table gives the area of the 95% confidence region ( $\alpha = 0.05$ ) for the different methods. When  $x$  belongs to the boundary of  $E_3$  the Bayesian method with Jeffrey prior gives the region with the smallest area. The Wald and the Level-set methods give regions with smaller volumes than the other methods when  $x$  is in the interior of  $E_3$ .

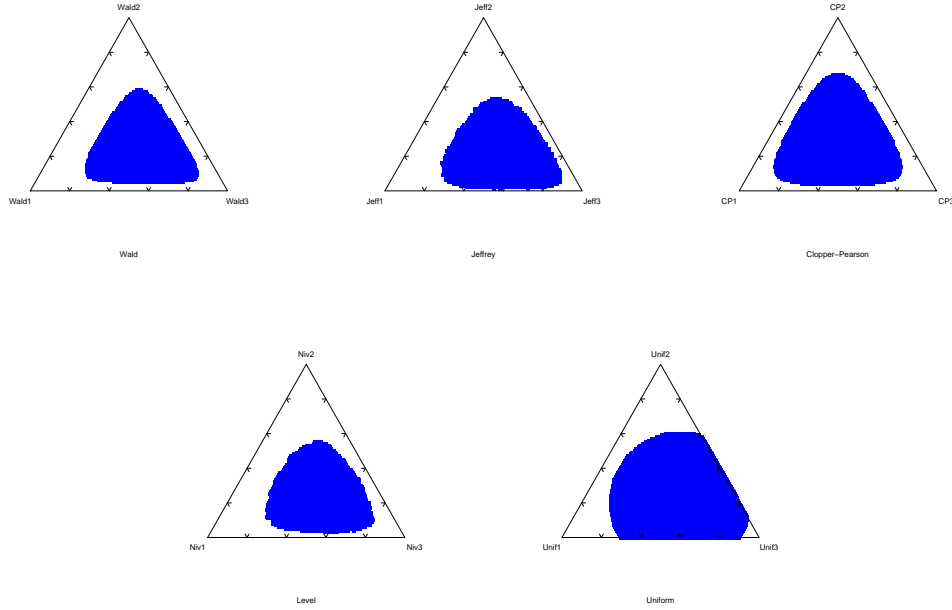


Figure 3: In barycentric coordinates, the 95% ( $\alpha = 0.05$ ) confidence regions obtained for  $p$  for the observation  $x = (3, 2, 5)$  of  $\mathcal{M}_3(10, p)$ . The Level-set and the Wald methods give regions with smaller volume than the others. The Uniform concentration method and the extended Clopper-Pearson method give large regions.

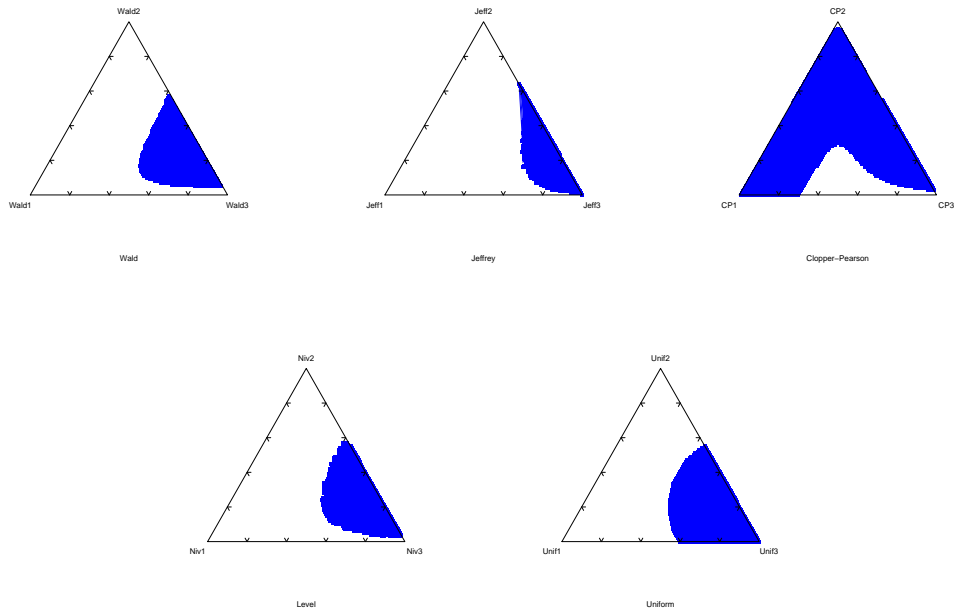


Figure 4: This figure represents in barycentric coordinates the 95% confidence regions ( $\alpha = 0.05$ ) for  $p$  obtained for the observation  $x = (0, 2, 8)$  of the trinomial  $\mathcal{M}_3(10, p)$ . Here the observation belongs to the boundary of  $E_3$ . The Bayesian method with Jeffrey prior gives the smallest region probably because it puts mass near the boundary of the parameter space. Unfortunately, the coverage probability of this method is not well controlled. The level-set and Wald methods give comparable regions. The Uniform concentration method and the extended Clopper-Pearson method give large regions.

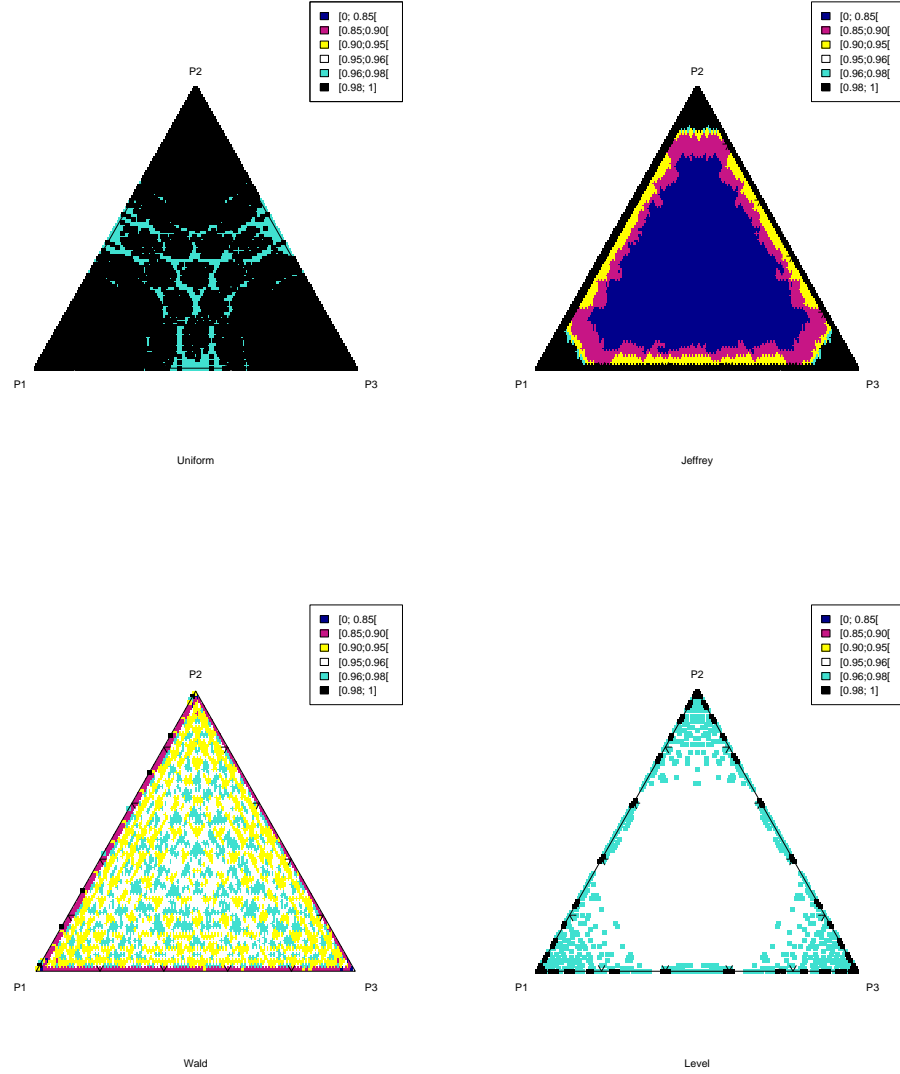


Figure 5: This figure represents in barycentric coordinates the actual coverage of each method. When the color is clear, the actual coverage is close to the prescribed coverage 0.95. As expected, the uniform-concentration method is very conservative. The Wald and the Bayesian methods fail to guarantee the prescribed coverage: there exist values of  $p$  for which the actual coverage is definitely too far from 0.95. The Level-set method is the only method that guarantees an actual coverage close to 0.95 for most of the points of the parameter space  $\Lambda_3$ .