Methods for Fast Estimation of Primary Activity Statistics in Cognitive Radio Systems

Miguel López-Benítez†‡, Ogeen H. Toma†, Dhaval K. Patel* and Kenta Umebayashi§
†Department of Electrical Engineering and Electronics, University of Liverpool, United Kingdom
‡ARIES Research Centre, Antonio de Nebrija University, Spain
*School of Engineering and Applied Science, Ahmedabad University, India
§Graduate School of Engineering, Tokyo University of Agriculture and Technology, Japan
Email: m.lopez-benitez@liverpool.ac.uk, ogeen.toma@liverpool.ac.uk, dhaval.patel@ahduni.edu.in, ume_k@cc.tuat.ac.jp

***Abstract*—Cognitive Radio (CR) is aimed at increasing the efficiency of spectrum utilisation by allowing unlicensed users to opportunistically access licensed spectrum bands during the inactivity periods of the licensed users. CR systems can benefit from an accurate knowledge of the spectrum occupancy patterns and their statistical properties. This statistical information can be obtained by periodically monitoring (sensing) the idle/busy state of the licensed channels. However, a reliable estimation of the primary activity statistics may require long observations times. This work proposes efficient methods to reduce the observation time required to produce a reliable estimation of the primary activity statistics. Furthermore, a method enabling CR users to quantify the accuracy of the estimated statistics is also proposed. Compared to other existing approaches, the proposed methods can provide accurate estimations of the primary activity statistics in significantly shorter observation times, thus allowing CR users to quickly adapt to new unknown operating channels.**

***Index Terms*—Cognitive radio, opportunistic spectrum access, spectrum occupancy modelling, primary activity statistics.**

I. Introduction

The Dynamic Spectrum Access (DSA) / Cognitive Radio (CR) paradigm [1], [2] has the potential to improve the currently inefficient exploitation of the radio spectrum that results from inflexible management policies. The basic underlying idea of DSA/CR is to allow unlicensed (secondary) users to opportunistically access allocated spectrum bands during the inactivity periods of the licensed (primary) users.

Owing to the opportunistic nature of this spectrum access paradigm, the performance of DSA/CR systems depends on the spectrum occupancy patterns of the licensed users. An accurate knowledge of such patterns and their statistical properties can therefore be useful to DSA/CR systems. The knowledge of the primary activity statistics can be exploited to predict future trends in the spectrum occupancy [3], decide the most convenient allocation of radio resources [4] and take appropriate actions to optimise the system performance and improve the overall spectrum efficiency [5].

The activity statistics of the primary channel are initially unknown to the DSA/CR system but can be estimated based on the sequence of spectrum sensing decisions, which can be used to estimate the durations of individual idle/busy periods and subsequently their moments, underlying distributions and other relevant statistics [6]. Unfortunately, an accurate estimation may require a long observation time until a sufficiently high number of samples (i.e., observed period durations) is available [7]. From a practical point of view, this means that a DSA/CR system may need a long period of adaptation every time the carrier frequency needs to switch to a new unknown radio channel or spectrum band. During such period of adaptation, advanced DSA/CR methods relying on primary activity statistics may experience a degraded performance. Adequate methods capable to shorten the observation time required to obtain reliable statistical information of an unknown channel can therefore help DSA/CR systems preserve the system performance after a spectrum handover. In this context, this paper proposes novel methods to estimate the activity statistics of a channel based on spectrum sensing observations. Closed-form expressions for the analysis, design and practical implementation of the proposed methods are derived and corroborated with simulations. Compared to other existing approaches, the proposed methods are able to produce statistically reliable estimations within shorter observation times.

The rest of this paper is organised as follows. First, Section II describes how primary activity statistics can be estimated based on spectrum sensing observations based on the direct calculation of the empirical distribution and the method of moments. The proposed methods are presented in Section III along with their corresponding analysis of required observation time (sample size) and resulting accuracy. The performance of the proposed methods is assessed and discussed in Section IV. Finally, Section V summarises and concludes the paper.

II. Estimation of primary activity statistics

Without loss of generality, this work focuses on the estimation of the distribution of idle/busy period durations. Other relevant statistics such as the minimum, mean (and other moments) or duty cycle can be derived from the estimated distribution. The distribution of period durations can be estimated accurately provided that a sufficiently large set of samples is available. The term *sample* is used in this work to refer to the duration of an observed period, estimated from the sequence of observed channel states (i.e., spectrum sensing decisions).

The duration of the channel idle/busy periods can be estimated from spectrum sensing decisions as illustrated in Fig. 1. DSA/CR users sense the channel with a finite sensing period T_s (assumed to be shorter than the minimum period duration). In every sensing event, a binary decision on the idle ($\mathcal{H}_0$) or busy ($\mathcal{H}_1$) state of the channel is made. Every time the observed channel state changes, the time interval elapsed since the last state change is computed as shown in Fig. 1 in order

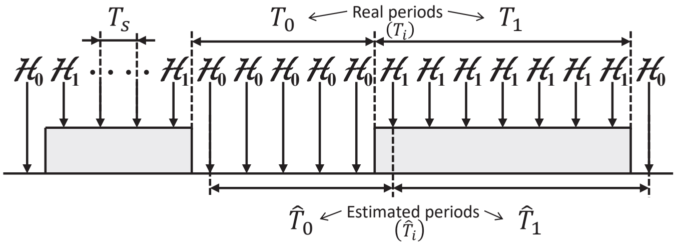

Fig. 1. Estimation of period durations from spectrum sensing decisions.

to make an estimation $\widehat{T}_i$ of the real idle/busy period duration T_i ($i = 0$ for idle periods, $i = 1$ for busy periods). Making use of the observed period durations $\widehat{T}_i$ (i.e., *samples* henceforth), the underlying statistical distribution can be estimated [6].

Various methods to estimate the distribution of the primary period durations can be employed as investigated in [6]. Given a set $\widehat{\mathcal{T}}_i = \{\widehat{T}_{i,n}\}_{n=1}^N$ of N observed period durations, an estimation, $F_{\widehat{T}_i}(\widehat{T})$, of the original Cumulative Distribution Function (CDF), $F_{T_i}(T)$, can be computed as follows:

$$F_{\widehat{T}_i}(\widehat{T}) = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{\widehat{\mathcal{T}}_i(\widehat{T})}\{\widehat{T}_{i,n}\} = \frac{|\widehat{\mathcal{T}}_i(\widehat{T})|}{N} \tag{1}$$

where $|\widehat{\mathcal{T}}_i(\widehat{T})|$ indicates the cardinality (number of elements) of $\widehat{\mathcal{T}}_i(\widehat{T}) = \{\widehat{T}_{i,n} : \widehat{T}_{i,n} \leq \widehat{T}, n = 1, \ldots, N\}$ (the subset of period durations lower than or equal to $\widehat{T}$), and $\mathbf{1}_A\{x\}$ is the indicator function of subset A, which is equal to one for the elements $x \in A$ and zero otherwise. Since the estimated period durations are integer multiples of the sensing period (i.e., $\widehat{T}_{i,n} = kT_s$, $k \in \mathbb{N}^+$), the distribution estimated in (1) is discrete. As a result, the direct calculation of the empirical CDF as shown in (1) may lead to significant estimation errors since the distribution of the original periods, $F_{T_i}(T)$, will in general be continuous. The accuracy of this estimation method was investigated analytically and experimentally in [6] and it was shown that the estimated distribution may differ significantly from the true distribution as a result of the limited time resolution imposed by the use of a finite sensing period.

An alternative is to make an assumption on the underlying distribution and then estimate its parameters based on statistical inference methods [6]. A commonly employed assumption is that idle/busy periods are exponentially distributed, meaning that the CDF of period durations, $F_{T_i}(T)$, could be writen as:

$$F_{T_i}^E(T) = 1 - e^{-\lambda_i^E (T - \mu_i^E)}, \quad T \geq \mu_i^E \tag{2}$$

where T_i represents the period type ($i \in \{0, 1\}$), T is the period duration, and $\mu_i^E > 0$ and $\lambda_i^E > 0$ are the location and rate parameters, respectively. Field measurements, however, have shown that the channel activity statistics of real systems are more accurately described by means of the Generalised Pareto (GP) distribution [8]–[10], whose CDF is given by:

$$F_{T_i}^{GP}(T) = 1 - \left[1 + \frac{\alpha_i^{GP}(T - \mu_i^{GP})}{\lambda_i^{GP}}\right]^{\frac{-1}{\alpha_i^{GP}}}, \quad T \geq \mu_i^{GP} \tag{3}$$

where $\mu_i^{GP} > 0$, $\lambda_i^{GP} > 0$, $\alpha_i^{GP} \in \mathbb{R}$ are the location, scale and shape parameters, respectively. These estimation strategies lead to a continuous estimated distribution, whose accuracy depends on the suitability of the assumed distribution model in describing the true distribution of period durations, $F_{T_i}(T)$.

TABLE I
PROPOSED DISTRIBUTION ESTIMATION METHODS

Method	$\widetilde{\mu}_i^E$	$\widetilde{\lambda}_i^E$
MMK	μ_i	$1/(\widetilde{m}_i - \mu_i)$
VMK	μ_i	$1/\sqrt{\widetilde{v}_i}$
MMU	$\min_n(\{\widehat{T}_{i,n}\}_{n=1}^N)$	$1/(\widetilde{m}_i - \widetilde{\mu}_i^E)$
VMU	$\min_n(\{\widehat{T}_{i,n}\}_{n=1}^N)$	$1/\sqrt{\widetilde{v}_i}$
MV	$\widetilde{m}_i - \sqrt{\widetilde{v}_i}$	$1/(\widetilde{m}_i - \widetilde{\mu}_i^E)$ or $1/\sqrt{\widetilde{v}_i}$

III. FAST ESTIMATION OF PRIMARY ACTIVITY STATISTICS

A. Proposed Methods

Based on the conclusions from previous work based on field measurements [8]–[10], the work reported in [6] analysed the estimation of the parameters μ_i^{GP}, λ_i^{GP} and α_i^{GP} of the model in (3) based on the Method of Moments (MoM)[1]. By contrast, this work analyses the performance of MoM-based methods under the assumption of exponentially distributed period durations as described by the model in (2) and shows the potential benefits of this alternative approach in terms of its lower sample size required to produce an accurate estimation.

Based on the sample mean $\widetilde{m}_i$ and sample variance $\widetilde{v}_i$ of the set of observed periods $\widehat{\mathcal{T}}_i$, given by [6, eqs. (21b) & (22)][2]:

$$\widetilde{m}_i = \frac{1}{N}\sum_{n=1}^{N} \widehat{T}_{i,n} \tag{4a}$$

$$\widetilde{v}_i = \frac{1}{N-1}\sum_{n=1}^{N} \left(\widehat{T}_{i,n} - \widetilde{m}_i\right)^2 - \frac{T_s^2}{6} \tag{4b}$$

an estimation $\widetilde{\mu}_i^E$ and $\widetilde{\lambda}_i^E$ of the parameters of the model in (2) can be made following various strategies (see Table I). When the true minimum period duration $\mu_i = \min(T_i)$ is known, then $\widetilde{\mu}_i^E = \mu_i$ and the rate parameter can be estimated from the sample mean (Mean-based with Minimum Known, MMK) or sample variance (Variance-based with Minimum Known, VMK). If the true minimum period duration is unknown, it can be estimated as $\widetilde{\mu}_i^E = \min_n(\{\widehat{T}_{i,n}\}_{n=1}^N)$, and the rate parameter can be estimated again based on the sample mean (Mean-based with Minimum Unknown, MMU) or sample variance (Variance-based with Minimum Unknown, VMU). Alternatively, when the minimum period duration is unknown, μ_i^E and λ_i^E can also be estimated solely from the sample mean and sample variance (Mean- and Variance-based, MV).

[1]The main advantage of MoM-based approaches is that the distribution parameters can be estimated from sample moments, which can be computed recursively based on the last sample [11], while other inference methods in general need the whole history of past observed period durations and therefore have much larger memory requirements in practical implementations.

[2]As the values of the observed periods $\widehat{T}_i$ are affected by the finite sensing period T_s, their sample moments are affected by T_s as well. A correction factor $T_s^2/6$ is required in (4b) to minimise the impact of T_s on $\widetilde{v}_i$ [6].

B. Required Observation Time (Sample Size)

This section provides closed-form expressions that can be used to evaluate, for each of the estimation strategies shown in Table I, the observation time required to produce a reliable estimation of the underlying distribution of period durations, assumed to be exponential as described by the model in (2). The methods shown in Table I make an estimation of the CDF based on the sample moments $\widetilde{m}_i$ and $\widetilde{v}_i$. As the number N of observed period durations (i.e., the sample size) increases, the estimated sample moments become more accurate and the estimated distribution converges asymptotically to the assumed exponential CDF given by (2). For a finite sample size, the sample moments have a certain estimation error, which also affects the estimated distribution. The estimated distribution $F_{\widehat{T}_i}(T) = 1 - e^{-\widetilde{\lambda}_i^E (T - \widetilde{\mu}_i^E)}$ can be considered to be reliable if it differs from the assumed distribution model $F_{T_i}^E(T)$ in (2) by no more than a predefined margin, which can be quantified in terms of the Kolmogorov-Smirnov (KS) distance as [12, eq. (14.3.17)]:

$$\begin{aligned} D_i^{KS} &= \sup_T \left| F_{T_i}^E(T) - F_{\widehat{T}_i}(T) \right| \\ &= \left| F_{T_i}^E\left(T_i^{KS}\right) - F_{\widehat{T}_i}\left(T_i^{KS}\right) \right| \\ &= \left| e^{-\lambda_i^E\left(T_i^{KS} - \mu_i^E\right)} - e^{-\widetilde{\lambda}_i^E\left(T_i^{KS} - \widetilde{\mu}_i^E\right)} \right| \end{aligned} \tag{5}$$

where T_i^{KS} is the value of the period duration T that maximises the absolute difference shown in (5), i.e., $T_i^{KS} = \{T : |F_{T_i}^E(T) - F_{\widehat{T}_i}(T)| = \sup_T |F_{T_i}^E(T) - F_{\widehat{T}_i}(T)|\}$.

The KS distance can be related to the sample size by first obtaining T_i^{KS} for each estimation method and then expressing the parameter estimates $\widetilde{\mu}_i^E$ and $\widetilde{\lambda}_i^E$ as a function of $\widetilde{m}_i$ and $\widetilde{v}_i$ (as shown in Table I), whose maximum relative errors are related to the sample size N as [7, eqs. (19) & (20)]:

$$\varepsilon_{r,\max}^{\widetilde{m}_i} \approx \frac{\sqrt{2}\,\mathrm{erf}^{-1}(\rho)}{\mathbb{E}(T_i)} \left[\frac{1}{N} \left(\mathbb{V}(T_i) + \frac{T_s^2}{6} \right) \right]^{\frac{1}{2}} \tag{6a}$$

$$\begin{aligned} \varepsilon_{r,\max}^{\widetilde{v}_i} \approx \frac{\sqrt{2}\,\mathrm{erf}^{-1}(\rho)}{\mathbb{V}(T_i)} \Bigg[\frac{1}{N} \Bigg(& \mathbb{M}_4(T_i) - \frac{N-3}{N-1} [\mathbb{V}(T_i)]^2 \\ & + \frac{2N}{3(N-1)} T_s^2 \mathbb{V}(T_i) + \frac{7N+3}{180(N-1)} T_s^4 \Bigg) \Bigg]^{\frac{1}{2}} \end{aligned} \tag{6b}$$

where ρ is the desired confidence interval, $\mathbb{E}(T_i) = \mu_i^E + 1/\lambda_i^E$, $\mathbb{V}(T_i) = 1/(\lambda_i^E)^2$ and $\mathbb{M}_4(T_i) = 9/(\lambda_i^E)^4$ are the mean, variance and fourth central moment of the period durations, respectively, assuming exponentially distributed periods.

For the methods where the minimum period is known (i.e., MMK and VMK), T_i^{KS} can be found by computing the value of T for which $d[F_{T_i}^E(T) - F_{\widehat{T}_i}(T)]/dT = 0$ and $d^2[F_{T_i}^E(T) - F_{\widehat{T}_i}(T)]/dT^2 < 0$, which yields:

$$T_i^{KS} = \mu_i^E + \frac{\ln \lambda_i^E - \ln \widetilde{\lambda}_i^E}{\lambda_i^E - \widetilde{\lambda}_i^E} \tag{7}$$

The KS distance in (5) becomes maximum (i.e., worst possible case) when $\widetilde{m}_i = \mathbb{E}(T_i)(1 - \varepsilon_{r,\max}^{\widetilde{m}_i})$ for the MMK method and when $\widetilde{v}_i = \mathbb{V}(T_i)(1 + \varepsilon_{r,\max}^{\widetilde{v}_i})$ for the VMK method.

For the methods where the minimum period is unknown (i.e., MMU, VMU and MV), it can be shown that $T_i^{KS} = \mu_i^E = \mu_i$ and (5) then simplifies to $D_i^{KS} = 1 - e^{-\widetilde{\lambda}_i^E(\mu_i - \widetilde{\mu}_i^E)}$. In the case of the MMU and VMU methods, the estimated minimum is related to the true minimum as $\widetilde{\mu}_i^E = \lfloor \mu_i / T_s \rfloor T_s$ [6, eq. (7)], where $\lfloor \cdot \rfloor$ represents the floor operator. The KS distance in (5) becomes maximum (i.e., worst possible case) when $\widetilde{m}_i = \mathbb{E}(T_i)(1 - \varepsilon_{r,\max}^{\widetilde{m}_i})$ for the MMU method, $\widetilde{v}_i = \mathbb{V}(T_i)(1 - \varepsilon_{r,\max}^{\widetilde{v}_i})$ for the VMU method, and $\widetilde{m}_i = \mathbb{E}(T_i)(1 - \frac{1}{2}\varepsilon_{r,\max}^{\widetilde{m}_i})$ and $\widetilde{v}_i = \mathbb{V}(T_i)(1 + \frac{1}{2}\varepsilon_{r,\max}^{\widetilde{v}_i})$ for the MV method.

These results can be used to evaluate D_i^{KS} as a function of the sample size N for each estimation method by introducing the appropriate T_i^{KS}, then expressing the parameter estimates $\widetilde{\mu}_i^E$ and $\widetilde{\lambda}_i^E$ as a function of $\widetilde{m}_i$ and $\widetilde{v}_i$ as shown in Table I, and finally expressing the sample moments ($\widetilde{m}_i$ and $\widetilde{v}_i$) as a function of the population moments ($\mathbb{E}(T_i)$ and $\mathbb{V}(T_i)$) and relative errors ($\varepsilon_{r,\max}^{\widetilde{m}_i}$ and $\varepsilon_{r,\max}^{\widetilde{v}_i}$) as indicated in each case, where the dependency on the sample size N is included in $\varepsilon_{r,\max}^{\widetilde{m}_i}$ and $\varepsilon_{r,\max}^{\widetilde{v}_i}$ as shown in (6). Notice that the algebraic complexity of the final analytical result prevents the derivation of a closed-form expression for N as a function of D_i^{KS}, which would be useful for a direct evaluation of the observation time (i.e., sample size) required to achieve a certain level of accuracy. However, by means of a numerical evaluation of D_i^{KS} in (5), which depends implicitly on N, the desired result can be obtained numerically.

C. Accuracy of the Estimated Distribution

As discussed in Section II, the real period durations are more accurately described by a generalised Pareto distribution than an exponential distribution. Consequently, the assumption of exponentially distributed periods leads to some error in the estimated distribution (exponential), which can be quantified in terms of the KS distance with respect to the true distribution (generalised Pareto) as follows:

$$\begin{aligned} D_i^{KS} &= \sup_T \left| F_{T_i}^{GP}(T) - F_{T_i}^E(T) \right| \\ &= \left| F_{T_i}^{GP}\left(T_i^{KS}\right) - F_{T_i}^E\left(T_i^{KS}\right) \right| \\ &= \left| \left[1 + \frac{\alpha_i^{GP}(T_i^{KS} - \mu_i^{GP})}{\lambda_i^{GP}} \right]^{\frac{-1}{\alpha_i^{GP}}} - e^{-\lambda_i^E(T_i^{KS} - \mu_i^E)} \right| \end{aligned} \tag{8}$$

which represents the KS distance between (2) and (3). The period duration T for which the maximum absolute difference is attained (i.e., T_i^{KS}) can be calculated as the value of T for which $d[F_{T_i}^{GP}(T) - F_{T_i}^E(T)]/dT = 0$ and $d^2[F_{T_i}^{GP}(T) - F_{T_i}^E(T)]/dT^2 < 0$. The resulting equation cannot be solved analytically as a result of the presence of some non-polynomial terms, however such terms can be approximated by their corresponding second-order Taylor series centred about $\mathbb{E}(T_i)$, which leads to the following tight approximation:

$$T_i^{KS} \approx \mathbb{E}(T_i) + [\mathbb{E}(T_i) - \mu_i] \times \frac{\left(1 - \alpha_i^{GP}\right)^{\frac{1}{\alpha_i^{GP}}} \left(1 + \alpha_i^{GP}\right) \left[1 - \sqrt{\left(1 + \alpha_i^{GP}\right)\left(1 + 2\alpha_i^{GP}\right)} \right]}{\left(1 - \alpha_i^{GP}\right)^{\frac{1}{\alpha_i^{GP}}} \left(1 + \alpha_i^{GP}\right)\left(1 + 2\alpha_i^{GP}\right) - 1/e} \tag{9}$$

The result in (8)–(9) can be used in the design of DSA/CR systems to quantify the best estimation accuracy that can be attained by the proposed estimation methods in Table I for a particular operation scenario. Moreover, the result in (8)–(9) can also be used by DSA/CR devices in a practical scenario to determine the accuracy of the estimation provided by the proposed methods. To this end, the parameters involved in (8)–(9), which would be unknown in a real operation scenario, can be evaluated based on their corresponding sample estimates according to the considered estimation method (see Table I and [6]). The resulting KS distance would then provide DSA/CR devices with an estimation of the maximum deviation between the so-far estimated distribution and the real one (assumed to be GP). Based on this metric, DSA/CR devices would thus be able to determine whether the distribution estimated with the available sample set is accurate enough for a particular application or if more *samples* (i.e., observed periods) would be required to reach the desired level of accuracy.

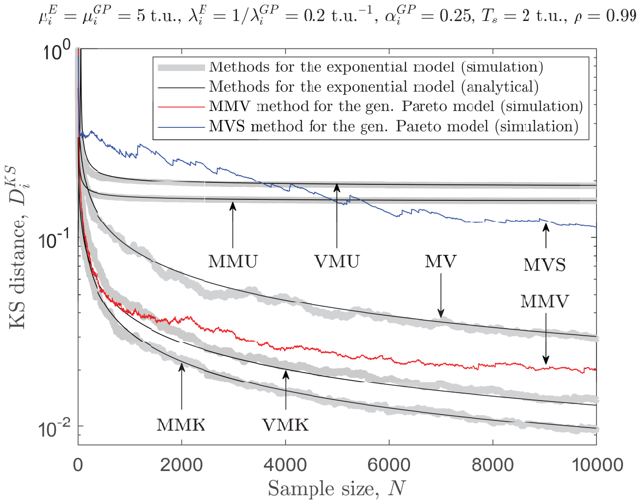

Fig. 2. KS distance as a function of the sample size.

IV. Performance Results

The performance of the proposed methods was evaluated based on two aspects, namely the sample size required to produce a reliable estimation (representative of the required observation time) and the resulting estimation accuracy.

A. Required Observation Time (Sample Size)

For the evaluation of the first aspect (sample size required to produce a reliable estimation assuming exponentially distributed periods), the analytical result in (5) was evaluated numerically for each of the proposed methods and compared with the counterpart results obtained from simulations, which were obtained based on the following steps:

1) Generate a set $\{T_{i,n}\}_{n=1}^N$ of N periods obtained as random numbers drawn from an exponential distribution (2) with the desired location (μ_i^E) and rate (λ_i^E) parameters.
2) From the sequence of periods obtained in step 1, determine the sequence of idle/busy states ($\mathcal{H}_0/\mathcal{H}_1$) that would be observed in the primary channel when a sensing period T_s is employed by the DSA/CR devices.
3) Based on the sequence $\mathcal{H}_0/\mathcal{H}_1$ obtained in step 2, compute, as depicted in Figure 1, the set $\{\widehat{T}_{i,n}\}_{n=1}^N$ of N period durations that would be observed by the DSA/CR system based on the spectrum sensing outcomes.
4) Based on the set $\{\widehat{T}_{i,n}\}_{n=1}^N$ obtained in step 3, estimate the sample mean $\widetilde{m}_i$ and sample variance $\widetilde{v}_i$ of the set of observed periods as indicated in (4).
5) Estimate the location ($\widetilde{\mu}_i^E$) and rate ($\widetilde{\lambda}_i^E$) parameters of the exponential model in (2) based on the sample mean and sample variance obtained in step 4 for each of the proposed methods as shown in Table I.
6) Evaluate the KS distance in (5) based on the true parameter values from step 1 (μ_i^E, λ_i^E) and the estimates from step 5 ($\widetilde{\mu}_i^E$, $\widetilde{\lambda}_i^E$) for the considered sample size N.

The simulation process described above was repeated for several values of the sample size parameter within the range $N \in [1, 10^4]$. For each value of N, simulation steps 1–6 were repeated for a total of 1000 iterations and the KS distance observed at the 99th percentile (i.e., $\rho = 0.99$) was selected. The same confidence interval was employed in the numerical evaluation of the analytical result in (5).

Fig. 2 evaluates (5) for the methods shown in Table I and corroborates its validity with simulation results (parameters are expressed in relative time units, t.u.). As it can be appreciated, the shortest observation intervals (i.e., sample size values) for a predefined KS distance are attained by the MMK and VMK methods. The observation interval required by MMK is lower than that required by VMK because it relies on a lower-order sample moment (i.e., the mean, which is a 1st-order moment, as opposed to the variance, which is a 2nd-order moment), thus requires a lower sample size to provide a similar estimation error [7]. This makes MMK a preferred option compared to VMK. Unfortunately, MMK can only be employed when the true minimum period μ_i is known. If μ_i is unknown, other methods must be used. In general, the MMU and VMU methods lead to a fixed and irreducible error that is independent of the sample size and can be significantly high, as shown in Fig. 2, depending on the particular relation between the true minimum period μ_i and the employed sensing period T_s. It can be shown that the estimation error of the MMU and VMU methods can be zero when the sensing period is an integer sub-multiple of the minimum period duration (i.e., $T_s = \mu_i/k$, with $k \in \mathbb{N}^+$). However, such ratio is unlikely to occur in practice and the estimation error provided by the MMU and VMU methods will in general be relatively high as shown in the example of Fig. 2. A more convenient approach when μ_i is unknown is the MV method, which guarantees a reliable estimation when μ_i is unknown at the expense of a higher observation interval compared to MMK.

If periods are assumed to be GP-distributed, the parameters μ_i^{GP}, λ_i^{GP} and α_i^{GP} of the model in (3) can be estimated based on the sample minimum, mean and variance (MMV method) if μ_i is known, or the sample mean, variance and skewness (MVS method) if μ_i is unknown (see [6] for details). The KS distance between the GP distribution estimated with these methods and the model in (3) as a function of the sample size have been included in Fig. 2 for comparison. As it can be appreciated, the MMK and MV methods provide reliable estimations within shorter observation intervals than their GP counterparts (i.e., the MMV and MVS methods, respectively). For instance, when the minimum period is known (i.e., either MMK or MMV can be used), MMK requires $N \approx 2500$ samples to achieve a target error $D_i^{KS} = 0.02$ as opposed to $N \approx 10000$ samples required by MMV for the same level of accuracy (i.e., MMK reduces the sample size required by MMV by about 75%). Similarly, if the minimum period is unknown (i.e., either MV or MVS need to be used), MV requires $N \approx 750$ samples to achieve a target error $D_i^{KS} = 0.10$ while MVS requires more than $N \approx 10000$ samples for the same level of accuracy (i.e., MV reduces the sample size required by MVS by more than 90%). Therefore, the proposed estimation methods can achieve a reliable estimation of the distribution of period durations (under the assumption of exponentially distributed periods) within significantly shorter observation intervals than those required by the estimation methods studied in [6] (under the assumption of GP-distributed periods).

B. Accuracy of the Estimated Distribution

For the evaluation of the second aspect (estimation accuracy resulting from the assumption of exponentially distributed periods), the analytical result in (8) was evaluated numerically (to produce an exact result) and based on the approximation of (9). The results are shown in Fig. 3 as a function of the GP distribution shape parameter (α_i^{GP}). Note that the exponential distribution in (2) is a particular case of the GP distribution in (3) with $\mu_i^{GP} = \mu_i^E$, $\lambda_i^{GP} = 1/\lambda_i^E$ and $\alpha_i^{GP} = 0$; therefore the divergence between both distribution models is mainly determined by the value of α_i^{GP}. Real values of α_i^{GP} for common radio technologies in practical scenarios have been observed to be within the range $\alpha_i^{GP} \in [0, 0.25]$ [10], which is the interval considered in Fig. 3. In general, the accuracy of the proposed methods (in particular, MMK and MV) improves as α_i^{GP} decreases. When the true period durations can be characterised by a GP distribution with $\alpha_i^{GP} \approx 0$, then $D_i^{KS} \approx 0$, meaning that MMK and MV lead to a nearly exact estimation of the true primary activity statistics, as it is also the case of MMV and MVS, but within significantly shorter observation intervals as discussed earlier. The worst accuracy is obtained for the upper bound of the interval (i.e., $\alpha_i^{GP} \approx 0.25$), where the maximum estimation error is $D_i^{KS} \approx 0.06$. Based on these results, it can be concluded that the methods proposed in this work are able to provide reasonably accurate estimations of the primary activity statistics within significantly shorter observation intervals, compared to other existing estimation approaches, thus enabling DSA/CR systems to quickly adapt to new unknown operating channels.

V. Conclusion

This work has proposed and analysed novel strategies to estimate the activity statistics of an unknown channel in the context of DSA/CR systems, along with a method enabling DSA/CR devices to quantify the accuracy of the estimated statistics. Compared to other existing approaches, the proposed strategies can provide reliable and accurate estimations of the primary activity statistics within significantly shorter observation times, thus allowing DSA/CR users to quickly adapt to new unknown operating channels.

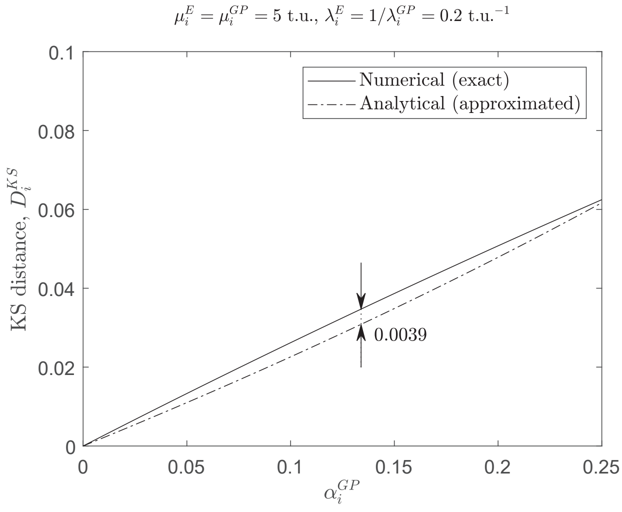

Fig. 3. KS distance as a function of the GP distribution shape parameter.

Acknowledgements

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

References

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

[2] M. López-Benítez, "Cognitive radio," in *Heterogeneous cellular networks*. Cambridge University Press, 2013, ch. 13, pp. 383–425.

[3] 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 Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 150–182, First Quarter 2018.

[4] D. G. C., K. Navaie, and Q. Ni, "Radio resource allocation in collaborative cognitive radio networks based on primary sensing profile," *IEEE Access*, vol. 6, pp. 50 344–50 357, Sep. 2018.

[5] W. Zhang, C.-X. Wang, X. Ge, and Y. Chen, "Enhanced 5G cognitive radio networks based on spectrum sharing and spectrum aggregation," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6304–6316, Dec. 2018.

[6] M. López-Benítez, A. Al-Tahmeesschi, D. K. Patel, J. Lehtomäki, and K. Umebayashi, "Estimation of primary channel activity statistics in cognitive radio based on periodic spectrum sensing observations," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 983–996, Feb. 2019.

[7] 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," *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 1, pp. 59–72, Mar. 2019.

[8] S. Geirhofer, L. Tong, and B. M. Sadler, "Dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 66–72, May 2007.

[9] 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.

[10] 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. Technol.*, vol. 62, no. 5, pp. 2091–2104, Jun. 2013.

[11] H. N. Phien, "On the computation of sample central moments," *Int'l. J. Math. Education Science Tech.*, vol. 19, no. 3, pp. 403–412, May 1988.

[12] 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.