Analysis of the Sample Size Required for an Accurate Estimation of Primary Channel Activity Statistics under Imperfect Spectrum Sensing

Ogeen H. Toma*†, Miguel López-Benítez*‡ and Dhaval K. Patel§
*Department of Electrical Engineering and Electronics, University of Liverpool, United Kingdom
†Department of Energy Engineering, Duhok Polytechnic University, Iraq
‡ARIES Research Centre, Antonio de Nebrija University, Spain
§School of Engineering and Applied Science, Ahmedabad University, India
Email: ogeen.toma@liverpool.ac.uk, M.Lopez-Benitez@liverpool.ac.uk

***Abstract*—Primary channel activity statistics play an important role in improving the performance of Dynamic Spectrum Access (DSA) / Cognitive Radio (CR) systems. The statistical information of the idle/busy periods of a primary channel can be estimated based on the outcomes of spectrum sensing. Recent studies have shown that these statistics can be estimated accurately even under Imperfect Spectrum Sensing (ISS) scenarios. Those studies, however, have assumed no constraints on the required sample size of observations of the idle/busy periods in order to provide accurate estimation (i.e., large sample size was assumed to test the accuracy of these statistics estimation methods). In real-world scenario, DSA/CR systems are limited to the hardware design capabilities, which include limited memory capacity, energy consumption and computational capability. As a result, it is very important to find how many samples of the idle/busy periods are required to provide an acceptable level of accuracy for the estimated statistics. Therefore, this work analyses the impact of the sample size on the estimation of the primary channel statistics under ISS and it finds closed-form expressions for the required sample size of the idle/busy periods to achieve a targeted accuracy. In addition, the analytical results achieved in this work are validated by means of simulations and hardware experiments.**

***Index Terms*—Cognitive radio, dynamic spectrum access, spectrum sensing, primary channel activity statistics.**

I. Introduction

As we are stepping into the era of the fifth generation (5G) of wireless communications, we are expecting to witness new emerging wireless technologies and a rapid growth in the number of wireless-connected devices. Such growing tendency also brings burdens of rising demand for frequency spectrum to support such tremendous number of interconnected terminals. Dynamic Spectrum Access (DSA) [1] based on Cognitive Radio (CR) [2] concept is an effective solution to overcome spectrum scarcity problem. In such system, Secondary Users (SUs) are able to exploit the unused patterns of the frequency channel that is assigned to a Primary User (PU). Spectrum sensing is the key enabler of the DSA/CR systems, which enables SUs to monitor and access the primary channel whenever the PU is idle and to evacuate the channel whenever the PU returns to use the channel (i.e., busy). It is important to highlight the two operation scenarios of spectrum sensing. First, Perfect Spectrum Sensing (PSS), which can be assumed when DSA/CR system operates under sufficiently high SNR conditions where no sensing errors could occur. Second, Imperfect Spectrum Sensing (ISS), which can be assumed when DSA/CR system operates under low SNR conditions where sensing errors are likely to occur. Due to the noise and fading channels in wireless communications, ISS is a more realistic assumption for DSA/CR systems.

978-1-7281-4490-0/20/$31.00 © 2020 IEEE

Although the objective of spectrum sensing is to monitor the state of the primary channel whether is idle or busy, the outcomes of spectrum sensing can also be exploited to obtain statistical information about the primary channel. These statistics are very beneficial for enhancing the performance of DSA/CR systems. For example, they can help in predicting the future trends of spectrum occupancy [3], [4], they can help in selecting the most underutilised primary channel [5]–[7] and they can also improve the performance by mitigating the interference between SUs and PUs [8].

Several works in the literature have analysed the estimation of the primary channel activity statistics based on the spectrum sensing decisions. The majority of the works, however, have focused on the estimation of the primary channel duty cycle as in [9]–[11]. The mean of the idle/busy periods has been studied in [12], [13]. A more comprehensive analysis for a broader range of primary channel statistics (minimum, mean, variance, skewness, kurtosis, duty cycle and distribution) has been studied under PSS in [14]. On the other hand, [15] has analysed the statistics (minimum, mean, duty cycle and distribution) under ISS. Another approach of reconstruction method in the form of algorithms have been proposed in [16]–[18] to correct the estimation of these statistics under ISS.

The majority of the previous works have analysed the primary channel activity statistics without any constraints on the sample size used to estimate these statistics (i.e., an arbitrarily large sample size as large as required to achieve the best attainable estimation accuracy). Although the work in [19] has analysed the impact of the sample size on the statistics estimation, it was conducted under the assumption that spectrum sensing is perfect (i.e., PSS). Since DSA/CR receivers are more likely to operate under low SNR conditions where sensing errors are likely to occur, this work analyses the impact of the sample size on the estimation of the primary channel statistics under (a realistic) ISS scenario. In addition, it finds closed-form expressions for the required sample size of the idle/busy periods under ISS to achieve a targeted level of accuracy.

The rest of the paper is organised as follows. First, Section II introduces the system model. Then the sample size of the idle/busy periods under ISS is analysed in Section III. The estimation of the mean, duty cycle and distribution as a function of the sample size under ISS are analysed in Section IV, V and VI, respectively. The validation of the analytical results is shown in Section VII. Finally, Section VIII concludes the paper.

II. System Model

Consider a single PU occupying a particular primary channel. The occupancy patterns of such PU can be represented by a sequence of idle and busy time periods. The durations of these periods are random depending on the activity behavior of the PU. However, the experimental measurements in [20] have shown that these idle/buy periods are best represented by Generalised Pareto (GP) distribution. A SU, on the other hand, monitors the activity patterns of the PU based on spectrum sensing. In spectrum sensing, periodic sensing events are performed using sensing period T_s. At each sensing event the instantaneous state of the channel is detected and reported as idle ($\mathcal{H}_0$) or busy ($\mathcal{H}_1$). The outcomes of spectrum sensing, therefore, would result in a set of binary decisions, based on which the durations of idle/busy periods can be calculated. The time duration elapsed between any two changes in the state of the channel represents an estimation of the original duration as shown in Fig. 1(a), where T_i refers to the original idle/busy periods ($i = 0$ for idle and $i = 1$ for busy), and $\widehat{T}_i$ refers to the estimated idle/busy periods under PSS (i.e., without sensing errors). It can be noticed that the accuracy of such estimation (i.e., under PSS) is only affected by the time resolution imposed by the employed sensing period T_s. Since DSA/CR systems are more likely to experience low SNR conditions, sensing errors are likely to occur in the sensing decisions. Therefore, ISS is a more realistic scenario, where sensing errors can occur either within the idle periods as false alarms or within busy periods as missed detections, as shown in Fig. 1(b). Due to sensing errors, the estimated idle/busy periods under ISS $\breve{T}_i$ can be highly inaccurate (as shorter fragments) with respect to the original periods T_i. Sensing errors can be modelled as independent and identically distributed (i.i.d.) random variables and their probability can be given by probability of false alarm P_{fa} and probability of missed detection P_{md}.

III. Sample Size analysis under ISS

Let us consider as a set of idle/busy periods $\{T_{i,n}\}_{n=1}^{N}$ to represent N samples of the real activity and inactivity duration times of a PU within a particular channel, where

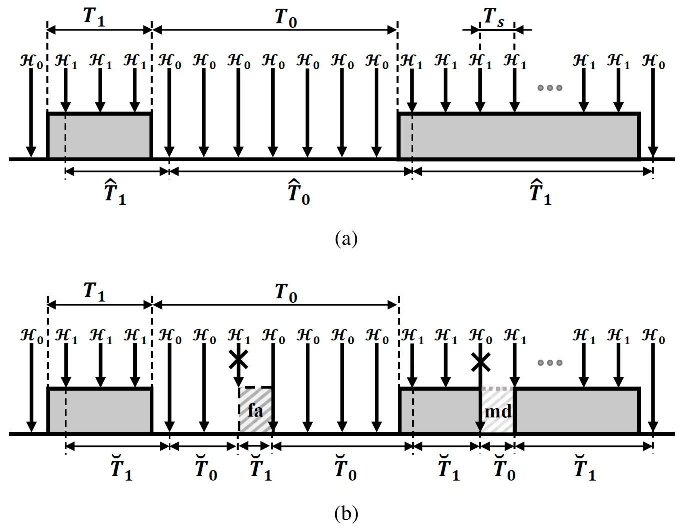

Fig. 1. Estimation of idle/busy periods based on spectrum sensing: (a) perfect spectrum sensing (PSS), (b) imperfect spectrum sensing (ISS).

i refers to the state of the channel, $i = 0$ for idle and $i = 1$ for busy. A SU on the other hand monitors the activity of the primary channel using spectrum sensing method. The outcomes of spectrum sensing can be used to calculate the duration of the idle/busy periods of the PU. Since spectrum sensing is not perfect in real-world operation as discussed in Section II, the calculated duration of the idle/busy periods can easily be corrupted by the presence of sensing errors. Sensing errors divide the original duration of the idle/busy periods into shorter fragments. The number of these fragments as a result is higher than the original number of periods N. In other words, the number of the observed idle/busy periods under ISS, N_{iss}, would be greater than (depending on the probability of spectrum sensing error) the original number of periods N (i.e., $N_{iss} > N$). These N_{iss} periods are short fragments of the original N periods. If probability of sensing error is zero, PSS can be assumed such that the original periods are observed without sensing errors and they are therefore not divided into fragments (i.e., $N_{pss} = N$). The only difference between the idle/busy periods observed under PSS and the original ones is the accuracy of calculating the duration of these periods, which depends on the resolution of the employed sensing period T_s. Since ISS is a more realistic scenario for the DSA/CR systems operating under low SNR conditions, it is very important to find a mathematical expression for the number of idle/busy periods observed under ISS, N_{iss}, as a function of the original number of periods N when probabilities of sensing errors, P_{fa} and P_{md}, are predefined by DSA/CR system.

To find the sample size N_{iss} as a function of the original sample size N, consider a single idle period T_0 and a single busy period T_1 which are observed under PSS as $\widehat{T}_0$ and $\widehat{T}_1$, respectively, based on K sensing events within each period (assuming they have same duration). If a single false alarm occurs within the K sensing events of the idle period and a single missed detection occurs within the K sensing events of the busy period as shown in Fig. 1(b), the total number of idle periods becomes three (the same applies to the busy periods as well). This is because the single false alarm divides the original idle period into two fragments of idle periods and the single missed detection produces another new idle period. As a result, the total number of idle periods becomes three (the two fragments plus the new one). Therefore, one can say:

$$N_{iss} = N + N_{fa} + N_{md} \tag{1}$$

where N_{fa} and N_{md} represent the number of false alarms and missed detections, respectively. They can be found by multiplying the number of sensing events by the probabilities of sensing errors as:

$$\begin{aligned} N_{fa} &= KP_{fa} = \frac{\sum_{n=1}^{N} \widehat{T}_{0,n}}{T_s} \cdot P_{fa} \\ &= \frac{N\mathbb{E}(\widehat{T}_0)}{T_s} P_{fa} = \frac{N\mathbb{E}(T_0)}{T_s} P_{fa} \end{aligned} \tag{2}$$

$$\begin{aligned} N_{md} &= KP_{md} = \frac{\sum_{n=1}^{N} \widehat{T}_{1,n}}{T_s} \cdot P_{md} \\ &= \frac{N\mathbb{E}(\widehat{T}_1)}{T_s} P_{md} = \frac{N\mathbb{E}(T_1)}{T_s} P_{md} \end{aligned} \tag{3}$$

Note that the estimated mean under PSS is equal to the true mean $\mathbb{E}(\widehat{T}_i) = \mathbb{E}(T_i)$ as discussed in [14].

Expression (1) is true if all false alarms and missed detecions occur as solo sensing errors within the idle/busy periods. This, however, is not the case as sensing errors can also appear attached to other periods or consecutive to other sensing errors. Therefore, we can correct (1) by taking the following two cases into consideration. Note that these two cases were discussed in [13] for a different purpose (to find an accurate mean estimator under ISS). However, here we discuss the impact of these two cases on the sample size.

- **Case I:** As shown in Fig. 2, when a sensing error occurs at the first (or last) sensing event within a period, it will be observed as a part of the next period attached to it, thus causing no fragments or additional periods. Therefore, these sensing errors should not be counted in expression (1). One can subtract 2 (the first and the last sensing events) from the total number of sensing events observed within a period when calculating N_{fa} and N_{md} in order to consider only the sensing errors which cause additional periods.
- **Case II:** As shown in Fig. 3, when a sensing error occurs consecutively to another one, it will have the same effect of a single sensing error in terms of the resulting number of fragments. Therefore, consecutive errors should be counted as single error in expression (1) (i.e., only the first sensing error in a burst should be counted while the remaining consecutive errors should not). In order to consider this in the calculation of N_{fa} and N_{md} the probability of having consecutive sensing errors should be subtracted from the original probability of sensing error as:

$$\grave{P}_{fa} = P_{fa} - \sum_{j=2}^{\infty} P_{fa}^{j} = P_{fa}\left(\frac{1-2P_{fa}}{1-P_{fa}}\right) \tag{4}$$

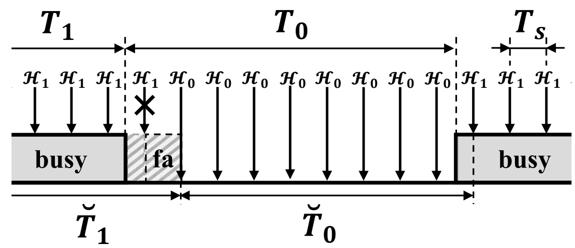

Fig. 2. Case I: A single sensing error at the edge of a period [13].

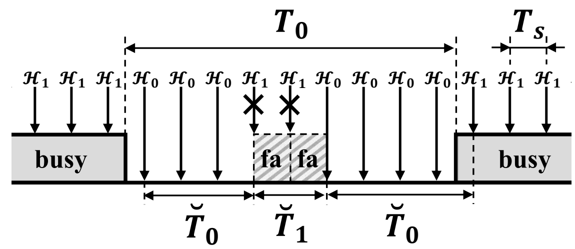

Fig. 3. Case II: Two consecutive sensing errors in the middle of a period [13].

$$\grave{P}_{md} = P_{md} - \sum_{j=2}^{\infty} P_{md}^{j} = P_{md}\left(\frac{1-2P_{md}}{1-P_{md}}\right) \tag{5}$$

From these two cases, the actual number of false alarms and missed detections which cause fragments or additional periods can be written with reference to (2) and (3) as:

$$\grave{N}_{fa} = N\left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} \tag{6}$$

$$\grave{N}_{md} = N\left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md} \tag{7}$$

The final expression for the sample size under ISS, as a function of the original sample size and probabilities of sensing errors is:

$$N_{iss} = N\left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}\right) \tag{8}$$

Similarly, we can also find the corresponding original sample size based on the observed sample size under ISS as:

$$N = \frac{N_{iss}}{\left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}\right)} \tag{9}$$

IV. Required Sample Size for the Mean Estimation under ISS

Given a set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$ of N idle/busy periods observed under PSS, the mean $\mathbb{E}(\widehat{T}_i)$ of the observed periods can be found based on the sample mean estimator $\widehat{m}_i$:

$$\mathbb{E}(\widehat{T}_i) \approx \widehat{m}_i = \frac{1}{N}\sum_{n=1}^{N}\widehat{T}_{i,n} \tag{10}$$

The maximum relative error of the mean estimator $\widehat{m}_i$ under PSS as a function of the sample size N is found as [19]:

$$\varepsilon_{max}^{\widehat{m}_i} \approx \frac{\kappa}{\mathbb{E}(T_i)} \left[\frac{1}{N} \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) \right]^{\frac{1}{2}} \tag{11}$$

where $\mathbb{V}(T_i)$ denotes the variance of the idle/busy periods, and κ is standard deviation interval defined by the confidence level ρ. For a given confidence level ρ, κ can be found from concentration inequalities as explained in [19]. However, concentration inequalities may lead to loose upper bounds of the maximum relative error. A more accurate result can be achieved by applying the central limit theorem on the mean estimator $\widehat{m}_i$ where κ can be obtained for a certain confidence level ρ as [19]:

$$\kappa \geq \sqrt{2}\ \mathrm{erf}^{-1}(\rho) \tag{12}$$

In order to find the maximum relative error of the mean estimator $\breve{m}_i$ under ISS as a function of the sample size N_{iss}, we use the obtained expression in (9), which finds the original sample size as a function of the ISS sample size, along with (11) as:

$$\varepsilon_{max}^{\breve{m}_i} \approx \frac{\kappa}{\mathbb{E}(T_i)} \left[\frac{\left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right) \grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right) \grave{P}_{md}\right)}{N_{iss}} \times \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) \right]^{\frac{1}{2}} \tag{13}$$

Note that if probabilities of sensing errors (P_{fa} and P_{md}) are zero, (13) will be the same as (11) for the calculated relative error under PSS. Therefore, the result shown in (13) can be considered as a general form to find the maximum relative error of the mean estimator based on the sample size under both scenarios.

The required sample size of the idle/busy periods observed under ISS to achieve a targeted maximum relative error of the mean estimator can be found from (13) as:

$$N_{iss}^{\breve{m}_i} \approx \left(\frac{\kappa}{\varepsilon_{max}^{\breve{m}_i} \mathbb{E}(T_i)} \right)^2 \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) \times \left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right) \grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right) \grave{P}_{md}\right) \tag{14}$$

V. Required Sample Size for the Duty Cycle Estimation under ISS

The duty cycle of the idle/busy periods of the primary channel is one of the most important statistical information for DSA/CR systems, which helps determine the amount of opportunities available in the primary channels. The channel duty cycle Ψ can be estimated based on the observed sample size under PSS as:

$$\widehat{\Psi} = \frac{\widehat{m}_1}{\widehat{m}_1 + \widehat{m}_0} \tag{15}$$

where $\widehat{m}_1$ and $\widehat{m}_0$ are the sample mean of the busy and idle periods, respectively. The maximum relative error of the duty cycle estimator $\widehat{\Psi}$ under PSS as a function of the sample size N is found as [19]:

$$\varepsilon_{max}^{\widehat{\Psi}} \approx \frac{\kappa}{\Psi} \left[\frac{1}{N\left[\mathbb{E}(T_0) + \mathbb{E}(T_1)\right]^4} \times \left\{ [\mathbb{E}(T_1)]^2 \left(\mathbb{V}(T_0) + \frac{T_s^2}{6} \right) + [\mathbb{E}(T_0)]^2 \left(\mathbb{V}(T_1) + \frac{T_s^2}{6} \right) \right\} \right]^{\frac{1}{2}} \tag{16}$$

In order to find the maximum relative error of the duty cycle estimator $\breve{\Psi}$ under ISS as a function of the sample size N_{iss}, we use the obtained expression in (9), which finds the original sample size as a function of the ISS sample size, along with (16) as:

$$\varepsilon_{max}^{\breve{\Psi}} \approx \frac{\kappa}{\Psi} \left[\frac{\left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right) \grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right) \grave{P}_{md}\right)}{N_{iss}\left[\mathbb{E}(T_0) + \mathbb{E}(T_1)\right]^4} \times \left\{ [\mathbb{E}(T_1)]^2 \left(\mathbb{V}(T_0) + \frac{T_s^2}{6} \right) + [\mathbb{E}(T_0)]^2 \left(\mathbb{V}(T_1) + \frac{T_s^2}{6} \right) \right\} \right]^{\frac{1}{2}} \tag{17}$$

The required sample size of the idle/busy periods observed under ISS to achieve a targeted maximum relative error of the duty cycle estimator can be found from (17) as:

$$N_{iss}^{\breve{\Psi}} \approx \left(\frac{\kappa}{\varepsilon_{max}^{\breve{\Psi}} \Psi} \right)^2 \times \frac{\Psi^2 \left(\mathbb{V}(T_0) + \frac{T_s^2}{6} \right) + (1-\Psi)^2 \left(\mathbb{V}(T_1) + \frac{T_s^2}{6} \right)}{\left[\mathbb{E}(T_0) + \mathbb{E}(T_1)\right]^2} \times \left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right) \grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right) \grave{P}_{md}\right) \tag{18}$$

VI. Required Sample Size for the Distribution Estimation under ISS

The distribution of the idle/busy periods can also be estimated based on the outcomes of spectrum sensing. Since the experimental measurements in [20] have shown that Generalised Pareto (GP) distribution is the best description for the original idle/busy periods of a PU, the CDF of these periods can be estimated based on the PSS observation as [14]:

$$F_{\widehat{T}_i}(T) = 1 - \left[1 + \frac{\widehat{\alpha}_i (T - \widehat{\mu}_i)}{\widehat{\lambda}_i} \right]^{-1/\widehat{\alpha}_i}, \quad T \geq \widehat{\mu}_i \tag{19}$$

where $\widehat{\mu}_i$, $\widehat{\lambda}_i$ and $\widehat{\alpha}_i$ are the location, scale and shape parameters of the GP distribution. The location $\widehat{\mu}_i$ also represents the minimum period duration and it can be assumed to be known $\widehat{\mu}_i \approx \mu_i$, while $\widehat{\lambda}_i$ and $\widehat{\alpha}_i$ can be found as [14]:

$$\widehat{\lambda}_i = \frac{1}{2} \left(1 + \frac{(\widehat{m}_i - \widehat{\mu}_i)^2}{\widetilde{v}_i} \right) (\widehat{m}_i - \widehat{\mu}_i) \tag{20a}$$

$$\widehat{\alpha}_i = \frac{1}{2} \left(1 + \frac{(\widehat{m}_i - \widehat{\mu}_i)^2}{\widetilde{v}_i} \right) \tag{20b}$$

where $\widehat{m}_i$ and $\widetilde{v}_i$ are the sample mean and variance estimators, respectively. Note that the variance estimator is found in [14] as $\widetilde{v}_i = \widehat{v}_i - \frac{T_s^2}{6}$. The accuracy of the CDF estimator $F_{\widehat{T}_i}(T)$ in (19) as a function of the sample size N under PSS is found in terms of the Kolmogorov-Smirnov (KS) distance as [19]:

$$D_{KS}^{F_{\widehat{T}}} = \kappa(1+\alpha_i)^{-\frac{1}{\alpha_i}-1} \times \left[\frac{1}{N}\left(\frac{1}{\lambda_i^2}\Omega(T_i) + \frac{[(1+\alpha_i)\ln(1+\alpha_i)-\alpha_i]^2}{\alpha_i^4}\Upsilon(T_i)\right)\right]^{\frac{1}{2}} \quad (21)$$

where $\Omega(T_i)$ and $\Upsilon(T_i)$ are given in [19, eq. (45)].

In order to find the KS distance of the CDF estimator $F_{\check{T}_i}(T)$ under ISS as a function of the sample size N_{iss}, we use the obtained expression in (9), which finds the original sample size as a function of the ISS sample size, along with (21) as:

$$D_{KS}^{F_{\check{T}}} = \kappa(1+\alpha_i)^{-\frac{1}{\alpha_i}-1} \times \left[\frac{\left(1+\left(\frac{\mathbb{E}(T_0)}{T_s}-2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s}-2\right)\grave{P}_{md}\right)}{N_{iss}} \times \left(\frac{1}{\lambda_i^2}\Omega(T_i) + \frac{[(1+\alpha_i)\ln(1+\alpha_i)-\alpha_i]^2}{\alpha_i^4}\Upsilon(T_i)\right)\right]^{\frac{1}{2}} \quad (22)$$

The required sample size of the idle/busy periods observed under ISS to achieve a targeted KS distance of the CDF estimator can be found from (22) as:

$$N_{iss}^{F_{\check{T}}} = \left(\frac{\kappa}{D_{KS}^{F_{\check{T}}}}\right)^2 (1+\alpha_i)^{-\frac{2}{\alpha_i}-2} \times \left(1+\left(\frac{\mathbb{E}(T_0)}{T_s}-2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s}-2\right)\grave{P}_{md}\right) \times \left(\frac{1}{\lambda_i^2}\Omega(T_i) + \frac{[(1+\alpha_i)\ln(1+\alpha_i)-\alpha_i]^2}{\alpha_i^4}\Upsilon(T_i)\right) \quad (23)$$

VII. Numerical, Simulation and Experimental Results

In this section we validate the obtained analytical results by means of simulations and experiments. First of all, the obtained expression in (9) is tested, which finds the sample size that would be observed under ISS for a given original sample size of the idle/busy periods. In simulations, we generate $N = 10^4$ samples of the idle/busy periods (drawn from the GP distribution). These periods can then be observed using spectrum sensing method with a sensing period of T_s t.u. (time units). Applying sensing errors on the sensing decisions (based on the predefined probability of sensing error) results in ISS observations for the generated idle/busy periods. Therefore, the sample size of the ISS observations can be calculated in the simulations for the $N = 10^4$ samples using different probabilities of sensing error ($P_{fa}, P_{md} \in \{0.001, 0.01, 0.1\}$) and compared with the theoretical one obtained by (9). As shown in Fig. 4, the analytical results match the simulation results for the calculated sample size under ISS with respect to T_s. It can also be noticed that the number of observed periods under ISS increases as the probabilities of error increase because the presence of more sensing errors produces a larger number of shorter fragments of the original periods.

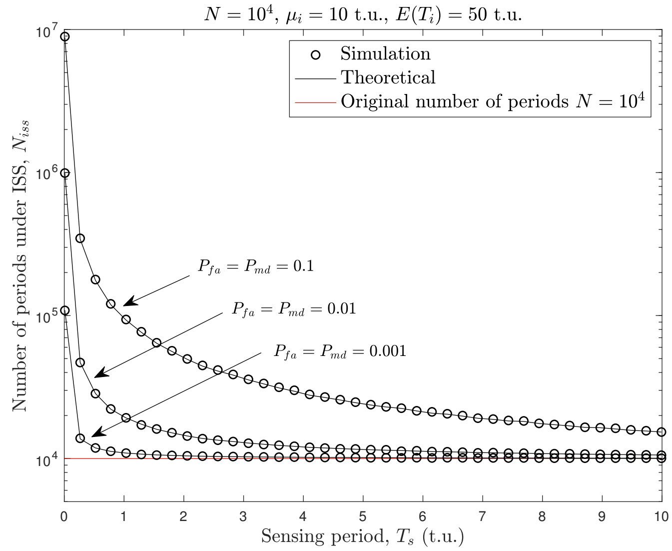

Fig. 4. The sample size under ISS as a function of the sensing period T_s, when the original sample size $N = 10^4$.

After validating (9), we can validate the obtained maximum relative error expressions (13) and (17) for the mean and variance estimators, respectively, and the KS distance expression (22) for the CDF estimator. In simulation, we estimate the mean, variance, and distribution of the periods (using the proposed estimators under ISS in [15]). Then the maximum relative error of these estimators (using confidence level $\rho = 0.95$) can be calculated with respect to the sample size of the ISS observations and compared with the analytical expressions. These analyses are also validated experimentally using a hardware Prototype for the Estimation of Channel Activity Statistics (PECAS), where a full description for the implementation of such prototype can be found in [21]. As shown in Fig. 5, the relative errors of the mean and variance estimators as well as the KS distance of the CDF estimator decrease as the sample size of the ISS increases. It can also be noted that the analytical results obtained in this work reproduce accurately the sample sizes required to achieve the desired estimation accuracies and can therefore be useful in DSA/CR system designs under any scenario of spectrum sensing (especially the realistic ISS scenario).

VIII. Conclusion

The significance of the primary channel activity statistics on the performance of the DSA/CR systems has led many research campaigns to study these statistics and to investigate different methods to accurately estimate such statistics from spectrum sensing observations. However, these studies were conducted without any constraints on the sample size used

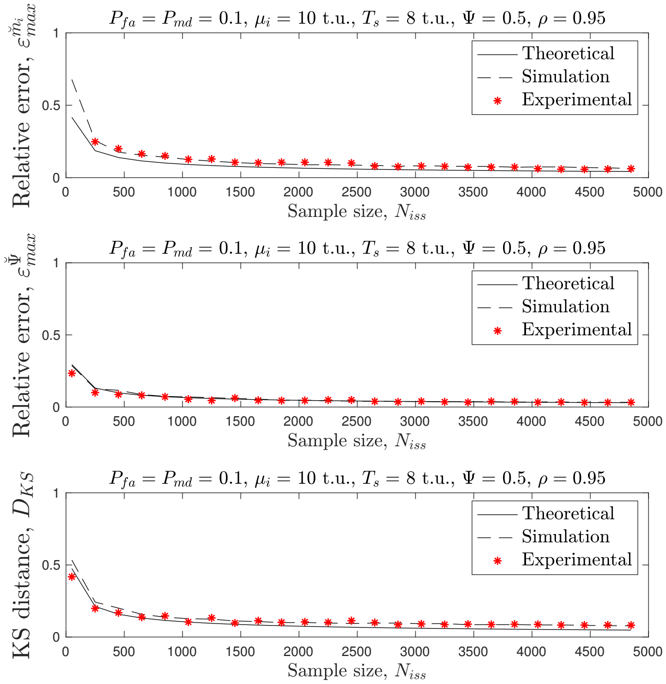

Fig. 5. (Top) Relative error of the mean estimator, (Middle) Relative error of the duty cycle estimator, and (Bottom) KS distance of the CDF estimator with respect to the sample size under ISS, when confidence level $\rho = 0.95$.

to estimate primary channel statistics (i.e., an arbitrarily large sample size as large as required to achieve the best attainable estimation accuracy), ignoring the fact that DSA/CR systems are constrained by the practical limitations of the hardware design. In this context, this work has analysed the impact of the sample size of the ISS observations on the estimation accuracy of the primary channel activity statistics, which, to the best of the authors' knowledge, has not been investigated in the literature. In addition, it has found closed-form expressions for the required sample size of the ISS observations to achieve a desired level of accuracy. The outcomes of this work will help DSA/CR system designs to select the minimum required sample size that can guarantee achieving a predefined level of accuracy for the estimation of the primary channel activity statistics under ISS.

Acknowledgment

This work was supported by the British Council under UKIERI DST Thematic Partnerships 2016-17 (ref. DST-198/2017).

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. López-Benítez, "Cognitive radio," in *Heterogeneous cellular networks: Theory, simulation and deployment.* Cambridge University Press, 2013, ch. 13.

[3] X. Liu, B. Krishnamachari, and H. Liu, "Channel selection in multi-channel 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.

[4] S. Sengottuvelan, J. Ansari, P. Mähönen, T. G. Venkatesh and M. Petrova, "Channel Selection Algorithm for Cognitive Radio Networks with Heavy-Tailed Idle Times," in *IEEE Transactions on Mobile Computing*, vol. 16, no. 5, pp. 1258–1271, 1 May 2017.

[5] M. B. Hosen, M. M. H. Mridha and M. A. Hamza, "Secondary User Channel Selection in Cognitive Radio Network Using Adaptive Method," *2018 3rd International Conference for Convergence in Technology (I2CT)*, Pune, 2018, pp. 1–6.

[6] Y. Chen and H.-S. Oh, "Spectrum measurement modelling and prediction based on wavelets," *IET Communications*, vol. 10, no. 16, pp. 2192–2198, Oct 2016.

[7] G. Ding *et al.*, "Spectrum Inference in Cognitive Radio Networks: Algorithms and Applications," in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 150–182, Firstquarter 2018.

[8] W. Zhang, C. Wang, X. Ge and Y. Chen, "Enhanced 5G Cognitive Radio Networks Based on Spectrum Sharing and Spectrum Aggregation," in *IEEE Trans. on Comms.*, vol. 66, no. 12, pp. 6304–6316, Dec. 2018.

[9] J. Lehtomäki, M. López-Benítez, K. Umebayashi and M. Juntti, "Improved Channel Occupancy Rate Estimation," in *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 643–654, March 2015.

[10] M. López-Benítez and J. Lehtomäki, "Energy detection based estimation of primary channel occupancy rate in cognitive radio," *2016 IEEE Wireless Communications and Networking Conference*, Doha, 2016, pp. 1–6.

[11] J. J. Lehtomaki, R. Vuohtoniemi and K. Umebayashi, "On the Measurement of Duty Cycle and Channel Occupancy Rate," in *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2555–2565, November 2013.

[12] W. Gabran, C. H. Liu, P. Pawelczak, and D. Cabric, "Primary user traffic estimation for dynamic spectrum access," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 544–558, March 2013.

[13] O. H. Toma, M. López-Benítez, D. K. Patel and K. Umebayashi, "Primary Channel Duty Cycle Estimation under Imperfect Spectrum Sensing Based on Mean Channel Periods," *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1–6.

[14] M. López-Benítez, A. Al-Tahmeesschi, D. Patel, J. Lehtomäki and K. Umebayashi, "Estimation of Primary Channel Activity Statistics in Cognitive Radio Based on Periodic Spectrum Sensing Observations," in *IEEE Trans. on Wirel. Comms.*, vol. 18, no. 2, pp. 983–996, Feb. 2019.

[15] O. H. Toma, M. López-Benítez, D. K. Patel and K. Umebayashi, "Estimation of Primary Channel Activity Statistics in Cognitive Radio Based on Imperfect Spectrum Sensing," in *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2016–2031, April 2020.

[16] A. Al-Tahmeesschi, M. López-Benítez, J. Lehtomäki and K. Umebayashi, "Improving primary statistics prediction under imperfect spectrum sensing," *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Barcelona, 2018, pp. 1–6.

[17] O. H. Toma, M. López-Benítez, D. K. Patel and K. Umebayashi, "Reconstruction Algorithm for Primary Channel Statistics Estimation Under Imperfect Spectrum Sensing," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, South Korea, 2020, pp. 1–5.

[18] M. López-Benítez, "Can primary activity statistics in cognitive radio be estimated under imperfect spectrum sensing?" in *Proc. 24th Annual IEEE Intl. Symp. Pers., Indoor and Mobile Radio Comms. (PIMRC 2013)*, Sep. 2013, pp. 750–755.

[19] A. Al-Tahmeesschi, M. López-Benítez, D. K. Patel, J. Lehtomäki and K. Umebayashi, "On the Sample Size for the Estimation of Primary Activity Statistics Based on Spectrum Sensing," in *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 1, pp. 59–72, March 2019.

[20] M. López-Benítez, and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE Trans. Veh. Tech.*, vol. 62, no. 5, pp. 2091–2104, Jun 2013.

[21] M. López-Benítez, A. Al-Tahmeesschi, K. Umebayashi and J. Lehtomäki, "PECAS: A low-cost prototype for the estimation of channel activity statistics in cognitive radio," in *Proc. IEEE Wirel. Comms. and Net. Conf. (WCNC 2017)*, 2017, pp. 1–6.