# On the Sensing Sample Size for the Estimation of Primary Channel Occupancy Rate in Cognitive Radio

Miguel López-Benítez
Dept. Electrical Engineering and Electronics
University of Liverpool, United Kingdom
Email: M.Lopez-Benitez@liverpool.ac.uk

Janne Lehtomäki
Centre for Wireless Communications
University of Oulu, Finland
Email: jannel@ee.oulu.fi

***Abstract*—Dynamic Spectrum Access (DSA)/Cognitive Radio (CR) systems can benefit from the knowledge of the activity statistics of primary channels. A particularly relevant statistic is the Channel Occupancy Rate (COR) of a primary channel, which represents the probability that a channel is occupied by a primary user. The COR can be estimated based on a set of binary (idle/busy) spectrum sensing decisions. However, an important practical question is how many sensing observations are necessary (i.e., the sensing sample size) in order to estimate the COR of a primary channel to a certain desired level of accuracy. This work analyses the problem of estimating the COR of a primary channel based on spectrum sensing decisions and derives a tight closed-form expression for the required sensing sample size. Moreover, an iterative algorithm is proposed to accurately determine the sensing sample size required to estimate the primary COR to a desired level of accuracy. The obtained results demonstrate the ability of the proposed algorithm to make an arbitrarily accurate estimation of the real COR of unknown primary channels using the essential minimum number of sensing samples.**

***Keywords*—Cognitive radio, dynamic spectrum access, primary activity statistics, channel occupancy rate, spectrum sensing.***

## I. Introduction

Dynamic Spectrum Access (DSA) [1, 2] based on the Cognitive Radio (CR) paradigm [3–5] has the potential to increase spectrum efficiency by enabling unlicensed (secondary) users to opportunistically access licensed bands that are temporarily and/or spatially unused by the licensed (primary) users. By sensing the primary spectrum periodically, DSA/CR users can identify spectrum gaps and transmit opportunistically without causing harmful interference to primary users. Spectrum sensing methods make a binary decision on the busy/idle state of a primary channel based on a set of signal samples of the channel. While the main purpose of spectrum sensing is the detection of transmission opportunities that can be exploited by the DSA/CR system, the sequence of spectrum sensing decisions can certainly be exploited in order to produce more elaborated information. In particular, the sequence of binary busy/idle decisions for a channel can be used to compute relevant channel activity statistics such as the duration of the busy/idle periods, their minimum value, mean and variance, or the underlying distribution [6]. The knowledge of primary activity statistics can be exploited by DSA/CR systems in several ways, including the prediction of future trends in the spectrum occupancy [7, 8], the selection of the most appropriate channel/band of operation [9–13], and other spectrum and radio resource management decisions to optimise the system performance and spectrum efficiency [14–18].

A particularly relevant activity statistic is the Channel Occupancy Rate (COR) of a primary channel. From a theoretical point of view, the COR can be defined as the probability that a channel is occupied by a primary user. From an empirical point of view, the COR can be defined as the fraction of time that a primary channel is busy, which can be estimated in practice as the fraction of sensing events where the channel is observed as busy. The main interest of the COR as a channel activity statistic relies on its ability to summarise in a single numerical quantity the amount of spectrum opportunities that a DSA/CR system can expect to find in a primary channel. A DSA/CR system can estimate the COR for several primary channels (based on spectrum sensing observations) and then select the channel with the lowest COR (i.e., the channel with the highest expected amount of transmission opportunities). Since the primary channel of operation selected by a DSA/CR system will have an important impact on the overall system performance, it is essential for the DSA/CR system that the CORs of the candidate primary channels are estimated accurately. In this context, an important practical question is how many sensing observations are necessary (i.e., the sensing sample size) in order to estimate the COR of a primary channel to a certain desired level of accuracy. This work analyses the problem of estimating the COR of a primary channel based on spectrum sensing decisions and derives a tight closed-form expression for the required sensing sample size. The required sample size is found to be dependent on the COR itself, which is unknown. To solve this problem, an iterative algorithm is proposed. The obtained results demonstrate that the proposed algorithm is capable to determine accurately the sensing sample size required to estimate the COR of a primary channel to a specified level of accuracy. Therefore, this algorithm enables DSA/CR systems to estimate the real COR of unknown primary channels to an arbitrary level of accuracy using the essential minimum number of sensing samples.

The rest of this work is organised as follows. First, Section II presents the system model considered in this work. Then, a theoretical analysis is performed in Section III, where a closed-form expression for the required sensing sample size is derived. The required sample size is found to be dependent on the COR itself. To address this problem, several practical solutions are proposed in Section IV. The performance of the proposed solutions is assessed by means of simulations. The obtained simulation results are presented and discussed in Section V. A discussion on the practical use of the proposed solution in DSA/CR systems is provided in Section VI. Finally, Section VII summarises and concludes the paper.

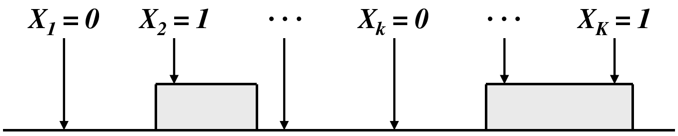


Fig. 1. Considered system model.

## II. System Model

A DSA/CR system makes an estimation, denoted as $\widehat{\Psi}$, of the real COR of a primary channel, denoted as $\Psi \in [0, 1]$. To this end, the DSA/CR system senses the primary channel periodically and takes a certain number of spectrum sensing samples, $K$, as illustrated in Figure 1. The set of sensing samples $\mathbf{X} = \{X_k\}_{k=1}^K$ contains a total of $K$ samples, each of which represents the outcome of one spectrum sensing event, which can be either an idle channel (denoted as $X_k = 0$) or a busy channel (denoted as $X_k = 1$). Based on the set $\mathbf{X}$ of spectrum sensing samples, the DSA/CR system makes an estimation $\widehat{\Psi}$ of the COR as:

$$\widehat{\Psi} = \frac{1}{K} \sum_{k=1}^{K} X_k \tag{1}$$

The estimation $\widehat{\Psi}$ obtained in (1) is considered by the DSA/CR system as a representative value of the real COR $\Psi$ of the primary channel and used in subsequent spectrum and radio resource management decisions (e.g., selection of the primary channel with lowest estimated COR value).

In this context, an important practical question is how many sensing samples $K$ are required (i.e., the sensing sample size) in order to guarantee that the estimated COR $\widehat{\Psi}$ does not differ from the real COR $\Psi$ by more than a specified maximum absolute estimation error $\varepsilon_{\text{abs}}^{\max}$ (i.e., $\varepsilon_{\text{abs}} = |\Psi - \widehat{\Psi}| \leq \varepsilon_{\text{abs}}^{\max}$).

## III. Theoretical Analysis

The elements of the set $\mathbf{X} = \{X_k\}_{k=1}^K$ can take one of two possible values, either $X_k = 0$ (idle channel) or $X_k = 1$ (busy channel). Therefore, each sensing sample $X_k$ is a random variable that follows a Bernoulli distribution with "success" probability given by the real COR, $P(X_k = 1) = \Psi$. The mean and variance of the distribution are given by $\mathbb{E}\{X_k\} = \Psi$ and $\mathbb{V}\{X_k\} = \Psi(1 - \Psi)$, respectively.

The COR is estimated in (1) as the sample mean of the set $\mathbf{X}$ by averaging a number of Bernoulli-distributed random variables. The estimated COR $\widehat{\Psi}$ can be thought of as the probability of obtaining $m$ "successes" (i.e., $X_k = 1$) out of $K$ Bernoulli trials, where each trial has probability of "success" $P(X_k = 1) = \Psi$ and probability of "failure" $P(X_k = 0) = 1 - \Psi$. Therefore, the estimated COR $\widehat{\Psi}$ follows a Binomial distribution with Probability Mass Function (PMF) given by:

$$P\left(\widehat{\Psi} = \frac{m}{K}\right) = \binom{K}{m} \Psi^m (1 - \Psi)^{K-m} \quad \text{for } m = 0, 1, \ldots, K \tag{2}$$

TABLE I. Relation between $\kappa$ and $\rho$ for various concentration inequalities [20].

| Inequality | Relation |
|---|---|
| Chebyshev | $\kappa \geq \frac{1}{\sqrt{1-\rho}}$ |
| Cantelli | $\kappa \geq \sqrt{\frac{\rho}{1-\rho}}$ |
| Vysochanskij-Petunin | $\kappa \geq \frac{2}{3\sqrt{1-\rho}}$ |
| Sobolev | $\kappa \geq \sqrt{-4\ln\left(\frac{1-\rho}{2}\right)}$ |
| Bernstein | $\kappa \geq \sqrt{-2\ln(1-\rho)}$ |

The mean and variance of the estimated COR are given by:

$$\mathbb{E}\{\widehat{\Psi}\} = \mathbb{E}\left\{\frac{1}{K}\sum_{k=1}^{K} X_k\right\} = \mathbb{E}\{X_k\} = \Psi \tag{3}$$

$$\mathbb{V}\{\widehat{\Psi}\} = \mathbb{V}\left\{\frac{1}{K}\sum_{k=1}^{K} X_k\right\} = \frac{\mathbb{V}\{X_k\}}{K} = \frac{\Psi(1-\Psi)}{K} \tag{4}$$

As it can be appreciated, the estimated COR has a certain distribution around the real COR value with a variance/width that depends on the sensing sample size. According to (4), the width of the distribution can be reduced by increasing the value of $K$, meaning that the estimated COR $\widehat{\Psi}$ can be made arbitrarily accurate by increasing the sensing sampling size.

The weak law of large numbers [19] states that it is possible to define a confidence interval of $\kappa$ standard deviations around the expected value of an estimator $\omega$ such that the estimated values are within that interval with a minimum probability $\rho$ (confidence level):

$$P\left(|\mathbb{E}\{\omega\} - \omega| \leq \kappa\sqrt{\mathbb{V}\{\omega\}}\right) \geq \rho \tag{5}$$

which for the estimator in (1), $\omega = \widehat{\Psi}$, reduces to:

$$P\left(|\mathbb{E}\{\widehat{\Psi}\} - \widehat{\Psi}| \leq \kappa\sqrt{\mathbb{V}\{\widehat{\Psi}\}}\right) \geq \rho \tag{6}$$

$$P\left(|\Psi - \widehat{\Psi}| \leq \kappa\sqrt{\frac{\Psi(1-\Psi)}{K}}\right) \geq \rho \tag{7}$$

$$P\left(\varepsilon_{\text{abs}} \leq \varepsilon_{\text{abs}}^{\max}\right) \geq \rho \tag{8}$$

where $\varepsilon_{\text{abs}} = |\Psi - \widehat{\Psi}|$ and $\varepsilon_{\text{abs}}^{\max} = \kappa\sqrt{\Psi(1-\Psi)/K}$. Hence, the sample size corresponding to a certain maximum absolute error is given by:

$$K = \left(\frac{\kappa}{\varepsilon_{\text{abs}}^{\max}}\right)^2 \Psi(1 - \Psi) \tag{9}$$

The value of $\kappa$ for a certain confidence level $\rho$ can be derived from concentration inequalities [20]. Some examples are shown in Table I. It is worth noting that the relations in Table I are obtained from inequalities, which lead to (rather loose) upper bounds on the maximum absolute error as shown in Figure 2. As result, these relations, when introduced in (9), lead to an overestimation of the required sample size.

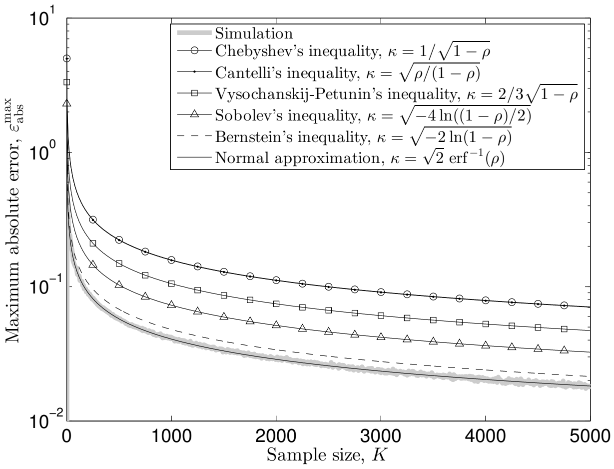


Fig. 2. Maximum absolute error of the estimated COR, $\varepsilon_{\text{abs}}^{\text{max}}$, observed in a fraction $\rho = 0.99$ of cases (i.e., at the 99% percentile) as a function of the sample size, $K$, for a channel with COR $\Psi = 0.5$.

A more accurate estimation of the required sample size can be obtained by noting that the distribution of the estimated COR $\widehat{\Psi}$ can be approximated (according to the Central Limit Theorem) by a normal distribution with the same mean $\mathbb{E}\{\widehat{\Psi}\}$ and variance $\mathbb{V}\{\widehat{\Psi}\}$ as the actual binomial distribution. The normal approximation of a binomial random variable is accurate when $K\Psi \gg 1$ and $K\Psi(1-\Psi) \gg 1$ [19], which is true is most practical cases. Based on this approximation, the inequalities (6)–(8) can be expressed as:

$$\begin{aligned} P\left(\varepsilon_{\text{abs}} \leq \varepsilon_{\text{abs}}^{\text{max}}\right) &= P\left(|\mathbb{E}\{\widehat{\Psi}\} - \widehat{\Psi}| \leq \kappa\sqrt{\mathbb{V}\{\widehat{\Psi}\}}\right) \\ &= \int_{\mathbb{E}\{\widehat{\Psi}\}-\kappa\sqrt{\mathbb{V}\{\widehat{\Psi}\}}}^{\mathbb{E}\{\widehat{\Psi}\}+\kappa\sqrt{\mathbb{V}\{\widehat{\Psi}\}}} \frac{e^{-\frac{1}{2}\left(\frac{\widehat{\Psi}-\mathbb{E}\{\widehat{\Psi}\}}{\sqrt{\mathbb{V}\{\widehat{\Psi}\}}}\right)^2}}{\sqrt{2\pi\mathbb{V}\{\widehat{\Psi}\}}} d\widehat{\Psi} \\ &= \mathrm{erf}\left(\frac{\kappa}{\sqrt{2}}\right) \geq \rho \end{aligned} \tag{10}$$

Solving (10) for $\kappa$ yields the relation $\kappa \geq \sqrt{2}\,\mathrm{erf}^{-1}(\rho)$, which provides a more accurate estimation of the maximum absolute error as illustrated in Figure 2. Introducing the relation $\kappa \geq \sqrt{2}\,\mathrm{erf}^{-1}(\rho)$ in (9) yields the much tighter bound:

$$K \geq 2\left(\frac{\mathrm{erf}^{-1}(\rho)}{\varepsilon_{\text{abs}}^{\text{max}}}\right)^2 \Psi(1-\Psi) \tag{11}$$

Thus, the minimum (optimum) number of sensing samples required to guarantee that the absolute estimation error of the COR of a primary channel does not exceed the value $\varepsilon_{\text{abs}}^{\text{max}}$ with a probability (confidence level) of $\rho$ is given by:

$$K^* = \left\lceil 2\left(\frac{\mathrm{erf}^{-1}(\rho)}{\varepsilon_{\text{abs}}^{\text{max}}}\right)^2 \Psi(1-\Psi) \right\rceil \tag{12}$$

where $\lceil\cdot\rceil$ denotes the ceil operator (i.e., rounding to the nearest integer towards infinity).

## IV. Proposed Practical Solutions

The result in (12) indicates that the minimum sensing sample size required to estimate the COR of a primary channel to a certain level of accuracy depends on the COR itself. This means that a DSA/CR system cannot determine in advance how many sensing samples need to be taken from a primary channel in order to produce a reliable estimation of its COR. To address this problem, two solutions are proposed, which are presented and discussed in the following subsections.

### A. *Worst-case design*

A possible solution is to adopt a worst-case design where the number of samples taken from a primary channel guarantees that the COR is estimated to the desired level of accuracy even in the worst possible case. The sample size required for the worst possible case can be determined by finding the value of $\Psi$ that maximises (12), which can be obtained by solving the equation $\partial K^*/\partial\Psi = 0$ and is found to be $\Psi = 1/2$. It can be shown that $\partial^2 K^*/\partial\Psi^2 < 0$. Therefore, $K^*$ has a global maximum at $\Psi = 1/2$, which is given by:

$$K_{\max}^* = \left\lceil \frac{1}{2}\left(\frac{\mathrm{erf}^{-1}(\rho)}{\varepsilon_{\text{abs}}^{\text{max}}}\right)^2 \right\rceil \tag{13}$$

The result in (13) does not require the knowledge of the real COR (it only requires the desired maximum absolute error $\varepsilon_{\text{abs}}^{\text{max}}$ and the probability $\rho$ that it is not exceeded) and is able to guarantee the desired level of accuracy irrespective of the real COR. However, an important drawback of this method is that the sample size predicted by (13) can be significantly higher than the sample size actually required to provide a reliable COR estimation. In practice, this means that the DSA/CR system might need a significantly long time in order to collect the number of samples $K_{\max}^*$ predicted by (13), even if such a high number is not actually needed in order to estimate the COR to the desired level of accuracy. To cope with this drawback, an iterative algorithm is proposed in Section IV-B.

### B. *Iterative algorithm*

In order to overcome the main drawback of the worst-case design discussed in Section IV-A, the iterative procedure of Algorithm 1 is proposed. The objective of this algorithm is to produce an empirical estimation of the required sample size as close to $K^*$ as possible, but not lower, so that the DSA/CR system can produce a reliable COR estimation within a period of time as short as allowed by the desired level of accuracy.

The proposed algorithm takes as input parameters the desired maximum absolute error $\varepsilon_{\text{abs}}^{\text{max}}$ and the probability $\rho$ that it is not exceeded along with the value selected for the algorithm parameter $\beta \in (0,1)$. The output provided by the algorithm is an estimation $\widehat{\Psi}$ of the COR that meets the desired level of accuracy as specified by $\varepsilon_{\text{abs}}^{\text{max}}$ and $\rho$. The main difference with respect to the method of Section IV-A is that the COR estimation $\widehat{\Psi}$ is produced based on a number of samples close to the minimum required value $K^*$ given by (12), thus avoiding the need to use the sample size $K_{\max}^*$ given by (13), which corresponds to the worst case ($K^* \leq K_{\max}^*$ ).

**Algorithm 1** Iterative algorithm

**Input:** $\varepsilon_{\text{abs}}^{\max} \in (0,1)$, $\rho \in (0,1)$, $\beta \in (0,1)$
**Output:** $\widehat{\Psi} \in [0,1]$
1: $K_{\max}^* \leftarrow \lceil (\text{erf}^{-1}(\rho)/\varepsilon_{\text{abs}}^{\max})^2/2 \rceil$
2: $K \leftarrow \lceil \beta K_{\max}^* \rceil$
3: Sense the channel until a sample set $\{X_k\}_{k=1}^K$ is available
4: $\widehat{\Psi} \leftarrow (1/K) \sum_{k=1}^K X_k$
5: $K_{\text{new}} \leftarrow \lceil 2\,(\text{erf}^{-1}(\rho)/\varepsilon_{\text{abs}}^{\max})^2\, \widehat{\Psi}(1-\widehat{\Psi}) \rceil$
6: **if** $K_{\text{new}} > K$ **then**
7: $\quad K \leftarrow K_{\text{new}}$
8: $\quad$ Go to step 3
9: **end if**

First, the algorithm computes, based on (13), the sample size $K_{\max}^*$ required in the worst possible case (step 1). The obtained $K_{\max}^*$ is then employed to compute an initial value of the sample size as $K = \lceil \beta K_{\max}^* \rceil$ (step 2), where $\beta \in (0,1)$ is an algorithm parameter that represents the fraction of $K_{\max}^*$ that is considered as the initial sample size. For example, a value $\beta = 0.1$ means that the initial sample size considered by the algorithm is one tenth of the maximum required sample size $K_{\max}^*$. In the first iteration of the algorithm, a set of $K = \lceil \beta K_{\max}^* \rceil$ sensing samples is collected from a primary channel (step 3) and used to produce, based on (1), an estimation $\widehat{\Psi}$ of the COR (step 4). In order to determine whether more sensing samples are required to attain the desired level of accuracy, a new sample size $K_{\text{new}}$ is computed (step 5) based on (12) and making use of the COR value $\widehat{\Psi}$ estimated in step 4. If the obtained sample size $K_{\text{new}}$ is greater than the sample size $K$ employed in the current iteration (step 6), then more sensing samples are required. In such a case, the value of the sample size $K$ is updated to the new estimated sample size $K_{\text{new}}$ (step 7) and a new iteration is executed (step 8). In the next iteration of the algorithm, more sensing samples are collected until the new/larger amount $K$ of sensing samples is available (step 3) and a new estimation of the COR $\widehat{\Psi}$ is computed based on the larger sample set (step 4). The most recent COR estimation $\widehat{\Psi}$ is used in each iteration in order to compute a new sample size $K_{\text{new}}$ (step 5) and determine whether more samples are required (step 6). The algorithm is designed to iterate until the employed sensing sample size $K$ is greater than (or equal to) the optimum value $K^*$ predicted by (12). When the employed sample size exceeds $K^*$, the condition in step 6 becomes false ($K_{\text{new}} \leq K$) and the algorithm stops. When the algorithm stops, the employed sample size is expected to be close to the optimum value $K^*$ and the COR value $\widehat{\Psi}$ estimated in the last iteration of the algorithm is expected to meet the desired level of accuracy as specified by $\varepsilon_{\text{abs}}^{\max}$ and $\rho$.

An important practical aspect of the algorithm is its convergence (i.e., its ability to stop instead of iterating indefinitely). A new iteration implies the collection of additional sensing samples and therefore a larger sample set, which in turn implies a lower COR estimation error. After a certain number of iterations, the employed sample size will be large enough to produce a sufficiently accurate estimation of the COR. When this occurs, the new sample size computed in step 5 will be lower than the sample size employed in the last iteration. In such a case, the conditional test of step 6 guarantees that no more iterations are executed and thus the algorithm converges.

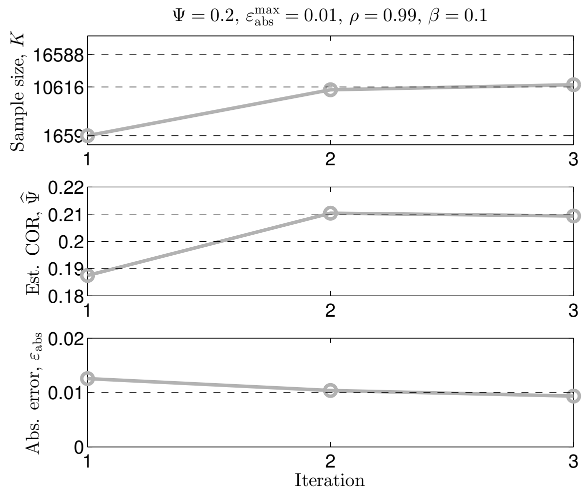


Fig. 3. Example of the operation of the proposed iterative algorithm.

TABLE II. EXAMPLE OF THE OPERATION OF THE PROPOSED ITERATIVE ALGORITHM.

| Iteration 1 | Iteration 2 | Iteration 3 | Target |
|---|---|---|---|
| $K_1$ = 1659 | $K_2$ = 10108 | $K_3$ = 11023 | $K^*$ = 10616 |
| $\widehat{\Psi}_1$ = 0.1875 | $\widehat{\Psi}_2$ = 0.2104 | $\widehat{\Psi}_3$ = 0.2093 | $\Psi$ = 0.2 |
| $\varepsilon_{\text{abs},1}$ = 0.0125 | $\varepsilon_{\text{abs},2}$ = 0.0104 | $\varepsilon_{\text{abs},3}$ = 0.0093 | $\varepsilon_{\text{abs}}^{\max}$ = 0.01 |
| $K_{\text{new},1}$ = 10108 | $K_{\text{new},2}$ = 11023 | $K_{\text{new},3}$ = 10981 | $K_{\text{new}} < K$ |

## V. SIMULATION RESULTS

The performance of the proposed iterative algorithm is evaluated by means of simulations. First, Section V-A presents an example illustrating the operation of the iterative algorithm. Afterwards, Section V-B presents a comprehensive performance evaluation of the proposed algorithm based on extensive simulations over a wide range of operation conditions.

### A. Operation of the iterative algorithm

Figure 3 and Table II show an example of the operation of the proposed iterative algorithm. The primary channel under study has a real COR of $\Psi = 0.2$ and the DSA/CR system aims at estimating such value with a maximum estimation error of $\varepsilon_{\text{abs}}^{\max} = 0.01$, which is to be guaranteed with a probability (confidence level) of $\rho = 0.99$. According to the theoretical results obtained in Section III, the sample size required for the desired level of accuracy in the worst case is $K_{\max}^* = 16588$ and the minimum sample size actually required for the real COR is $K^* = 10616$ (these values are shown in Figure 3).

As it can be appreciated, the iterative algorithm stops after 3 iterations. In the first iteration, the initial sample size is $K_1 = \lceil \beta K_{\max}^* \rceil = \lceil 0.1 \cdot 16588 \rceil = 1659$. After collecting $K_1$ sensing samples, the COR estimated in the first iteration is $\widehat{\Psi}_1 = 0.1875$. This estimation differs from the real COR by $\varepsilon_{\text{abs},1} = |\Psi - \widehat{\Psi}_1| = |0.2 - 0.1875| = 0.0125$, which is greater than the target $\varepsilon_{\text{abs}}^{\max}$. In order to determine whether more samples are required, the new sample size computed in the first iteration is $K_{\text{new},1} = 10108$, which is greater than the used sample size $K_1 = 1659$. Therefore, a new

TABLE III. SIMULATION RESULTS FOR $\Psi = 0.05$.

| $\varepsilon_{\text{abs}}^{\max}$ | $\rho$ | $K_{\max}^*$ | $K^*$ | $K$ | Red. | $\widehat{\Psi}$ | $\widehat{\varepsilon}_{\text{abs}}^{\max}$ |
|---|---|---|---|---|---|---|---|
| 0.01 | 0.90 | 6764 | 1286 | 1342 | 80.2% | 0.0493 | 0.0101 |
| | 0.95 | 9604 | 1825 | 1892 | 80.3% | 0.0495 | 0.0101 |
| | 0.99 | 16588 | 3152 | 3240 | 80.5% | 0.0497 | 0.0102 |
| 0.02 | 0.90 | 1691 | 322 | 345 | 79.6% | 0.0471 | 0.0240 |
| | 0.95 | 2401 | 457 | 482 | 79.9% | 0.0476 | 0.0242 |
| | 0.99 | 4147 | 788 | 828 | 80.0% | 0.0488 | 0.0242 |
| 0.05 | 0.90 | 271 | 52 | 62 | 77.1% | 0.0382 | 0.0500 |
| | 0.95 | 385 | 73 | 80 | 79.2% | 0.0397 | 0.0500 |
| | 0.99 | 664 | 127 | 140 | 78.9% | 0.0435 | 0.0500 |

TABLE IV. SIMULATION RESULTS FOR $\Psi = 0.1$.

| $\varepsilon_{\text{abs}}^{\max}$ | $\rho$ | $K_{\max}^*$ | $K^*$ | $K$ | Red. | $\widehat{\Psi}$ | $\widehat{\varepsilon}_{\text{abs}}^{\max}$ |
|---|---|---|---|---|---|---|---|
| 0.01 | 0.90 | 6764 | 2435 | 2521 | 62.7% | 0.0996 | 0.0099 |
| | 0.95 | 9604 | 3458 | 3562 | 62.9% | 0.0998 | 0.0097 |
| | 0.99 | 16588 | 5972 | 6106 | 63.2% | 0.0998 | 0.0100 |
| 0.02 | 0.90 | 1691 | 609 | 650 | 61.6% | 0.0988 | 0.0196 |
| | 0.95 | 2401 | 865 | 914 | 61.9% | 0.0990 | 0.0201 |
| | 0.99 | 4147 | 1493 | 1560 | 62.4% | 0.0994 | 0.0203 |
| 0.05 | 0.90 | 271 | 98 | 108 | 60.2% | 0.0891 | 0.0537 |
| | 0.95 | 385 | 139 | 151 | 60.8% | 0.0910 | 0.0544 |
| | 0.99 | 664 | 239 | 262 | 60.5% | 0.0961 | 0.0540 |

TABLE V. SIMULATION RESULTS FOR $\Psi = 0.2$.

| $\varepsilon_{\text{abs}}^{\max}$ | $\rho$ | $K_{\max}^*$ | $K^*$ | $K$ | Red. | $\widehat{\Psi}$ | $\widehat{\varepsilon}_{\text{abs}}^{\max}$ |
|---|---|---|---|---|---|---|---|
| 0.01 | 0.90 | 6764 | 4329 | 4419 | 34.7% | 0.1999 | 0.0101 |
| | 0.95 | 9604 | 6147 | 6256 | 34.9% | 0.1999 | 0.0097 |
| | 0.99 | 16588 | 10616 | 10758 | 35.2% | 0.1999 | 0.0100 |
| 0.02 | 0.90 | 1691 | 1083 | 1126 | 33.4% | 0.1994 | 0.0196 |
| | 0.95 | 2401 | 1537 | 1591 | 33.7% | 0.1996 | 0.0198 |
| | 0.99 | 4147 | 2654 | 2724 | 34.3% | 0.1998 | 0.0200 |
| 0.05 | 0.90 | 271 | 174 | 189 | 30.3% | 0.1965 | 0.0491 |
| | 0.95 | 385 | 246 | 265 | 31.2% | 0.1976 | 0.0500 |
| | 0.99 | 664 | 425 | 452 | 31.9% | 0.1985 | 0.0516 |

TABLE VI. SIMULATION RESULTS FOR $\Psi = 0.5$.

| $\varepsilon_{\text{abs}}^{\max}$ | $\rho$ | $K_{\max}^*$ | $K^*$ | $K$ | Red. | $\widehat{\Psi}$ | $\widehat{\varepsilon}_{\text{abs}}^{\max}$ |
|---|---|---|---|---|---|---|---|
| 0.01 | 0.90 | 6764 | 6764 | 6764 | 0.00% | 0.4999 | 0.0099 |
| | 0.95 | 9604 | 9604 | 9603 | 0.01% | 0.5000 | 0.0099 |
| | 0.99 | 16588 | 16588 | 16587 | 0.01% | 0.5000 | 0.0100 |
| 0.02 | 0.90 | 1691 | 1691 | 1691 | 0.00% | 0.5002 | 0.0198 |
| | 0.95 | 2401 | 2401 | 2400 | 0.04% | 0.4999 | 0.0200 |
| | 0.99 | 4147 | 4147 | 4146 | 0.02% | 0.5000 | 0.0202 |
| 0.05 | 0.90 | 271 | 271 | 270 | 0.37% | 0.5003 | 0.0502 |
| | 0.95 | 385 | 385 | 384 | 0.26% | 0.5002 | 0.0497 |
| | 0.99 | 664 | 664 | 663 | 0.15% | 0.5002 | 0.0495 |

iteration is executed. In the second iteration, the considered sample size is $K_2 = K_{\text{new},1} = 10108$. After collecting the required extra samples to reach the new sample size $K_2$, a new COR estimation $\widehat{\Psi}_2 = 0.2104$ is made. The absolute error of this new estimation is $\varepsilon_{\text{abs},2} = |0.2 - 0.2104| = 0.0104$, which is still greater than the target $\varepsilon_{\text{abs}}^{\max}$. Based on the estimated $\widehat{\Psi}_2$, the new sample size is $K_{\text{new},2} = 11023 > K_2 = 10108$, so a new iteration is executed. In the third iteration, $K_3 = K_{\text{new},2} = 11023 > K^*$ and the new estimated COR is $\widehat{\Psi}_3 = 0.2093$, whose estimation error is $\varepsilon_{\text{abs},3} = |0.2 - 0.2093| = 0.0093 < \varepsilon_{\text{abs}}^{\max}$. Note that the third iteration is the first iteration where the used sample size is greater than the minimum required value $K^*$ and the absolute estimation error is lower than the target $\varepsilon_{\text{abs}}^{\max}$. In the third iteration, $K_{\text{new},3} = 10981 < K_3 = 11023$ and the algorithm correctly stops since no more samples are required and the desired level of accuracy has been achieved. As it can be observed, the proposed algorithm is able to estimate the COR with the desired level of accuracy, making use of a sample size ($K = 11023$) that is close to the minimum required value ($K^* = 10616$) and significantly lower than the sample size corresponding to a worst-case design ($K_{\max}^* = 16588$).

### B. Performance of the iterative algorithm

The performance of the iterative algorithm is evaluated by means of extensive simulations for $\Psi = \{0.05, 0.1, 0.2, 0.5\}$, $\varepsilon_{\text{abs}}^{\max} = \{0.01, 0.02, 0.05\}$ and $\rho = \{0.90, 0.95, 0.99\}$. For each possible combination of $\Psi$, $\varepsilon_{\text{abs}}^{\max}$ and $\rho$, a total of 10000 statistically independent repetitions of the same experiment are performed. The obtained results are summarised in Tables III, IV, V and VI for $\Psi = 0.05$, $\Psi = 0.1$, $\Psi = 0.2$ and $\Psi = 0.5$, respectively. It is worth noting that the results obtained for $\Psi = \zeta$ are similar to those obtained for $\Psi = 1 - \zeta$. Therefore, only results for COR values $\Psi \leq 0.5$ are shown. The results in Tables III ($\Psi = 0.05$), IV ($\Psi = 0.1$) and V ($\Psi = 0.2$) are also valid for $\Psi = 0.95$, $\Psi = 0.9$ and $\Psi = 0.8$, respectively.

The columns of each table show the following information: desired maximum absolute error $\varepsilon_{\text{abs}}^{\max}$, desired level of confidence $\rho$, sample size required in the worst case $K_{\max}^*$, sample size required for the actual COR $K^*$, average sample size predicted by the proposed iterative algorithm $K$ (i.e., sample size at which the algorithm stops, averaged over all repetitions), percentage by which the sample size $K$ reduces the sample size required in the worst case $K_{\max}^*$, average estimated COR $\widehat{\Psi}$ (i.e., COR estimated by the algorithm, averaged over all repetitions), and maximum absolute error observed in the simulation in a fraction $\rho$ of the cases $\widehat{\varepsilon}_{\text{abs}}^{\max}$.

As it can be appreciated, the iterative algorithm is able to predict accurately the minimum sample size required to estimate the COR. The predicted sample size $K$ is always greater than (or equal to) the minimum required size $K^*$, which is necessary in order to estimate the COR with the desired level of accuracy. However, $K$ is just slightly greater than the optimum value $K^*$ (in most cases, $K$ is only about 2-5% greater than $K^*$). Therefore, the sample size $K$ predicted by the algorithm is a very close approximation to the optimum value $K^*$. As a result, the proposed algorithm enables a significant reduction of the number of samples $K_{\max}^*$ that would have been taken under a worst-case design, and therefore a significant reduction in the time required by a DSA/CR system to estimate the COR. For $\Psi = 0.5$ there is no reduction since this is indeed the worst possible case. However, the reduction can be very important for lower (higher) CORs: 30-35% for $\Psi = 0.2$ ($\Psi = 0.8$), 60-63% for $\Psi = 0.1$ ($\Psi = 0.9$), and 77-81% for $\Psi = 0.05$ ($\Psi = 0.95$). However, this significant reduction of the employed sample size does not come at the expense of a degraded estimation accuracy since, at it can be appreciated, the estimated COR is in all cases very close to the real COR and its degree of accuracy $\hat{\varepsilon}_{\text{abs}}^{\max}$ is in line with the target $\varepsilon_{\text{abs}}^{\max}$. Therefore, the proposed algorithm is able not only to estimate the COR of an unknown primary channel with a specified level of accuracy but also to do so with a sample size that is very close to the minimum/optimum size required for the actual COR, thus providing significant reductions in the amount of time required to provide a reliable COR estimation.

It is also worth noting that the average number of iterations required by the algorithm to find the right value of $K$ is 2.75, which does not depend on the operation conditions. Therefore, the computational cost of the algorithm itself is very low and can therefore be employed even if the DSA/CR system operates over a large number of primary channels. Furthermore, the computational cost of collecting a large number of sensing samples and computing their average in order to estimate the COR can be significantly higher that the computational cost of iterating the proposed algorithm 2-3 times. Since the proposed algorithm leads to a significant reduction of the required sample size, it can be stated that its application also results in a reduction of the overall computational cost of estimating the COR of a primary channel.

## VI. Discussion

An interesting observation from (12) is that the primary channels that are more attractive to DSA/CR systems (i.e., with low CORs) require a smaller sample size and can be identified faster. The algorithm proposed in this work enables DSA/CR systems to effectively find primary channels with low CORs faster, regardless of the employed search strategy.

For instance, the DSA/CR system may scan several primary channels in parallel until the channel with the lowest COR is identified reliably (i.e., with the desired level of accuracy). The proposed algorithm guarantees that the first COR estimated reliably will be the lowest one. The DSA/CR system can then select that channel, thus finding the channel with the lowest COR in the shortest time. It may also happen that the first COR estimated reliably corresponds to a channel with a high COR (e.g., $\Psi = 0.05$ and $\Psi = 0.95$ require the same sample size). In such a case, the DSA/CR system can discard channels with high COR values and keep scanning the rest of channels, which also helps reducing the channel search time.

The DSA/CR system may also scan several primary channels sequentially until a channel with a COR lower than a certain target is found (e.g.,$\Psi \leq 0.2$). The sample size $K^*$ required for the target COR can be determined based on (12). Once $K^*$ samples have been collected, if the proposed algorithm determines that more samples are needed (based on the test performed in step 6), then that means that the actual COR is greater than the target COR. In such a case the DSA/CR system can discard the channel and scan a new one (i.e., there is no need to spend more time scanning the current channel), which in practice reduces the total time required to find a channel that meets the needs of the DSA/CR system.

## VII. Conclusions

This work has analysed the sample size required to provide a reliable estimation of the COR of a primary channel and has proposed an algorithm that enables DSA/CR systems to make arbitrarily accurate estimations of the real COR of unknown primary channels using the essential minimum number of sensing samples, which can reduce significantly the search time required to find an adequate channel of operation.

## References


[1] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[2] M. M. Buddhikot, "Understanding dynamic spectrum access: Taxonomy, models and challenges," in *Proc. 2nd IEEE Int'l. Symp. Dyn. Spect. Access Networks (DySPAN 2007)*, Apr. 2007, pp. 649–663.

[3] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. A. Comms.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[4] Y.-C. Liang, K.-C. Chen, G. Y. Li, and P. Mähönen, "Cognitive radio networking and communications: An overview," *IEEE Trans. Vehic. Tech.*, vol. 60, no. 7, pp. 3386–3407, Sep. 2011.

[5] M. López-Benítez, "Cognitive radio," in *Heterogeneous cellular networks: Theory, simulation and deployment*. Cambridge University Press, 2013, ch. 13, pp. 383–425.

[6] ——, "Can primary activity statistics in cognitive radio be estimated under imperfect spectrum sensing?" in *Proc. 24th Annual IEEE Int'l. Symp. Pers., In. Mob. R. Comms. (PIMRC 2013)*, Sep. 2013, pp. 750–755.

[7] S. Yarkan and H. Arslan, "Binary time series approach to spectrum prediction for cognitive radio," in *Proc. IEEE 66th Vehic. Tech. Conf. (VTC 2007 Fall)*, Sep. 2007, pp. 1563–1567.

[8] V. K. Tumuluru, P. Wang, and D. Niyato, "A neural network based spectrum prediction scheme for cognitive radio," in *Proc. 2010 IEEE Int'l. Conf. Comms. (ICC 2010)*, May 2010, pp. 1–5.

[9] X. Liu, B. Krishnamachari, and H. Liu, "Channel selection in multichannel opportunistic spectrum access networks with perfect sensing," in *Proc. 2010 IEEE Int'l. Symp. Dyn. Spect. Access Networks (DySPAN 2010)*, Apr. 2010, pp. 1–8.

[10] J. Vartiainen, M. Höyhtyä, J. Lehtomäki, and T. Bräysy, "Priority channel selection based on detection history database," in *Proc. Fifth Int'l. Conf. Cognitive Radio Oriented Wireless Networks & Comms. (CROWNCOM 2010)*, Jun. 2010, pp. 1–5.

[11] M. Höyhtyä, S. Pollin, and A. Mämmelä, "Classification-based predictive channel selection for cognitive radios," in *Proc. 2010 IEEE Int'l. Conf. Comms. (ICC 2010)*, May 2010, pp. 1–6.

[12] F. Bouali, O. Sallent, J. Pérez-Romero, and R. Agustí, "A novel spectrum selection strategy for matching multi-service secondary traffic to heterogeneous primary spectrum opportunities," in *Proc. IEEE 22nd Int'l. Symp. Pers. In. Mob. R. Comms. (PIMRC 2011)*, Sep. 2011, pp. 417–422.

[13] ——, "Strengthening radio environment maps with primary-user statistical patterns for enhancing cognitive radio operation," in *Proc. Sixth Int'l. ICST Conf. Cognitive Radio Oriented Wireless Networks and Comms. (CROWNCOM 2011)*, Jun. 2011, pp. 256–260.

[14] E. Jung and X. Liu, "Opportunistic spectrum access in multiple-primary-user environments under the packet collision constraint," *IEEE/ACM Trans. Networking*, vol. 20, no. 2, pp. 501–514, Apr. 2012.

[15] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive medium access: constraining interference based on experimental models," *IEEE J. Sel. Areas Comms.*, vol. 26, no. 1, pp. 95–105, Jan. 2008.

[16] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Sel. Areas Comms.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.

[17] H. Su and X. Zhang, "Cross-layer-based opportunistic MAC protocols for QoS provisioning over cognitive radio wireless networks," *IEEE J. Sel. Areas Comms.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.

[18] A. Anandkumar, N. Michael, A. K. Tang, and A. Swami, "Distributed algorithms for learning and cognitive medium access with logarithmic regret," *IEEE J. Sel. Areas Comms.*, vol. 29, no. 4, pp. 731–745, 2011.

[19] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*, 4th ed. Boston: McGraw-Hill, 2002.

[20] R. I. Savage, "Probability inequalities of the Tchebycheff type," *J. Research of the National Bureau of Standards-B.Mathematics and Mathematical Physics*, vol. 65B, no. 3, pp. 211–226, Jul. 1961.