

On the Sample Size for the Estimation of Primary Activity Statistics Based on Spectrum Sensing

Ahmed Al-Tahmeesschi, *Student Member, IEEE*, Miguel López-Benítez, *Senior Member, IEEE*, Dhaval Patel, *Member, IEEE*, Janne Lehtomäki, *Member, IEEE*, and Kenta Umebayashi, *Member, IEEE*

Abstract—Dynamic Spectrum Access (DSA) / Cognitive Radio (CR) systems can benefit from the knowledge of the activity statistics of primary channels, which can use this information to intelligently adapt their spectrum use to the operating environment. Particularly relevant statistics are the minimum, mean and variance of the on/off period durations, the channel duty cycle and the governing distribution. However, most DSA/CR systems have limited resources (power consumption, memory capacity, computational capability) and an important question arises of how many on/off period observations are required (i.e., the number of observed on/off periods, referred to as *observation sample size* in this work) to estimate the statistics of the primary channel to a certain desired level of accuracy. In this work, closed-form expressions to link such sample size with the accuracy of the observed primary activity statistics are proposed. A comprehensive theoretical analysis is performed on the required number of observed on/off periods to obtain a specific estimation accuracy. The accuracy of the obtained analytical results is validated and corroborated with both simulation and experimental results, showing a perfect agreement. The analytical results derived in this work can be used in the design and dimensioning of DSA/CR systems in which the spectrum awareness function relies on spectrum sensing.

Keywords— *Cognitive radio, spectrum sensing, spectrum awareness, primary activity statistics, sample size.*

I. INTRODUCTION

Dynamic Spectrum Access (DSA) principle [1–3], relying on the Cognitive Radio (CR) paradigm, has been proposed as a promising solution to improve spectral efficiency [4]. A CR is designed for autonomous reconfiguration of the transmitter parameters based on changing radio environment [5]. CRs can increase spectrum efficiency by allowing unlicensed/secondary users (SU) to opportunistically access licensed/primary user (PU) bands. By sensing the primary channel periodically, DSA/CR users can identify the idle/busy times of the channel and transmit with minimal, non-harmful interference to PUs.

While the main role of spectrum sensing is to determine the instantaneous on/off state of the primary channel, spectrum

sensing decisions can also be exploited to estimate the duration of past busy/idle periods and hence primary activity statistics such as the minimum, mean and variance of the on/off period durations, the channel duty cycle and the underlying distribution. This information can in turn be employed to access the spectrum more effectively and improve the CR system performance [6, 7] by selecting the most appropriate channel for transmission [8], reducing the switching time delay [9, 10], selecting an appropriate threshold for energy detection [11], forecasting the primary occupancy pattern to minimise the interference [12, 13], and thus increase the overall spectrum efficiency. This information could be obtained from other alternatives such as databases. However, the sensing-based approach has significant advantages including lower cost and complexity, independence of external systems and better suitability for highly dynamic radio environments [14].

In the literature, several studies have considered the problem of estimating primary activity statistics based on sensing decisions, mostly focusing on the estimation of the channel duty cycle. In [15], the estimation of the channel occupancy rate (duty cycle) based on different approaches was studied analytically in the presence of sensing errors. A mathematical analysis on the estimation of the mean on/off durations as well as the duty cycle under DSA was presented in [16] and references therein. An analytical study on the estimation of the distribution of busy/idle periods based on spectrum sensing was presented in [17], while several methods for the classification of such distribution were proposed in [1] under the assumption of no sensing errors. To overcome the degrading effect of sensing errors on the estimated primary activity statistics, several algorithms were proposed in [18, 19].

The immense majority of previous work relies on a *sample-based* approach where primary activity statistics are estimated based on individual sensing decisions. This approach can be used to estimate the channel duty cycle (typically calculated as the quotient between the number of busy sensing decisions and the total number of sensing decisions), which has been the parameter of interest in most previous work. In contrast, this work considers a broader range of relevant primary activity statistics that includes not only the channel duty cycle but also the minimum on/off periods, the mean and variance of on/off periods and the underlying distributions. These statistics cannot be estimated based on the sample-based approach (at least not in an evident manner), which motivates this work to consider a *period-based* approach where the individual on/off sensing decisions are first processed to estimate the durations of each on/off period observed in the channel. Once

Manuscript received March 11, 2018. M. López-Benítez and D. Patel would like to thank the financial support received from UKIERI under the DST Thematic Partnerships 2016-17 (grant ref. DST-198/2017). K. Umebayashi would like to thank the supports received from the MIC/SCOPE #165003006, and JSPS KAKENHI Grant Numbers JP15K06053, JP15KK0200.

Al-Tahmeesschi and M. López-Benítez are with the Department of Electrical Engineering and Electronics, University of Liverpool, United Kingdom (email: {Ahmed.Al-Tahmeesschi, M.Lopez-Benitez}@liverpool.ac.uk).

D. Patel is with the School of Engineering and Applied Science, Ahmedabad University, India (email: dhaval.patel@ahduni.edu.in).

J. Lehtomäki is with the Centre for Wireless Communications, University of Oulu, Finland (email: jannel@ee.oulu.fi).

K. Umebayashi is with the Graduate School of Engineering, Tokyo University of Agriculture and Technology, Japan (email: ume_k@cc.tuat.ac.jp).

a sufficient number of periods have been observed, the whole set of activity statistics (not only the duty cycle) is estimated. The main advantage of the period-based approach considered in this work is the broader range of activity statistics that can be estimated compared to the sample-based approach.

An important practical question in this context is how many busy/idle (on/off) periods need to be observed by a DSA/CR system (i.e., the observation sample size) in order to estimate the primary channel activity statistics (in this work: minimum, mean and variance of on/off times, channel duty cycle and the underlying distribution) to a certain desired level of accuracy. To the best of the authors' knowledge, the only work that has considered this problem is the study presented in [20], where only the channel duty cycle is considered (estimated based on individual on/off sensing decisions) and the required number of individual sensing events is analysed. On the other hand, this work considers a more general approach where primary statistics are estimated based on a set of observed busy/idle time durations (instead of individual sensing decisions), which can be used to estimate a broader range of primary activity parameters (not only the channel duty cycle).

This work presents a comprehensive analytical study on the estimation of primary channel statistics based on spectrum sensing decisions and determines the relation between the number of observed on/off periods and the accuracy of the resulting primary activity statistics. To the best of authors' knowledge, this problem has not received a rigorous treatment before in the literature. The accuracy of the derived closed-form expressions is validated and corroborated with simulation results as well as experimental results from a hardware prototype developed to this end.

The contributions of this work are outlined as follows:

- 1) Closed-form expressions are derived for the maximum error of the minimum, mean and variance of the estimated on/off periods, the channel duty cycle and the distribution as a function of the observation sample size (and other relevant parameters such as the sensing period). These analytical results are useful to determine the number of observed periods/samples required to guarantee a desired maximum error in the estimated statistics. Expressions for the required observation sample sizes are derived as well.
- 2) Practical validation of the obtained analytical results, not only with simulations but also with experiments, is provided using a hardware prototype specifically designed to replicate inexpensive low-end DSA/CR devices. This scenario is closer to real-life DSA/CR scenarios than using advanced and costly laboratory equipment.
- 3) A comprehensive study about the effect of the observation sample size on the estimation of primary activity statistics is carried out considering a practical primary activity model based on the Generalised Pareto distribution, which is proven to be a more accurate model compared to the commonly used exponential distribution models [21].

The remainder of this paper is organised as follows. First, Section II presents the considered system model and formulates the problem addressed in this work. Subsequent sections provide a detailed mathematical analysis on the sample size

TABLE I
MATHEMATICAL NOTATION USED IN THIS WORK.

Parameter	Definition
N (N_x)	Sample size (for parameter x)
T_s	Spectrum sensing period
T_e	Error in the estimation of a period
T_i / \hat{T}_i	True/estimated period
μ_i / $\hat{\mu}_i$	True/estimated minimum period
Ψ / $\hat{\Psi}$	True/estimated channel duty cycle
$f_{T_i}(\cdot)$ / $f_{\hat{T}_i}(\cdot)$	PDF of true/estimated periods
$F_{T_i}(\cdot)$ / $F_{\hat{T}_i}(\cdot)$	CDF of true/estimated periods
λ / $\hat{\lambda}$	True/estimated scale parameter
α / $\hat{\alpha}$	True/estimated shape parameter
$P(x)$	Probability of x
$\mathbb{E}(x)$	Expected value (mean) of x
$\mathbb{V}(x)$	Variance of x
$\mathbb{M}_3(x)$	Third central moment of x
$\mathbb{M}_4(x)$	Fourth central moment of x
$\mathbb{C}(x, y)$	Covariance of x and y
\hat{m}_i	Sample mean of estimated periods
\hat{v}_i	Sample variance of estimated periods
\tilde{v}_i	Corrected version of \hat{v}_i
$\mathcal{E}_{r,\max}^x$	Maximum relative error of parameter x
D_{KS}	Kolmogorov-Smirnov distance
$U(a, b)$	Uniform distribution within $[a, b]$
ξ	Bernoulli random variable
ρ	Confidence level

Sub-indices: $i = 0$ for idle periods, $i = 1$ for busy periods.

required to provide an arbitrarily accurate estimation of the minimum period (Section III), the mean and variance of the observed periods (Section IV), the channel duty cycle (Section V) and the underlying distribution of the observed periods (Section VI). The obtained analytical results depend on the real/actual parameters of the PU traffic, which are unknown to the SU; this problem is overcome with an iterative stopping algorithm that enables an accurate estimation of the required sample size in practical implementations, which is presented in Section VII. The simulation and experimental methodology employed in this work to validate the correctness and accuracy of the analytical results is described in Section VIII, while the obtained numerical results are presented in Section IX. Finally, Section XI summarises and concludes this work. Table I summarises the mathematical notation used in this work.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this work, without loss of generality, a single SU is considered to detect PU activity on a single frequency band and channel through spectrum sensing decisions. The SU senses the channel with a certain periodicity of T_s time units (t.u.), which is referred to as *sensing period*. The sensing decisions are introduced as a binary alternating state: busy (H_1) when the PU signal is present and idle (H_0) when the PU signal is absent. The computed elapsed time (at SU) between two PU state changes is considered as an estimation \hat{T}_i of the real period duration T_i ($i = 0$ for idle periods and $i = 1$ for

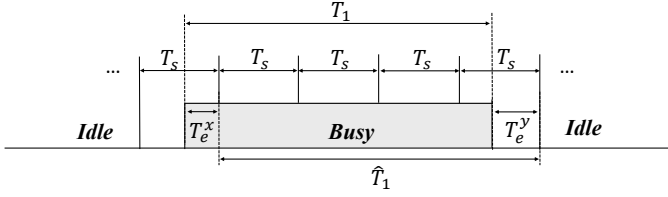


Fig. 1. Considered model. T_s , T_i and \hat{T}_i represent the sensing period, original period duration and estimated period duration, respectively. T_e^x and T_e^y are the errors in the estimated period.

busy periods) as illustrated in Fig. 1, where the estimation of the duration of a busy period is shown (idle periods can be estimated using the same method). Following the approach shown in Fig. 1, the SU estimates a set $\{\hat{T}_{i,n}\}_{n=1}^N$ of N observed periods of the same type, which is used to calculate the primary activity statistics of interest. A wide range of statistics is considered in this work, including the minimum, mean and variance of the observed periods \hat{T}_i , the PU channel duty cycle, and the underlying distribution (this includes the activity statistics most commonly used in the DSA/CR literature).

The main objective of this work is to explore the relation between the number of observed periods N (the observation sample size) and the accuracy of the estimated primary activity statistics mentioned above. To this end, explicit closed-form expressions are derived for the estimation error of each primary activity statistic as a function of the sample size N (and other involved parameters). This is essential to enable DSA/CR systems determine whether the estimated statistics of the primary user activity are sufficiently accurate, or more observations are required instead.

III. ESTIMATION OF THE MINIMUM PERIOD

The estimation of the minimum primary activity time (both busy/on and idle/off times) has a great deal of importance when it comes to spectrum sensing and DSA/CR in general [17, 22], as it determines the minimum amount of time that a DSA/CR will have to wait before the primary channel is available (minimum busy time) and the minimum amount of time it will be available for transmission (minimum idle time). In some cases, the true minimum period is known, for example in the case of primary systems that use some form of regional beacon signals with real-time information [14] or when the radio technology of the primary system is standardised (e.g., the slot duration of GSM). Otherwise, it needs to be estimated.

Based on a set $\{\hat{T}_{i,n}\}_{n=1}^N$ of N observed periods \hat{T}_i estimated as shown in Fig. 1, the minimum period duration or minimum primary (in)activity time, denoted as $\hat{\mu}_i$, can be obtained as:

$$\hat{\mu}_i = \min_n \left(\{\hat{T}_{i,n}\}_{n=1}^N \right) \quad (1)$$

Note that the periods \hat{T}_i estimated as shown in Fig. 1 are integer multiples of the sensing period T_s , however the same original period T_i can lead to two possible estimated periods, either $\hat{T}_i = \lfloor T_i/T_s \rfloor T_s = kT_s$, $k \in \mathbb{N}^+$ (where $\lfloor \cdot \rfloor$ is the floor operator) or $\hat{T}_i = \lceil T_i/T_s \rceil T_s = (k+1)T_s$, $k \in \mathbb{N}^+$ (where $\lceil \cdot \rceil$ is the ceil operator). The actual estimated period depends on the relative (random) position of the sensing events with respect

to the beginning/end of the original period T_i . Based on this observation, the estimated periods can be modelled as:

$$\hat{T}_i = \left(\left\lfloor \frac{T_i}{T_s} \right\rfloor + \xi \right) T_s \quad (2)$$

where $\xi \in \{0, 1\}$ is a Bernoulli random variable. Introducing (2) into (1), and noting that $\min(T_i) = \mu_i$ and $\min(\xi) = 0$, it can be seen that the estimated minimum period is given by:

$$\hat{\mu}_i = \min(\hat{T}_i) = \min \left[\left(\left\lfloor \frac{T_i}{T_s} \right\rfloor + \xi \right) T_s \right] = \left\lfloor \frac{\mu_i}{T_s} \right\rfloor T_s \quad (3)$$

The main question that this work aims to answer is how many primary periods N need to be observed in order to estimate each primary activity statistic to a certain degree of accuracy. From (3) it can be observed that, in the particular case of the estimation of the minimum period duration, the estimation error is mainly given by the employed sensing period T_s and increasing the sample size N will not improve the accuracy of the estimated minimum $\hat{\mu}_i$. However, if the observation sample size N is not sufficiently large, a longer period $\hat{T}_i > \hat{\mu}_i = \lfloor \mu_i/T_s \rfloor T_s$ might be selected as the minimum observed period, thus potentially leading to a less accurate estimation. Therefore, in the particular case of this section, the relevant question is how many primary periods N need to be observed to ensure that the estimated minimum period is the most accurate possible estimation, in other words, ensure that at least one instance of the period $\hat{\mu}_i = \lfloor \mu_i/T_s \rfloor T_s$ is observed in the set $\{\hat{T}_{i,n}\}_{n=1}^N$.

To answer this question, let first determine the probability that an observed period \hat{T}_i is equal to the best possible estimation of the minimum period given by (3), i.e., $\hat{\mu}_i = \lfloor \mu_i/T_s \rfloor T_s$. A period with duration $\hat{T}_i = \hat{\mu}_i = \lfloor \mu_i/T_s \rfloor T_s$ will be observed if $T_i \in [\lfloor \mu_i/T_s \rfloor T_s, \lceil \mu_i/T_s \rceil T_s]$ and $\xi = 0$, therefore the probability to observe one instance of $\hat{\mu}_i$ can be obtained as:

$$\begin{aligned} P(\hat{T}_i = \hat{\mu}_i) &= P\left(\left\lfloor \frac{\mu_i}{T_s} \right\rfloor T_s \leq T_i \leq \left\lceil \frac{\mu_i}{T_s} \right\rceil T_s\right) \mathbb{E}(P(\xi=0)) \\ &= P\left(\mu_i \leq T_i \leq \left\lceil \frac{\mu_i}{T_s} \right\rceil T_s\right) \mathbb{E}(P(\xi=0)) \chi_0 \\ &= \left[F_{T_i}\left(\left\lceil \frac{\mu_i}{T_s} \right\rceil T_s\right) - F_{T_i}(\mu_i)\right] \mathbb{E}(P(\xi=0)) \chi_0 \\ &= F_{T_i}\left(\left\lceil \frac{\mu_i}{T_s} \right\rceil T_s\right) \mathbb{E}(P(\xi=0)) \chi_0 \end{aligned} \quad (4)$$

where $F_{T_i}(\cdot)$ is the Cumulative Distribution Function (CDF) of the original periods T_i and $F_{T_i}(\mu_i) = 0$, $P(\xi=0)$ is the probability that $\xi=0$ in the model of (2) which can be calculated as:

$$P(\xi=0) = P\left(\hat{T}_i = \left\lfloor \frac{T_i}{T_s} \right\rfloor T_s\right) = \frac{\left\lfloor \frac{T_i}{T_s} \right\rfloor T_s - T_i}{T_s} = \left\lfloor \frac{T_i}{T_s} \right\rfloor - \frac{T_i}{T_s} \quad (5)$$

the expected value of which is given by:

$$\begin{aligned} \mathbb{E}(P(\xi=0)) &= \int_T P(\xi=0) f_{T_i}(T) dT \\ &= \sum_{m=0}^{\infty} (m+1) [F_{T_i}((m+1)T_s) - F_{T_i}(mT_s)] - \frac{\mathbb{E}(T_i)}{T_s} \end{aligned} \quad (6)$$

where $f_{T_i}(\cdot)$ is the Probability Density Function (PDF) of the original periods T_i , and χ_0 is a correction factor for $P(\xi = 0)$ given by:

$$\chi_0 = \frac{\left\lceil \frac{\mu_i}{T_s} \right\rceil T_s - \mu_i}{T_s} = \left\lceil \frac{\mu_i}{T_s} \right\rceil - \frac{\mu_i}{T_s} \quad (7)$$

Notice that for any arbitrary period T_i , the width of the interval $[\lfloor T_i/T_s \rfloor T_s, \lceil T_i/T_s \rceil T_s]$ is T_s . However, around the minimum period μ_i it holds that $T_i \in [\mu_i, \lceil \mu_i/T_s \rceil T_s]$ (since $T_i \geq \mu_i \geq \lfloor \mu_i/T_s \rfloor T_s$) and the width of such interval is $\lceil \mu_i/T_s \rceil T_s - \mu_i$ instead of T_s . Therefore, a scaling coefficient χ_0 is required for the probability $P(\xi = 0)$ as shown in (7).

The probability to observe at least one instance of $\hat{\mu}_i$ in the N observed periods can be related to the binomial distribution:

$$P_{\text{obs}}^{\hat{\mu}_i} = 1 - \left[1 - P(\hat{T}_i = \hat{\mu}_i) \right]^N \quad (8)$$

Finally, by specifying a probability of occurrence of the minimum, $P_{\text{obs}}^{\hat{\mu}_i}$, the minimum number of periods required to ensure the observation of the minimum period $\hat{\mu}_i$ is:

$$N_{\hat{\mu}_i} = \frac{\log(1 - P_{\text{obs}}^{\hat{\mu}_i})}{\log(1 - P(\hat{T}_i = \hat{\mu}_i))} \quad (9)$$

Notice that increasing the sample size will not improve the accuracy of the estimated minimum $\hat{\mu}_i$ itself, but the probability that a period $\hat{\mu}_i$ is observed (otherwise a longer period $\hat{T}_i > \hat{\mu}_i$ might be selected as the minimum period, thus potentially leading to a more inaccurate estimation).

IV. ESTIMATION OF THE MEAN AND VARIANCE

Given a set of N observed periods $\{\hat{T}_{i,n}\}_{n=1}^N$, the mean $\mathbb{E}(\hat{T}_i)$ and variance $\mathbb{V}(\hat{T}_i)$ of the provided durations can be estimated directly from the (unbiased) sample moments:

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

$$\mathbb{V}(\hat{T}_i) \approx \hat{v}_i = \frac{1}{N-1} \sum_{n=1}^N (\hat{T}_{i,n} - \hat{m}_i)^2 \quad (11)$$

Since the estimated periods $\{\hat{T}_{i,n}\}_{n=1}^N$ are integer multiples of the sensing period T_s as discussed in Section III, the sample moments obtained as shown in (10) and (11) will be affected by an error associated with the employed finite sensing period T_s . For the purposes of the analysis carried out in this section, the estimated periods can be modelled as $\hat{T}_i = T_i + T_e$, where T_e represents the above mentioned estimation error. Such error can be represented by the sum of two error components T_e^x and T_e^y shown in Fig. 1 (i.e., $T_e = T_e^y - T_e^x$). Both error components can take values within the interval $[0, T_s]$ and can be assumed to be uniformly distributed (i.e., $T_e^x, T_e^y \sim U(0, T_s)$); the analysis of simulation results indicated that this assumption

is valid. The impact on the estimated moments can thus be modelled as:

$$\begin{aligned} \mathbb{E}(\hat{T}_i) &= \mathbb{E}(T_i) + \mathbb{E}(T_e) \\ &= \mathbb{E}(T_i) + \mathbb{E}(T_e^y) - \mathbb{E}(T_e^x) = \mathbb{E}(T_i) \end{aligned} \quad (12)$$

$$\begin{aligned} \mathbb{V}(\hat{T}_i) &= \mathbb{V}(T_i) + \mathbb{V}(T_e) \\ &= \mathbb{V}(T_i) + \mathbb{V}(T_e^y) + \mathbb{V}(T_e^x) = \mathbb{V}(T_i) + \frac{T_s^2}{6} \end{aligned} \quad (13)$$

where $\mathbb{E}(T_e^y) - \mathbb{E}(T_e^x) = 0$ since T_e^x and T_e^y are identically distributed, $\mathbb{V}(T_e) = \mathbb{V}(T_e^y) + \mathbb{V}(T_e^x)$ assuming that T_e^x and T_e^y are independent (the same assumption can be found in [1, 17]), and $\mathbb{V}(T_e^x) = \mathbb{V}(T_e^y) = T_s^2/12$ is the variance of the uniform distribution $U(0, T_s)$ of T_e^x and T_e^y . As appreciated in (12), the estimated sample mean \hat{m}_i is not affected by the employed sensing period. On the other hand, as observed in (13), the estimated sample variance \hat{v}_i is affected by an error factor of $T_s^2/6$, which is constant and known. Based on (13), the effect of T_s can be removed by applying to (11) the appropriate correction factor:

$$\tilde{v}_i = \hat{v}_i - \frac{T_s^2}{6} \quad (14)$$

where \tilde{v}_i is the observed variance after correction (i.e., the corrected sample variance). This approach eliminates the impairments imposed by the use of a finite sensing period T_s in the estimated moments and is able to provide an accurate estimation of the real moments of T_i based on the estimated period durations $\{\hat{T}_{i,n}\}_{n=1}^N$, provided that a sufficiently large number of periods N is captured.

Notice that the estimation error T_e resulting from the use of a finite sensing period T_s cannot be reduced by increasing the observation sample size N , therefore it is necessary to first correct the sample moments (in this case, the sample variance) as discussed above in order to remove such error. Once the sample moments have been corrected, the resulting estimations can then be made arbitrarily close to the true population moments of the original periods T_i by taking a sufficiently large number of period observations N . Thus, the question to answer in this section is how large does the set $\{\hat{T}_{i,n}\}_{n=1}^N$ need to be (i.e., what is the required value of N) so that the sample mean \hat{m}_i and the corrected sample variance \tilde{v}_i of the set $\{\hat{T}_{i,n}\}_{n=1}^N$ are as close as desired to the original population moments, i.e., $\hat{m}_i \approx \mathbb{E}(T_i)$ and $\tilde{v}_i \approx \mathbb{V}(T_i)$.

Since the estimators \hat{m}_i in (10) and \tilde{v}_i in (14) are unbiased (i.e., $\mathbb{E}(\hat{m}_i) = \mathbb{E}(T_i)$ and $\mathbb{E}(\tilde{v}_i) = \mathbb{V}(T_i)$), the sample size required for a certain estimation error can be determined based on the standard errors of the estimators, which are related to their variances [23, p.229] as:

$$\mathbb{V}(\hat{m}_i) = \frac{\mathbb{V}(\hat{T}_i)}{N} \quad (15)$$

$$\mathbb{V}(\tilde{v}_i) = \mathbb{V}(\hat{v}_i) = \frac{1}{N} \left(\mathbb{M}_4(\hat{T}_i) - [\mathbb{V}(\hat{T}_i)]^2 \frac{N-3}{N-1} \right) \quad (16)$$

where $\mathbb{M}_4(\widehat{T}_i)$ is the fourth central moment of \widehat{T}_i , given by:

$$\begin{aligned}\mathbb{M}_4(\widehat{T}_i) &= \mathbb{E}([\widehat{T}_i - \mathbb{E}(\widehat{T}_i)]^4) \\ &= \mathbb{E}([T_i - \mathbb{E}(T_i) + T_e - \mathbb{E}(T_e)]^4) \\ &= \mathbb{M}_4(T_i) + 6\mathbb{V}(T_i)\mathbb{V}(T_e) + \mathbb{M}_4(T_e) \\ &= \mathbb{M}_4(T_i) + \mathbb{V}(T_i)T_s^2 + \frac{T_s^4}{15}\end{aligned}\quad (17)$$

where $\mathbb{V}(T_e) = \mathbb{V}(T_e^x) + \mathbb{V}(T_e^y) = T_s^2/6$ and $\mathbb{M}_4(T_e) = \mathbb{M}_4(T_e^x) + 6\mathbb{V}(T_e^x)\mathbb{V}(T_e^y) + \mathbb{M}_4(T_e^y) = T_s^4/15$, since $\mathbb{V}(T_e^x) = \mathbb{V}(T_e^y) = T_s^2/12$ and $\mathbb{M}_4(T_e^x) = \mathbb{M}_4(T_e^y) = T_s^4/80$.

Given an estimator ω , it is possible to define a confidence interval of κ standard deviations around the expected value of the estimator such that the estimated values are within that interval with a minimum probability ρ (confidence level):

$$P\left(|\omega - \mathbb{E}(\omega)| \leq \kappa\sqrt{\mathbb{V}(\omega)}\right) \geq \rho \quad (18)$$

If the estimator ω is unbiased, its relative error can then be bounded by $\varepsilon_{r,\max}^\omega \approx \kappa\sqrt{\mathbb{V}(\omega)}/\mathbb{E}(\omega)$. Based on this, the relative errors of the estimators \widehat{m}_i in (10) and \widehat{v}_i in (14), which are unbiased (i.e., $\mathbb{E}(\widehat{m}_i) = \mathbb{E}(T_i)$, $\mathbb{E}(\widehat{v}_i) = \mathbb{V}(T_i)$) are:

$$\varepsilon_{r,\max}^{\widehat{m}_i} \approx \frac{\kappa}{\mathbb{E}(T_i)} \left[\frac{1}{N} \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) \right]^{\frac{1}{2}} \quad (19)$$

$$\begin{aligned}\varepsilon_{r,\max}^{\widehat{v}_i} &\approx \frac{\kappa}{\mathbb{V}(T_i)} \left[\frac{1}{N} \left(\mathbb{M}_4(T_i) - \frac{N-3}{N-1} [\mathbb{V}(T_i)]^2 + \right. \right. \\ &\quad \left. \left. + \frac{2N}{3(N-1)} T_s^2 \mathbb{V}(T_i) + \frac{7N+3}{180(N-1)} T_s^4 \right) \right]^{\frac{1}{2}}\end{aligned}\quad (20)$$

which relates the maximum relative error of the estimators \widehat{m}_i and \widehat{v}_i to the observation sample size N (as well as the sensing period T_s and the moments of the original periods T_i).

The value of κ for a certain confidence level ρ can be derived from concentration inequalities (some examples are shown in Table II). However, this approach usually leads to loose upper bounds on the maximum relative error as it will be shown in Section IX. A much tighter result can be found by noting that \widehat{m}_i and \widehat{v}_i can be assumed to be normally distributed by the central limit theorem. The inequality in (18) can be rewritten for a normal distribution as follows:

$$\begin{aligned}P\left(\varepsilon_{\text{abs}}^\omega \leq \varepsilon_{\text{abs,max}}^\omega\right) &= P\left(|\omega - \mathbb{E}(\omega)| \leq \kappa\sqrt{\mathbb{V}(\omega)}\right) \\ &= \int_{\mathbb{E}(\omega) - \kappa\sqrt{\mathbb{V}(\omega)}}^{\mathbb{E}(\omega) + \kappa\sqrt{\mathbb{V}(\omega)}} \frac{e^{-\frac{1}{2}\left(\frac{\omega - \mathbb{E}(\omega)}{\sqrt{\mathbb{V}(\omega)}}\right)^2}}{\sqrt{2\pi\mathbb{V}(\omega)}} d\omega \\ &= \text{erf}\left(\frac{\kappa}{\sqrt{2}}\right) \geq \rho\end{aligned}\quad (21)$$

Solving (21) for κ yields the relation $\kappa \geq \sqrt{2}\text{erf}^{-1}(\rho)$.

It is worth noting that the approach employed to determine the relation between κ and ρ has a significant impact on the accuracy of (19) and (20) as well as the mathematical results derived later on for other PU activity statistics. The relations shown in Table II are concentration inequalities and therefore provide *bounds* on the true value of the maximum estimation error, while the relation $\kappa \geq \sqrt{2}\text{erf}^{-1}(\rho)$ obtained from (21) is an *approximation* to the true value of the maximum estimation

TABLE II
RELATION BETWEEN κ AND ρ FOR VARIOUS CONCENTRATION INEQUALITIES [24].

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

error. As such, the latter can be expected to be more accurate. This will be shown and discussed in detail in Section IX.

The observation sample size required to guarantee a predefined maximum relative error follows from (19) and (20):

$$\begin{aligned}N_{\widehat{m}_i} &\approx \left(\frac{\kappa}{\mathbb{E}(T_i) \varepsilon_{r,\max}^{\widehat{m}_i}} \right)^2 \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) \\ N_{\widehat{v}_i} &\approx \left(\frac{\kappa}{\mathbb{V}(T_i) \varepsilon_{r,\max}^{\widehat{v}_i}} \right)^2 \left(\mathbb{M}_4(T_i) - [\mathbb{V}(T_i)]^2 + \frac{2T_s^2\mathbb{V}(T_i)}{3} + \frac{7T_s^4}{180} \right)\end{aligned}\quad (22)$$

where $N \gg 3$ has been assumed in (20). This assumption is reasonable as a relatively large observation sample size is usually required for an accurate estimation.

It is worth noting that the required sample size for the estimation of moments (sample mean \widehat{m}_i and corrected sample variance \widehat{v}_i) may in some cases be relatively high, depending on the desired level of estimation accuracy, the statistics of the original periods T_i and the employed sensing period T_s . In DSA/CR devices with limited memory capabilities, this problem can be overcome by computing the sample moments based on recurrence formulae [25] instead of storing the complete history of the N past observed period durations.

V. ESTIMATION OF THE DUTY CYCLE

The duty cycle is an important statistic commonly used to characterise the level of occupancy of a primary channel or frequency band (defined as the probability that the primary channel is busy, i.e., occupied by a PU signal).

The duty cycle can be estimated based on individual spectrum sensing decisions as the ratio of the number of sensing events with a busy/on (H_1) decision to the total number of sensing events. The work reported in [20] provides an analytical study on the sample size (understood in this case as the number of individual sensing events) required for an arbitrarily accurate estimation of the duty cycle based on this approach. However, taking into account that the duty cycle, denoted as Ψ , can be related to the mean value of idle and busy periods as follows:

$$\Psi = \frac{\mathbb{E}(T_1)}{\mathbb{E}(T_0) + \mathbb{E}(T_1)} \quad (24)$$

an estimation thereof, denoted as $\widehat{\Psi}$, can also be obtained based on the sample mean estimator \widehat{m}_i in (10) as shown below:

$$\widehat{\Psi} = \frac{\widehat{m}_1}{\widehat{m}_0 + \widehat{m}_1} \quad (25)$$

where \widehat{m}_0 and \widehat{m}_1 represent the sample mean of idle/off and busy/on periods, respectively. This section provides an analytical study on the observation sample size N (i.e., number of periods in the observed set $\{\widehat{T}_{i,n}\}_{n=1}^N$) required for an arbitrarily accurate estimation of the duty cycle based on (25).

The standard error (i.e., the standard deviation, or equivalently the variance) of the estimated sample mean can be propagated through (24)–(25) to obtain the standard error (or variance) of the estimated duty cycle [26, 27]:

$$\mathbb{V}(\widehat{\Psi}) = \left(\frac{\partial \widehat{\Psi}}{\partial \widehat{m}_0} \right)^2 \mathbb{V}(\widehat{m}_0) + \left(\frac{\partial \widehat{\Psi}}{\partial \widehat{m}_1} \right)^2 \mathbb{V}(\widehat{m}_1) \quad (26)$$

Based on (18), the relative error of the duty cycle estimated from (25) is obtained as:

$$\begin{aligned} \varepsilon_{r,\max}^{\widehat{\Psi}} &\approx \frac{\kappa}{\mathbb{E}(\widehat{\Psi})} \sqrt{\mathbb{V}(\widehat{\Psi})} \\ &= \frac{\kappa}{\widehat{\Psi}} \left[\frac{1}{N[\mathbb{E}(T_0) + \mathbb{E}(T_1)]^4} \left\{ [\mathbb{E}(T_1)]^2 \left(\mathbb{V}(T_0) + \frac{T_s^2}{6} \right) + \right. \right. \\ &\quad \left. \left. + [\mathbb{E}(T_0)]^2 \left(\mathbb{V}(T_1) + \frac{T_s^2}{6} \right) \right\} \right]^{\frac{1}{2}} \end{aligned} \quad (27)$$

where $\mathbb{E}(\widehat{\Psi}) = \Psi$ and $\mathbb{V}(\widehat{\Psi})$ is obtained by introducing (13) and (15) into (26) and solving the derivatives.

The observation sample size required to guarantee a given maximum relative error follows from solving (27) for N :

$$N_{\widehat{\Psi}} \approx \left(\frac{\kappa}{\Psi \varepsilon_{r,\max}^{\widehat{\Psi}}} \right)^2 \frac{\Psi^2 \left[\mathbb{V}(T_0) + \frac{T_s^2}{6} \right] + (1 - \Psi)^2 \left[\mathbb{V}(T_1) + \frac{T_s^2}{6} \right]}{[\mathbb{E}(T_0) + \mathbb{E}(T_1)]^2} \quad (28)$$

where κ can be obtained from concentration inequalities (Table II) or the normal approximation as described in Section IV.

VI. ESTIMATION OF THE DISTRIBUTION

The distribution of the busy/on and idle/off times provides a complete characterisation of the PU activity statistics and its accurate estimation is therefore of great importance in the context of DSA/CR systems. Several methods can be used to estimate the distribution of the PU busy/idle times based on a finite set $\{\widehat{T}_{i,n}\}_{n=1}^N$ of N observed periods \widehat{T}_i . The most commonly used method is the direct calculation of the empirical CDF of the set $\{\widehat{T}_{i,n}\}_{n=1}^N$. The main drawback of this method is that the estimated distribution has a discrete domain (even though the original distribution is in general continuous) since the periods \widehat{T}_i estimated as shown in Fig. 1 are integer multiples of the sensing period T_s . This leads to an irreducible estimation error that depends on the employed sensing period T_s and cannot be improved by increasing the sample size N [17]. This limitation can be overcome by the method of moments proposed in [22], where the parameters of the primary distribution are estimated based on the sample moments of the set $\{\widehat{T}_{i,n}\}_{n=1}^N$. This alternative method relies on the assumption of a particular model for the PU distribution.

A common assumption frequently employed in the literature is that idle/busy periods follow an exponential distribution [16, 28–30], which simplifies analytical studies. Field measurements, however, have shown that this model is unrealistic

[31–33]. A more realistic and broader model is the Generalised Pareto (GP) distribution [21]. While many models are available in the literature for the PU traffic, most of them have been developed based on a single radio technology (see [21] and references therein). On the other hand, the work conducted in [21] demonstrated that the GP model can provide a very accurate fit to the distribution of PU periods over wide range of radio technologies (including amateur, paging, TETRA, mobile communications, DECT, ISM band and others). This may be explained by the versatility of the GP distribution, since many of the models widely used in the literature can be obtained as particular cases (exponential or Pareto) or approximated very accurately (generalised exponential, gamma or Weibull, among others). The analysis presented in this section considers the GP distribution model, which leads to more realistic and widely applicable results.

According to the GP model, the CDF of the original periods T_i is given by [34, ch. 20]:

$$F_{T_i}(T) = 1 - \left[1 + \frac{\alpha_i(T - \mu_i)}{\lambda_i} \right]^{-1/\alpha_i}, \quad T \geq \mu_i \quad (29)$$

where $\mu_i > 0$, $\lambda_i > 0$, $\alpha_i \in \mathbb{R}$ are the location, scale and shape parameters, respectively. Based on (29), the method proposed in [22] provides an estimated distribution:

$$F_{\widehat{T}_i}(T) = 1 - \left[1 + \frac{\widehat{\alpha}_i(T - \widehat{\mu}_i)}{\widehat{\lambda}_i} \right]^{-1/\widehat{\alpha}_i}, \quad T \geq \widehat{\mu}_i \quad (30)$$

Parameter $\widehat{\mu}_i$ represents the estimated minimum period and its value is assumed to be known to a reasonable level of accuracy ($\widehat{\mu}_i \approx \mu_i$). According to the method of moments, the scale and shape parameters are estimated as [34, ch. 20]:

$$\widehat{\lambda}_i = \frac{1}{2} \left(1 + \frac{(\widehat{m}_i - \widehat{\mu}_i)^2}{\widehat{v}_i} \right) (\widehat{m}_i - \widehat{\mu}_i) \quad (31a)$$

$$\widehat{\alpha}_i = \frac{1}{2} \left(1 - \frac{(\widehat{m}_i - \widehat{\mu}_i)^2}{\widehat{v}_i} \right) \quad (31b)$$

where the sample mean \widehat{m}_i and the corrected sample variance \widehat{v}_i are obtained as shown in (10) and (14), respectively, and $\widehat{\mu}_i \approx \mu_i$. Introducing the MoM estimates provided by (31) into (30) provides a continuous estimation of the distribution of PU on/off times. Such estimation can be made arbitrarily close to the true distribution in (29) by increasing the number of periods N used to calculate \widehat{m}_i and \widehat{v}_i .

To determine the observation sample size N required for an arbitrarily accurate estimation, it is necessary to express the error in the estimated distribution as a function of N . This can be achieved by propagating the standard error of \widehat{m}_i and \widehat{v}_i , $\mathbb{V}(\widehat{m}_i)$ and $\mathbb{V}(\widehat{v}_i)$ in (15) and (16) respectively, through (31):

$$\mathbb{V}(\widehat{\lambda}_i) = \left(\frac{\partial \widehat{\lambda}_i}{\partial \widehat{m}_i} \right)^2 \mathbb{V}(\widehat{m}_i) + \left(\frac{\partial \widehat{\lambda}_i}{\partial \widehat{v}_i} \right)^2 \mathbb{V}(\widehat{v}_i) + 2 \frac{\partial \widehat{\lambda}_i}{\partial \widehat{m}_i} \frac{\partial \widehat{\lambda}_i}{\partial \widehat{v}_i} \mathbb{C}(\widehat{m}_i, \widehat{v}_i) \quad (32a)$$

$$\mathbb{V}(\widehat{\alpha}_i) = \left(\frac{\partial \widehat{\alpha}_i}{\partial \widehat{m}_i} \right)^2 \mathbb{V}(\widehat{m}_i) + \left(\frac{\partial \widehat{\alpha}_i}{\partial \widehat{v}_i} \right)^2 \mathbb{V}(\widehat{v}_i) + 2 \frac{\partial \widehat{\alpha}_i}{\partial \widehat{m}_i} \frac{\partial \widehat{\alpha}_i}{\partial \widehat{v}_i} \mathbb{C}(\widehat{m}_i, \widehat{v}_i) \quad (32b)$$

Since the sample mean and sample variance are not independent, the covariance between both, denoted as $\mathbb{C}(\widehat{m}_i, \widehat{v}_i)$, needs to be included in (32). Such covariance is given by $\mathbb{C}(\widehat{m}_i, \widehat{v}_i) = \mathbb{M}_3(T_i)/N$ [35] where $\mathbb{M}_3(T_i)$ is the third central moment of the population distribution.

The standard errors obtained from (32) can be further propagated through (30) to obtain (assuming $\widehat{\mu}_i \approx \mu_i$):

$$\mathbb{V}(F_{\widehat{T}_i}(T)) = \left(\frac{\partial F_{\widehat{T}_i}(T)}{\partial \widehat{\lambda}_i} \right)^2 \mathbb{V}(\widehat{\lambda}_i) + \left(\frac{\partial F_{\widehat{T}_i}(T)}{\partial \widehat{\alpha}_i} \right)^2 \mathbb{V}(\widehat{\alpha}_i) \quad (33)$$

and the absolute error of the estimated distribution can then be written as:

$$|F_{T_i}(T) - F_{\widehat{T}_i}(T)| = \kappa \sqrt{\mathbb{V}(F_{\widehat{T}_i}(T))} \quad (34)$$

A common metric typically employed to quantify the difference between two CDFs is the Kolmogorov-Smirnov (KS) distance, which is defined as [36]:

$$D_{KS} = \sup_T |F_{T_i}(T) - F_{\widehat{T}_i}(T)| \quad (35)$$

and can be obtained from (34) by solving:

$$\frac{\partial}{\partial T} |F_{T_i}(T) - F_{\widehat{T}_i}(T)| = \frac{\partial}{\partial T} [\kappa \sqrt{\mathbb{V}(F_{\widehat{T}_i}(T))}] = 0 \quad (36)$$

Unfortunately (36) cannot be solved in closed form. However, it can be shown that the value of (33) is mainly dominated (and can be approximated) by its first term. As a result, an approximated result to (36) can be obtained by solving:

$$\frac{\partial}{\partial T} |F_{T_i}(T) - F_{\widehat{T}_i}(T)| \approx \frac{\partial}{\partial T} \left[\kappa \sqrt{\left(\frac{\partial F_{\widehat{T}_i}(T)}{\partial \widehat{\lambda}_i} \right)^2 \mathbb{V}(\widehat{\lambda}_i)} \right] = 0 \quad (37)$$

which is maximised for $T = \widehat{\mu}_i + \widehat{\lambda}_i$.

Combining (30)–(35) and evaluating the resulting expression in $T = \widehat{\mu}_i + \widehat{\lambda}_i$ yields the final expression for D_{KS} :

$$D_{KS} = \kappa (1 + \alpha_i)^{-\frac{1}{\alpha_i} - 1} \left[\frac{1}{\lambda_i^2} \mathbb{V}(\widehat{\lambda}_i) + \frac{[(1 + \alpha_i) \ln(1 + \alpha_i) - \alpha_i]^2}{\alpha_i^4} \mathbb{V}(\widehat{\alpha}_i) \right]^{\frac{1}{2}} \quad (38)$$

where $\mathbb{V}(\widehat{\lambda}_i)$ and $\mathbb{V}(\widehat{\alpha}_i)$ are obtained from (32) and given by (44). The relevant moments in (44) for the GP distribution are:

$$\mathbb{E}(T_i) = \mu_i + \frac{\lambda_i}{1 - \alpha_i} \quad (39)$$

$$\mathbb{V}(T_i) = \frac{\lambda_i^2}{(1 - \alpha_i)^2(1 - 2\alpha_i)} \quad (40)$$

$$\mathbb{M}_3(T_i) = \frac{2\lambda_i^3(\alpha_i + 1)}{(1 - \alpha_i)^3(1 - 2\alpha_i)(1 - 3\alpha_i)} \quad (41)$$

$$\mathbb{M}_4(T_i) = \frac{3\lambda_i^4(2\alpha_i^2 + \alpha_i + 3)}{(1 - \alpha_i)^4(1 - 2\alpha_i)(1 - 3\alpha_i)(1 - 4\alpha_i)} \quad (42)$$

The result in (38) provides a closed-form relation between the observation sample size, N , and the error of the estimated CDF, $F_{\widehat{T}_i}(T)$, in terms of the KS distance with respect to the true distribution $F_{T_i}(T)$. The observation sample size required

to guarantee a given KS distance follows from solving (38) for N :

$$N_{F_{\widehat{T}_i}(T)} = \left(\frac{\kappa}{D_{KS}} \right)^2 (1 + \alpha_i)^{-\frac{2}{\alpha_i} - 2} \left[\frac{1}{\lambda_i^2} \Omega(T_i) + \frac{[(1 + \alpha_i) \ln(1 + \alpha_i) - \alpha_i]^2}{\alpha_i^4} \Upsilon(T_i) \right] \quad (43)$$

where $\Omega(T_i)$ and $\Upsilon(T_i)$ are obtained from (44), assuming $N \gg 3$ and therefore $(N - 3)/(N - 1) \approx 1$, and are given by (45).

VII. ITERATIVE STOPPING ALGORITHM

The analytical results obtained in previous sections provide closed-form expressions for the required observation sample size as a function of the desired estimation error for the minimum period (9), mean (22), variance (23), duty cycle (28) and distribution (43). Notice that such expressions depend not only on the desired estimation error but also on the real moments and/or distribution parameters of the PU traffic, which are unknown to the SU (and indeed the parameters to be estimated). As a result, such expressions cannot be used in a real implementation directly, without further considerations.

To overcome this problem, an iterative stopping algorithm is here proposed. The proposed algorithm is composed of three steps that are executed every time a new idle/busy period is observed (i.e., every time a new sample becomes available)¹:

- 1) Update the calculated sample moments (mean, variance, third/fourth central moments) based on appropriate recurrence equations [25] and apply any required correction factors as appropriate (see the appendix).
- 2) Evaluate the expression for the estimation error – i.e., (19) for the mean, (20) for the variance, (27) for the duty cycle, and (38) for the distribution – based on the most recent sample estimates of the moments obtained in previous step (instead of the true moments as shown in such expressions) and the current sample size (number of observed periods).
- 3) If the value obtained in previous step is lower than the desired estimation error then stop, otherwise continue.

An example of how this algorithm would be implemented in a real system for the estimation of the mean period is shown in Algorithm 1 (the algorithm would need to be adapted to other PU activity statistics according to the different obtained analytical results but the operation principle would be the same). The key idea is that the theoretical error expressions are calculated by replacing the true moments (unknown to the DSA/CR system) with their last (known) sample estimates. In the shown example, (19) is evaluated by replacing $\mathbb{E}(T_i)$ and $\mathbb{V}(T_i)$ with their corresponding sample estimates \widehat{m}_i and \widehat{v}_i , respectively. It can be verified that such estimation error decreases as the number of observed periods used to calculate the sample estimates increases. When such value is lower than the target/desired estimation error, the algorithm will indicate

¹As discussed in Section III, the estimated minimum period cannot be made arbitrarily accurate by increasing the sample size. Therefore the algorithm proposed in this section is applicable to the rest of PU activity statistics (i.e., mean/variance of the estimated periods, channel duty cycle and distribution.)

$$\begin{aligned} \mathbb{V}(\hat{\lambda}_i) = & \frac{1}{N} \left\{ \frac{1}{4} \left(1 + \frac{3[\mathbb{E}(T_i) - \mu_i]^2}{\mathbb{V}(T_i)} \right)^2 \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) - \frac{\mathbb{M}_3(T_i) [\mathbb{E}(T_i) - \mu_i]^3}{2 [\mathbb{V}(T_i)]^2} \left(1 + \frac{3[\mathbb{E}(T_i) - \mu_i]^2}{\mathbb{V}(T_i)} \right) + \right. \\ & \left. + \frac{1}{4} \frac{[\mathbb{E}(T_i) - \mu_i]^6}{[\mathbb{V}(T_i)]^4} \left[\mathbb{M}_4(T_i) + \mathbb{V}(T_i) T_s^2 + \frac{T_s^4}{15} - \frac{N-3}{N-1} \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right)^2 \right] \right\} \end{aligned} \quad (44a)$$

$$\begin{aligned} \mathbb{V}(\hat{\alpha}_i) = & \frac{1}{N} \left\{ \left(\frac{\mathbb{E}(T_i) - \mu_i}{\mathbb{V}(T_i)} \right)^2 \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) - \left(\frac{\mathbb{E}(T_i) - \mu_i}{\mathbb{V}(T_i)} \right)^3 \mathbb{M}_3(T_i) + \right. \\ & \left. + \frac{1}{4} \left(\frac{\mathbb{E}(T_i) - \mu_i}{\mathbb{V}(T_i)} \right)^4 \left[\mathbb{M}_4(T_i) + \mathbb{V}(T_i) T_s^2 + \frac{T_s^4}{15} - \frac{N-3}{N-1} \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right)^2 \right] \right\} \end{aligned} \quad (44b)$$

$$\begin{aligned} \Omega(T_i) = & \frac{1}{4} \left(1 + \frac{3[\mathbb{E}(T_i) - \mu_i]^2}{\mathbb{V}(T_i)} \right)^2 \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) - \frac{\mathbb{M}_3(T_i) [\mathbb{E}(T_i) - \mu_i]^3}{2 [\mathbb{V}(T_i)]^2} \left(1 + \frac{3[\mathbb{E}(T_i) - \mu_i]^2}{\mathbb{V}(T_i)} \right) + \\ & + \frac{1}{4} \frac{[\mathbb{E}(T_i) - \mu_i]^6}{[\mathbb{V}(T_i)]^4} \left[\mathbb{M}_4(T_i) + \mathbb{V}(T_i) T_s^2 + \frac{T_s^4}{15} - \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right)^2 \right] \end{aligned} \quad (45a)$$

$$\begin{aligned} \Upsilon(T_i) = & \left(\frac{\mathbb{E}(T_i) - \mu_i}{\mathbb{V}(T_i)} \right)^2 \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) - \left(\frac{\mathbb{E}(T_i) - \mu_i}{\mathbb{V}(T_i)} \right)^3 \mathbb{M}_3(T_i) + \\ & + \frac{1}{4} \left(\frac{\mathbb{E}(T_i) - \mu_i}{\mathbb{V}(T_i)} \right)^4 \left[\mathbb{M}_4(T_i) + \mathbb{V}(T_i) T_s^2 + \frac{T_s^4}{15} - \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right)^2 \right] \end{aligned} \quad (45b)$$

that the number of observed periods is sufficient to estimate the PU activity statistic of interest with the desired level of accuracy. The SU receiver can then use the set of observed periods to produce an accurate estimation, after which the value of N can be reset to zero in order to start capturing a new set of samples for the next estimation (notice that when the proposed algorithm is executed in real-time the value of N at which it will stop may not necessarily be identical in every execution, even though it will be similar). Notice that all the required input information would be known in a real implementation (desired estimation accuracy in terms of the target error $\varepsilon_{r,\max}^{\hat{m}_i}$ and confidence interval ρ along with the employed sensing period T_s). Therefore, this iterative stopping algorithm enables the practical implementation of the theoretical results obtained in this work in a practical context.

VIII. SIMULATION AND EXPERIMENTAL METHODOLOGY

The analytical results obtained in previous sections were compared to results obtained from both software simulations and hardware experiments. Simulations were performed in MATLAB by generating several sequences with a sufficiently large number of interleaved on/busy and off/idle periods from a GP distribution with predefined location (μ_i), scale (λ_i) and shape (α_i) parameters. The generated periods T_i were sensed with a specified sensing period T_s in order to calculate the corresponding sequence of estimated periods \hat{T}_i that would be observed by a DSA/CR receiver following the principle shown in Fig. 1. The set of observed periods, $\{\hat{T}_{i,n}\}_{n=1}^N$, was used to estimate the primary activity statistics as described in

Algorithm 1 Iterative stopping algorithm (for the mean period)

Input: $\varepsilon_{r,\max}^{\hat{m}_i}$, ρ , T_s

Output: N

```

1: continue = true,  $N = 1$ 
2: while continue == true do
3:   Sense PU channel until a set  $\{\hat{T}_{i,n}\}_{n=1}^N$  is available
4:   Calculate  $\hat{m}_i = \frac{1}{N} \sum_{n=1}^N \hat{T}_{i,n}$  from (10)
5:   Calculate  $\tilde{v}_i = \left[ \frac{1}{N-1} \sum_{n=1}^N \left( \hat{T}_{i,n} - \hat{m}_i \right)^2 \right] - \frac{T_s^2}{6}$  from (14)
6:   Calculate  $\varepsilon_{r,\text{new}}^{\hat{m}_i} = \frac{\sqrt{2} \text{erf}^{-1}(\rho)}{\hat{m}_i} \left[ \frac{1}{N} \left( \tilde{v}_i + \frac{T_s^2}{6} \right) \right]^{\frac{1}{2}}$  from (19)
7:   if  $\varepsilon_{r,\text{new}}^{\hat{m}_i} \leq \varepsilon_{r,\max}^{\hat{m}_i}$  then
8:     continue = false
9:   else
10:     $N = N + 1$ 
11:   end if
12: end while
```

previous sections and compared to the original true statistics as a function of the observation sample size N .

The hardware experiments were conducted with a Prototype for the Estimation of Channel Activity Statistics (PECAS) [37]. This prototype is implemented with common low-cost components with the aim to reproduce a realistic scenario with inexpensive DSA/CR devices and introduce typical hardware sources of error and inaccuracies. This prototype is based on free open source code². The hardware experiments are

²Available at: www.lopezbenitez.es/misc/PECAS.zip

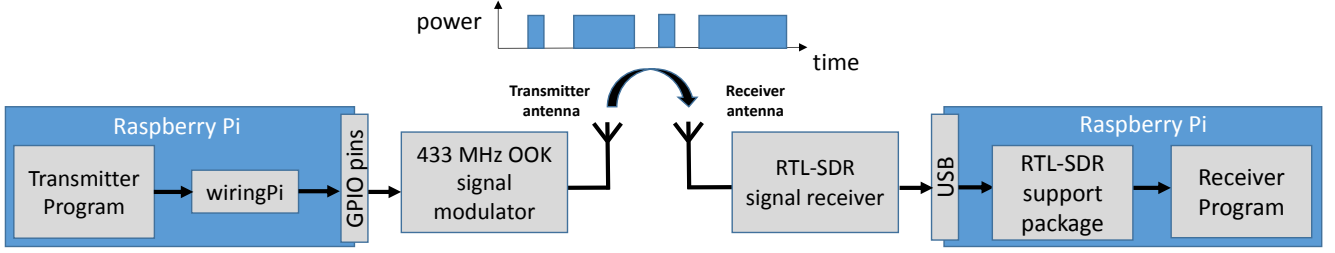


Fig. 2. Block diagram of the PECAS prototype employed for hardware experiments [37].

based on the same principle as the simulations but using a real transmitter and a real receiver. The block diagram is shown in Fig. 2. The transmitter (primary user) sends a sequence of GP-distributed idle/busy periods utilising a 433 MHz ON-OFF Keying (OOK) modulator with an output power of 2 dBm (controlled from a C program based on the `wiringPi` library). The receiver (secondary DSA/CR user), placed 1 metre apart, uses a Software-Defined Radio (SDR) with a gain of 20 dB to monitor the transmitter activity (idle/busy) at 433 MHz every T_s seconds. At every sensing event, signal samples are captured at a sample rate of 10^6 samples per second, which are processed to decide the instantaneous channel state (idle/busy) using energy detection. The outcomes of the energy detection decisions are used to estimate the durations of the observed idle/busy periods as shown in Fig. 1 and compute the primary activity statistics. While transmitter and receiver are controlled by C programs running on the same Raspberry Pi microcomputer, both programs run independently without synchronisation (as it would be the case of primary/secondary users in a real scenario). Real-time operation is achieved by a patched version of the Linux kernel and running the programs with real-time priority.

IX. SIMULATION AND EXPERIMENTAL RESULTS

In this section, a comprehensive analysis of the obtained analytical results as well as their validation with simulation and experimental results are presented. The value considered for each parameter is shown in the title of each figure in terms of generic time units (t.u.). In the case of experimental results, where a particular time unit needs to be selected according to the real-time capabilities of the employed hardware platform, the reference unit is the second (i.e., 1 t.u. = 1 second).

Fig. 3 shows the required observation sample size for the estimation of the minimum period as a function of the sensing period based on (9). As it can be observed, the required sample size increases with the desired probability of observation $P_{\text{obs}}^{\mu_i}$. It is worth noting that local minima are observed for T_s values that are integer submultiples of the true minimum (i.e., for $T_s = \mu_i/k$ with $k \in \mathbb{N}^+$). This can be explained by the fact that for such values of T_s it is possible to provide an exact estimation of the true minimum (see [18]). However for slightly higher values the required sample size tends to infinity, since in such a case $\chi_0 \approx 0$, $P(\hat{T}_i = \hat{\mu}_i) \approx 0$, and the denominator of (9) tends to zero as well. It is worth noting that the analytical result in (9) provides a perfect match with both simulation and experimental results.

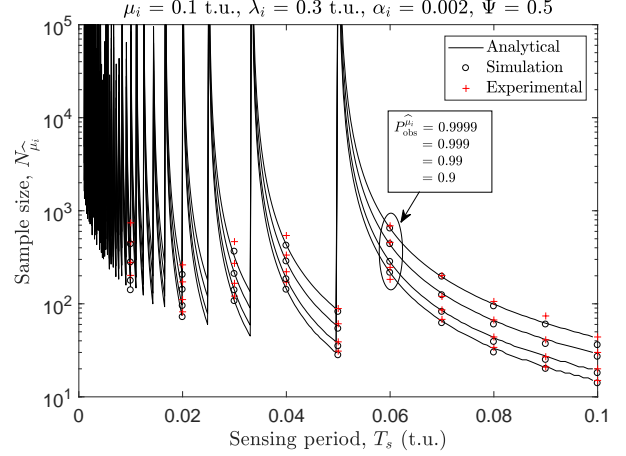


Fig. 3. Required observation sample size for the estimation of the minimum period as a function of the sensing period (duty cycle $\Psi = 0.5$).

Fig. 4 shows the maximum relative error of the estimated mean, $\varepsilon_{r,\text{max}}^{\hat{\mu}_i}$, observed at the 95% percentile ($\rho = 0.95$) as a function of the sample size when the channel duty cycle is $\Psi = 0.5$ (i.e., idle and busy periods have the same average duration). As it can be appreciated, when κ is determined based on concentration inequalities as the ones shown in Table II the result in (19) represents a loose upper bound to the true relative error. On the other hand, if κ is calculated assuming that the sample mean estimates are normally distributed as shown in (21), then the result in (19) represents a very accurate expression for the relative error of the estimated mean (as corroborated by both simulation and experimental results), which can then be used to precisely determine the observation sample size required for an accurate estimation as indicated in (22). The counterpart results for the estimated variance are shown in Fig. 5. In this case, the analytical result in (20) does not follow the simulation and experimental results when the observation sample is low, even if the normal approximation is considered for the calculation of κ . This can be explained by the fact that the assumption of normally distributed values of the sample variance (based on the central limit theorem) considered in (21) is valid only for a sufficiently large number of samples. As a result, if the sample size is not sufficiently large (for the particular evaluation conditions considered in the example of Fig. 5 this corresponds approximately to $N < 2000$) then the expression in (20) differs slightly from the true relative error. However, for a sufficiently large observation

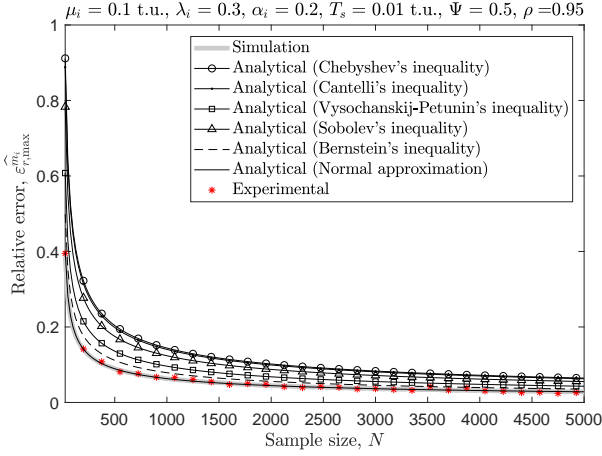


Fig. 4. Maximum relative error of the estimated mean observed at the 95% percentile ($\rho = 0.95$) as a function of the sample size (duty cycle $\Psi = 0.5$).

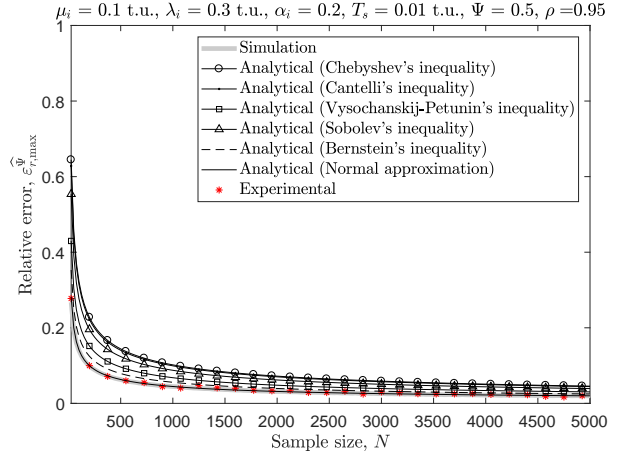


Fig. 6. Maximum relative error of the estimated duty cycle at the 95% percentile ($\rho = 0.95$) as a function of the sample size (duty cycle $\Psi = 0.5$).

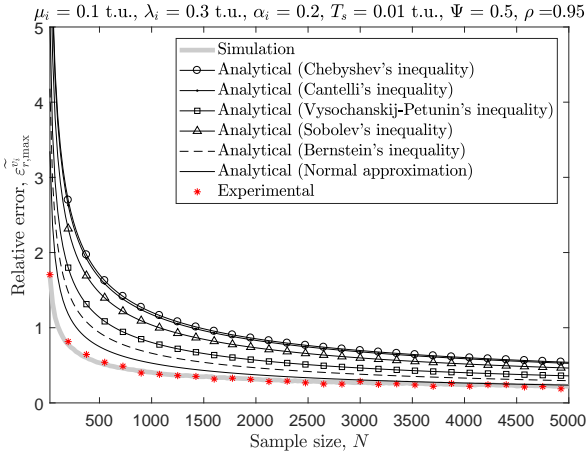


Fig. 5. Maximum relative error of the estimated variance observed at the 95% percentile ($\rho = 0.95$) as a function of the sample size (duty cycle $\Psi = 0.5$).

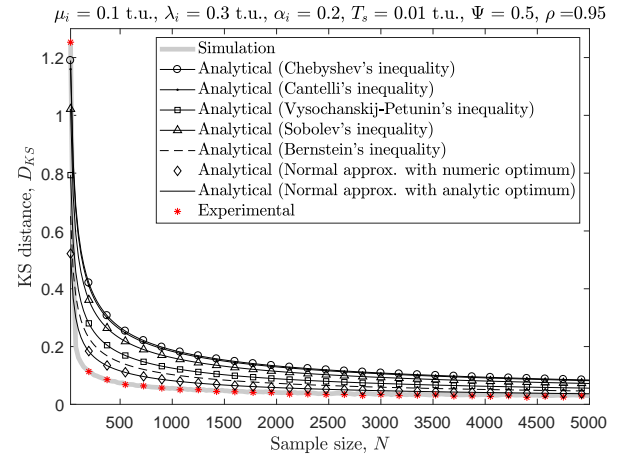


Fig. 7. KS distance of the estimated distribution at the 95% percentile ($\rho = 0.95$) as a function of the sample size (duty cycle $\Psi = 0.5$).

sample size (approximately $N > 2000$ in the example of Fig. 5) the result in (20) provides a very tight approximation for the true relative error, as shown by the perfect agreement with simulation and experimental results in this region of the figure. Since an accurate estimation of primary activity statistics will in general require a large sample size, the result in (20) is in practice accurate where it needs to be, and the observation sample size required for an accurate estimation of the variance can therefore be determined precisely based on (23).

Fig. 6 shows the maximum relative error of the estimated duty cycle observed at the 95% percentile ($\rho = 0.95$) as a function of the sample size when the channel duty cycle is $\Psi = 0.5$. Since the duty cycle is calculated based on the estimated mean value of idle and busy periods as shown in (25), the observed results and the accuracy of (27) and (28) can be explained based on the same arguments as those for Fig. 4.

Fig. 7 shows the maximum KS distance of the estimated distribution at the 95% percentile ($\rho = 0.95$) as a function of the sample size when the channel duty cycle is $\Psi = 0.5$. The normal approximation for the calculation of κ is depicted when the value of T that maximises (35) is calculated by numerical

evaluation of (36) (numeric optimum) and analytically based on the approximation considered in (37) (analytic optimum). As it can be appreciated, the approximation considered in (37) leads to very accurate results. The results observed in Fig. 7 show that the analytical result in (38) provides a very accurate evaluation of the KS distance for sufficiently large sample sizes (approximately $N > 2000$), similar to the trend observed in Fig. 5. Following the same argument, it can be stated that the analytical result in (43) provides in practice a close prediction of the observation sample size required for an accurate estimation of the distribution.

Fig. 8 compares the analytical predictions of the required sample size as a function of the desired estimation error for the mean (22), variance (23), duty cycle (28) and distribution (43) with the estimations provided by the algorithm of Section VII (based on simulations and experiments). As appreciated, the execution of the algorithm terminates at values of N very close to the analytical predictions. Therefore the proposed algorithm can be used by real SU devices to determine in real-time how many samples (observed periods) are required to accurately estimate the activity statistics of an unknown PU channel.

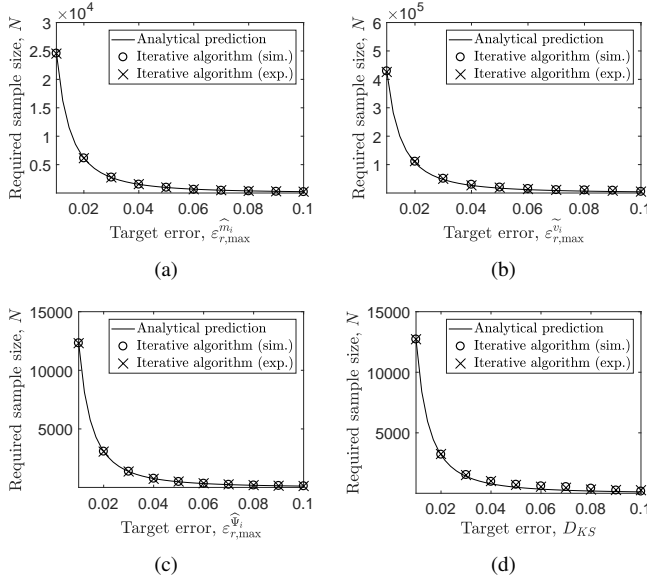


Fig. 8. Required sample size as a function of the desired estimation error for: (a) mean, (b) variance, (c) channel duty cycle, and (d) distribution ($\mu_i = 0.1$ t.u., $\lambda_i = 0.3$ t.u., $\alpha_i = 0.05$, $T_s = 0.01$ t.u., $\Psi = 0.5$, $\rho = 0.95$).

X. DISCUSSION OF PRACTICAL ASPECTS

This section discusses several important aspects related with the practical application of the analytical results obtained in this work in a real context.

An important practical aspect is the degree to which the accuracy of the estimated primary activity statistics can affect the performance of DSA/CR systems. To illustrate this, let us consider as a practical example the problem of channel selection, which is one of the cases where primary activity statistics can be useful as discussed in Section I. A simple channel selection approach is to select the channel that provides the highest expected opportunistic bit-rate (\hat{R}_b) which can be expressed as a function of the estimated duty cycle ($\hat{\Psi}$) as $\hat{R}_b = (1 - \hat{\Psi})W\eta$, where W is the primary channel bandwidth and η is the spectrum efficiency associated to the modulation and coding schemes used by the DSA/CR system. Fig. 9 shows the estimated available bit-rate (with $W = 20$ MHz and $\eta = 2$ bit/s/Hz) as a function of the estimated channel duty cycle for three cases: $N = 100$ (insufficient sample size), $N = 2000$ (sufficiently large sample size) and $N \rightarrow \infty$ (infinite sample size and therefore perfect duty cycle estimation). The lines shown represent the worst-case upper and lower bounds corresponding to $\hat{\Psi} = \Psi(1 \pm \varepsilon_{r,\max}^{\hat{\Psi}})$, with $\varepsilon_{r,\max}^{\hat{\Psi}}$ given by (27). As it can be appreciated in this example, an insufficient sample size ($N = 100$) can lead to estimation errors of up to ± 5 Mbit/s in the expected data rate, which for $\Psi = 0.5$ (where the true available data rate is 20 Mbit/s) represents an error of 25%. On the other hand, with a sufficiently large sample size ($N = 2000$), the estimation error is less than 1 Mbit/s (5% error for $\Psi = 0.5$), which can be made arbitrarily low by further increasing the sample size. This simple numerical example illustrates the potential impact that inaccurate primary activity statistics can have in the performance of DSA/CR systems and highlights the practical importance of the analytical results

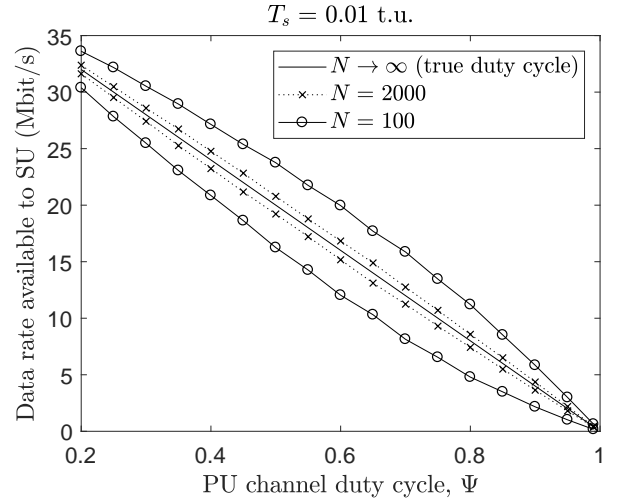


Fig. 9. Estimated available data rate as a function of the estimated duty cycle.

obtained in this work to determine the sample size required for an accurate estimation of the primary activity statistics.

Another aspect of practical interest is for how long (in absolute time units) the DSA/CR system needs to observe a PU channel before an accurate estimation of the PU statistics is available, and how this compares to the operation time scale of the DSA/CR system. All mathematical results for the considered PU activity statistics are provided in terms of the number of periods that need to be observed (N), which can be easily tracked by the DSA/CR system in a practical implementation (increasing a counter every time a new PU period is observed until the required sample size is reached). The total observation time in absolute time units, if it needs to be known, can be readily obtained as $N \cdot \mathbb{E}(T_i)$, where $\mathbb{E}(T_i)$ is the average PU period duration (which would also need to be estimated as described in Section IV). Notice that the required observation time for a particular PU channel depends on the PU channel itself and its specific PU occupancy, quantified through $\mathbb{E}(T_i)$. This is illustrated in Fig. 10, which shows the required observation time for different PU activity statistics as a function of the average PU period duration. In most practical cases, the required observation time can be expected to be greater than the typical operation time scale of the DSA/CR network as a result of the need to observe hundreds/thousands of periods for an accurate estimation (as illustrated in the results obtained in Section IX). This is compatible with the notion that the estimated PU activity statistics will normally be exploited in the long term and once they are estimated for the first time they only need to be updated sporadically, which can be done while the DSA/CR system is in normal operation.

It is also worth mentioning that an accurate estimation of the PU activity statistics requires a careful consideration not only of the sample size N , which is the aspect of interest investigated in this work, but also the sensing period T_s . In the estimation of the moments and related metrics (mean, variance, duty cycle) the impact of the sensing period can be removed by introducing appropriate correction factors as discussed in Section IV. In the estimation of the distribution

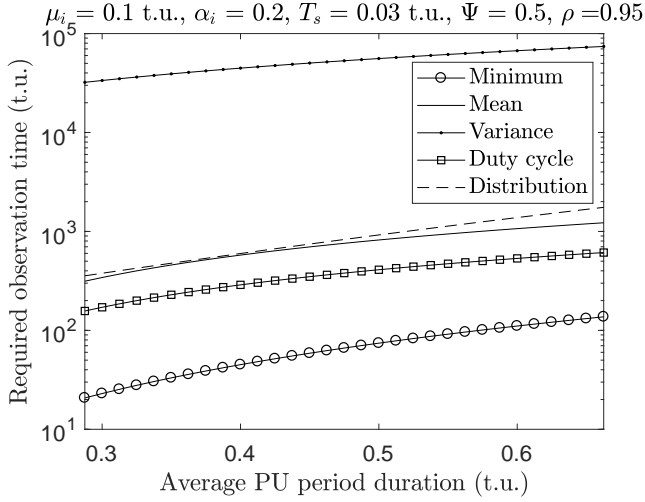


Fig. 10. Required observation time for different PU activity statistics as a function of the average PU period duration (target estimation error of 0.05).

the problem is more challenging since the accuracy of the estimated distribution depends directly on the accuracy of the estimated minimum period (as shown in detail in [22]), which in turn can only be estimated accurately if the sensing period is an integer submultiple of the true (and often unknown) minimum period. To enable a fair evaluation of the individual impact of the sample size, which is the aspect of interest in this work, on the accuracy of the estimated statistics (in particular, the distribution), the sensing period has been selected as an integer submultiple of the true minimum period. In a practical implementation it may be necessary to use more sophisticated methods for an accurate estimation of the minimum period such as those proposed in [38].

Finally, a useful extension of this work would be the investigation of how the obtained analytical results could be extended by including the relative locations of the PU and DSA/CR users and how they would be affected by specific fading scenarios. While this was investigated in [39] for the particular case of the observed duty cycle, the equivalent study for the rest of PU activity statistics considered in this work and the implications on the required observation sample size has not been investigated yet (to the best of the authors' knowledge), which is suggested as future work. Moreover, while the sample size analysis carried out in this work is based on intuitive and simple estimation methods for the considered activity statistics, the investigation of other estimation approaches based on more complex techniques that might potentially lead to accurate estimations with lower sample sizes would be a plausible extension to this work as well. In this context, machine learning techniques [40, 41] have already demonstrated their potential benefits in other areas of wireless networks and their application to the problem here considered is also suggested as further future work.

XI. CONCLUSIONS

DSA/CR systems can benefit from the knowledge of the primary traffic statistics, which can be exploited by DSA/CR

users to achieve a more efficient utilisation of the free primary spectrum. This information can be obtained from spectrum sensing by estimating the duration of individual idle/busy periods (from the sequence of spectrum sensing decisions) and then processing a sufficiently large number of observed periods (referred to in this work as the observation sample size) to calculate relevant statistics on the primary activity such as the minimum period duration, the mean and variance of the observed periods, the channel duty cycle or the underlying distribution. An important practical question is how many periods need to be observed in order to guarantee that the estimated statistics will meet a predefined level of accuracy. In this context, this work has performed a detailed mathematical analysis on the observation sample size required for an accurate estimation of each of the above mentioned primary activity statistics, and provided closed-form expressions for the estimation error as a function of the sample size as well as the required sample size as a function of the desired estimation error. The obtained analytical results have been compared to both simulation and experimental results, showing an excellent agreement in all cases. The expressions provided in this work can be used in practical DSA/CR systems to guarantee that PU activity statistics are estimated to the desired level of accuracy. Moreover, an iterative stopping algorithm has been proposed to enable SU perform a real-time calculation of the required sample size, which has been shown to be very accurate.

APPENDIX

CORRECTION OF SAMPLE MOMENTS

As discussed in Section IV, the sample mean needs no correction and the sample variance needs to be corrected as shown in (14). A similar analysis can be carried out for the third and fourth central moments.

For the third central moment of the estimated periods:

$$\begin{aligned}\mathbb{M}_3(\widehat{T}_i) &= \mathbb{E}[(\widehat{T}_i - \mathbb{E}(\widehat{T}_i))^3] = \mathbb{E}[(T_i - \mathbb{E}(T_i) + T_e - \mathbb{E}(T_e))^3] \\ &= \mathbb{E}[(T_i - \mathbb{E}(T_i))^3] + \mathbb{E}[(T_e - \mathbb{E}(T_e))^3] \\ &= \mathbb{M}_3(T_i) + \mathbb{M}_3(T_e)\end{aligned}\quad (46)$$

Thus the third sample central moment needs no correction since $\mathbb{M}_3(T_e) = 0$ for a symmetric triangular distribution.

For the fourth central moment of the estimated periods:

$$\begin{aligned}\mathbb{M}_4(\widehat{T}_i) &= \mathbb{E}[(\widehat{T}_i - \mathbb{E}(\widehat{T}_i))^4] = \mathbb{E}[(T_i - \mathbb{E}(T_i) + T_e - \mathbb{E}(T_e))^4] \\ &= \mathbb{E}[(T_i - \mathbb{E}(T_i))^4] + 6\mathbb{E}[(T_i - \mathbb{E}(T_i))^2] \cdot \\ &\quad \cdot \mathbb{E}[(T_e - \mathbb{E}(T_e))^2] + \mathbb{E}[(T_e - \mathbb{E}(T_e))^4] \\ &= \mathbb{M}_4(T_i) + 6\mathbb{V}(T_i)\mathbb{V}(T_e) + \mathbb{M}_4(T_e)\end{aligned}\quad (47)$$

where $\mathbb{V}(T_e) = T_s^2/6$ and $\mathbb{M}_4(T_e) = T_s^4/15$ for a symmetric triangular distribution. Solving (47) for $\mathbb{M}_4(T_i)$ it can be shown that a sample estimate $\widehat{c}_{4,i}$ of the fourth central moment needs to be corrected as $\widetilde{c}_{4,i} = \widehat{c}_{4,i} - \widehat{v}_i T_s^2 - T_s^4/15$.

REFERENCES

- [1] C. H. Liu, P. Pawelczak, and D. Cabric, "Primary user traffic classification in dynamic spectrum access networks," *IEEE Journal on Selected Areas in Comms.*, vol. 32, no. 11, pp. 2237–2251, November 2014.
- [2] M. López-Benítez and F. Casadevall, *Spectrum usage models for the analysis, design and simulation of cognitive radio networks*. Dordrecht: Springer Netherlands, 2012, ch. 2, pp. 27–73.

- [3] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Comput. Netw.*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006.
- [4] J. Mitola, *Cognitive radio architecture: The engineering foundations of radio XML*. Hoboken, N.J.: Wiley-Interscience, 2006., 2006.
- [5] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE Journal on Selected Areas in Comms.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [6] M. Wellens, A. D. Baynast, and P. Mahonen, "Performance of dynamic spectrum access based on spectrum occupancy statistics," *IET Communications*, vol. 2, no. 6, pp. 772–782, July 2008.
- [7] K. Umebayashi, Y. Suzuki, and J. J. Lehtomäki, "Dynamic selection of CWmin in cognitive radio networks for protecting IEEE 802.11 primary users," in *Proc. 2011 6th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, June 2011, pp. 266–270.
- [8] 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.
- [9] A. Kaur, S. Sharma, and A. Mishra, "Sensing period adaptation for multiobjective optimisation in cognitive radio using Jaya algorithm," *Electronics Letters*, vol. 53, no. 19, pp. 1335–1336, 2017.
- [10] H. Kim and K. G. Shin, "Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 533–545, May 2008.
- [11] K. Umebayashi, K. Hayashi, and J. J. Lehtomäki, "Threshold-setting for spectrum sensing based on statistical information," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1585–1588, July 2017.
- [12] C. H. Liu, W. Gabran, and D. Cabric, "Prediction of exponentially distributed primary user traffic for dynamic spectrum access," in *Proc. 2012 IEEE Global Communications Conference (GLOBECOM)*, Dec 2012, pp. 1441–1446.
- [13] G. Ding, Y. Jiao, J. Wang, Y. Zou, Q. Wu, Y. D. Yao, and L. Hanzo, "Spectrum inference in cognitive radio networks: Algorithms and applications," *IEEE Communications Surveys Tutorials*, vol. PP, no. 99, pp. 1–1, 2017.
- [14] M. López-Benítez, *Cognitive radio*. Cambridge University Press, 2013, ch. 13, pp. 383–425.
- [15] J. J. Lehtomäki, M. López-Benítez, K. Umebayashi, and M. Juntti, "Improved channel occupancy rate estimation," *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 643–654, March 2015.
- [16] 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.
- [17] A. Al-Tahmeesschi, M. López-Benítez, K. Umebayashi, and J. Lehtomäki, "Analytical study on the estimation of primary activity distribution based on spectrum sensing," in *Proc. 28th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2017), Workshop on Cognitive Radio and Innovative Spectrum Sharing Paradigms for Future Networks (CRAFT 2017)*, Montreal, Quebec, Canada, 8–13 Oct. 2017, pp. 1–5.
- [18] M. López-Benítez, "Can primary activity statistics in cognitive radio be estimated under imperfect spectrum sensing?" in *Proc. 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 750–755.
- [19] A. Al-Tahmeesschi, M. López-Benítez, J. Lehtomäki, and K. Umebayashi, "Improving primary statistics prediction under imperfect spectrum sensing," in *Proc. 2018 IEEE Wireless Comms. & Networking Conf. (WCNC), 4th Int'l. Workshop of Smart Spectrum (IWSS 2018)*, Apr. 2018, pp. 1–6.
- [20] M. López-Benítez and J. Lehtomäki, "On the sensing sample size for the estimation of primary channel occupancy rate in cognitive radio," in *Proc. 2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–6.
- [21] M. López-Benítez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2091–2104, Jun 2013.
- [22] A. Al-Tahmeesschi, M. López-Benítez, J. Lehtomäki, and K. Umebayashi, "Accurate estimation of primary user traffic based on periodic spectrum sensing," in *Proc. 2018 IEEE Wireless Comms. & Networking Conf. (WCNC)*, Apr. 2018, pp. 1–6.
- [23] A. Mood, F. Graybill, and D. Boes, *Introduction to the Theory of Statistics*. McGraw-Hill, 1974.
- [24] I. R. Savage, "Probability inequalities of the Tchebycheff type," *Journal of Research of the National Bureau of Standards-B. Mathematics and Mathematical Physics B*, vol. 65, no. 3, pp. 211–222, 1961.
- [25] H. N. Phien, "On the computation of sample central moments," *Int'l. J. Mathematical Education in Science and Tech.*, vol. 19, no. 3, pp. 403–412, May 1988.
- [26] I. Hughes and T. P. A. Hase, *Measurements and their uncertainties: A practical guide to modern error analysis*. Oxford University Press, 2010.
- [27] H. H. Ku, "Notes on the use of propagation of error formulas," *Journal of Research of the National Bureau of Standards. Section C: Engineering and Instrumentation*, vol. 70C, no. 4, pp. 263–273, Oct. 1966.
- [28] Q. Liang, M. Liu, and D. Yuan, "Channel estimation for opportunistic spectrum access: uniform and random sensing," *IEEE Trans. Mobile Computing*, vol. 11, no. 8, pp. 1304–1316, Aug. 2012.
- [29] P. Tehrani, L. Tong, and Q. Zhao, "Asymptotically efficient multichannel estimation for opportunistic spectrum access," *IEEE Trans. Signal Processing*, vol. 60, no. 10, pp. 5347–5360, Oct. 2012.
- [30] W. Saad, Z. Han, H. V. Poor, T. Basar, and J. B. Song, "A cooperative bayesian nonparametric framework for primary user activity monitoring in cognitive radio networks," *IEEE J. Sel. Areas Comms.*, vol. 30, no. 9, pp. 1815–1822, Oct. 2012.
- [31] S. Geirhofer, L. Tong, and B. M. Sadler, "A measurement-based model for dynamic spectrum access in WLAN channels," in *Proc. IEEE Military Comms. Conf. (MILCOM 2006)*, Oct. 2006, pp. 1–7.
- [32] L. Stabellini, "Quantifying and modeling spectrum opportunities in a real wireless environment," in *Proc. IEEE Wireless Comms. and Networking Conf. (WCNC 2010)*, Apr. 2010, pp. 1–6.
- [33] M. Wellens, J. Riihijärvi, and P. Mähönen, "Empirical time and frequency domain models of spectrum use," *Physical Comm.*, vol. 2, no. 1–2, pp. 10–32, Mar. 2009.
- [34] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*, 2nd ed. Wiley, Nov. 1994, vol. 1.
- [35] L. Zhang, "Sample mean and sample variance: Their covariance and their (in)dependence," *The American Statistician*, vol. 61, no. 2, pp. 159–160, 2007.
- [36] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes: The art of scientific computing*, 3rd ed. Cambridge University Press, 2007.
- [37] 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. 2017 IEEE Wireless Comms. & Networking Conf. (WCNC)*, March 2017, pp. 1–6.
- [38] M. López-Benítez, A. Al-Tahmeesschi, and D. Patel, "Accurate estimation of the minimum primary channel activity time in cognitive radio based on periodic spectrum sensing observations," in *Proc. 24th European Wireless Conference (EW 2018)*, May 2018, pp. 131–136.
- [39] M. López-Benítez and F. Casadevall, "Space-dimension models of spectrum usage for cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 306–320, Jan. 2017.
- [40] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," Available: <https://arxiv.org/abs/1710.02913>, Oct. 2017.
- [41] O. Simeone, "A brief introduction to machine learning for engineers," Available: <https://arxiv.org/abs/1709.02840>, Sep. 2017.

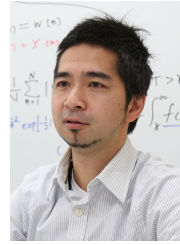


Ahmed Al-Tahmeesschi (S'16) received the B.Sc. and M.Sc. degrees in communications engineering from the University of Technology, Baghdad, Iraq, in 2009 and 2011, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K. His research interests include cognitive radio networks, dynamic spectrum access techniques, algorithm design and machine learning.



Miguel López-Benítez (S'08, M'12, SM'17) received the B.Sc. and M.Sc. degrees in Telecommunication Engineering (both with Distinction) from Miguel Hernández University, Elche, Spain in 2003 and 2006, respectively, and a Ph.D. degree in Telecommunication Engineering (2011 Outstanding Ph.D. Thesis Award) from the Technical University of Catalonia, Barcelona, Spain in 2011. From 2011 to 2013, he was a Research Fellow in the Centre for Communication Systems Research of the University of Surrey, Guildford, UK. Since 2013, he has been

a Lecturer (Assistant Professor) in the Department of Electrical Engineering and Electronics of the University of Liverpool, UK. His research interests include the field of wireless communications and networking, with special emphasis on cellular mobile communications and dynamic spectrum access in cognitive radio systems. He is/has been the principal investigator or co-investigator of research projects funded by the EPSRC, British Council and Royal Society, and has been involved in the European-funded projects AROMA, NEWCOM++, FARAMIR, QoS MOS and CoRaSat. He is Associate Editor of IEEE Access, IET Communications, and Wireless Communications and Mobile Computing, and has been a member of the Organising Committee for the IEEE WCNC International Workshop on Smart Spectrum (IWSS 2015-18). Please visit <http://www.lopezbenitez.es> for more details.



Kenta Umebayashi (S'00, M'04) received the LL.B. degree from Ritsumeikan University in 1996 and the B.E., M.E., and Ph.D. degrees from the Yokohama National University in 1999, 2001, and 2004, respectively. From 2004 to 2006, he was a research scientist at the University of Oulu, Centre for Wireless Communications (CWC). He is currently an associate professor at the Tokyo University of Agriculture and Technology. He was an associate editor of IEICE Transaction on Communications from May 2015 to May 2017. He was a principal

investigator in the four Grants-in-Aid for Scientific Research projects and three Strategic Information and Communications R&D Promotion Programme projects. His research interests lie in the areas of signal detection and estimation theories for wireless communication, signal processing for multiple antenna systems, cognitive radio networks, and terahertz band wireless communication. He received the Best Paper Award at IEEE WCNC 2012 for a paper he authored, and the Best Paper Award at IEEE WCNC workshop IWSS 2015 for a paper he co-authored.



Dhaval K Patel is currently working as Assistant Professor at School of Engineering and Applied Science - Ahmedabad University, India since 2014. He was also a visiting faculty at Franklin W. Olin College of Engineering-Massachusetts, USA. He worked as Junior Research Fellow in the Post Graduate Lab for Communication Systems at the Nirma University-India from 2011 to 2014. He received the B.E. and M.E. degrees in Communication Systems Engineering (both with Distinction/First-Class) from Gujarat University in 2003 and 2010, respectively,

and a Ph.D. degree in Electronics and Communications from the Institute of Technology-Nirma University, India in 2014. His research area of interest includes Vehicular Cyber Physical System, 5G Wireless Networks, Non-parametric statistics and Physical Layer Security. He is the principal investigator of research projects funded by Department of Science and Technology (DST), UK-India Education and Research Initiative (UKIERI), Association of Southeast Asian Nations (ASEAN)-India Collaborative R&D Project and Gujarat Council on Science and Technology (GUJCOST).



Janne Lehtomäki (S'03, M'06) received the Ph.D. degree from the University of Oulu, Oulu, Finland, in 2005. He was with Georgia Tech, Atlanta, GA, USA, as a Visiting Scholar, in 2013. He is currently an Adjunct Professor with the Centre for Wireless Communications, University of Oulu. His current research interest focuses on terahertz band wireless communication. Dr. Lehtomäki was a co-recipient of the Best Paper Award of IEEE WCNC, 2012. He is an Editorial Board member of Physical Communication. He was a General Co-Chair of the IEEE WCNC

2017 & 2018 International Workshop on Smart Spectrum, a TPC Co-Chair of the IEEE WCNC 2015 and 2016 International Workshop on Smart Spectrum, and a Publicity/Publications Co-Chair of ACM NANOCOM in 2015-2018. He has served as a Guest Associate Editor for two IEICE Transactions on Communications Special Sections and as a Managing Guest Editor of the Nano Communication Networks Special Issue in 2016.