SAMPLING SCHEMES FOR SEQUENTIAL DETECTION IN COLORED NOISE

Ruixin Niu and Pramod K. Varshney

Syracuse University
Department of EECS, 121 Link Hall
Syracuse, NY 13244

ABSTRACT

In this paper, four sampling schemes for sequential detection in colored noise are introduced. Two of them use uniform sampling procedures with high and low sampling rates, respectively. The other two employ groups of samples, which are separated by long inter-group gaps such that the inter-group correlations are negligible. Their performances, in terms of average termination time, are derived analytically. Under the assumption that all the schemes have the same power and sampling interval (x), their efficiencies are compared through analytical and numerical methods. Our results show that the scheme using group sampling with an optimal signal is the most efficient.

1. INTRODUCTION

In many detection applications, the sensor has a very low signal to noise ratio (SNR) and a decision can not be made with high reliability based on a single sample. To improve the detection performance, detection can be performed based on multiple observation samples. A powerful multiple-sample detection procedure is Wald's sequential probability ratio test (SPRT) [1]. It is well known that to get a required detection performance, SPRT, on an average, needs much fewer samples than fixed-sample-size (FSS) detection procedures. However, the SPRT is based on the assumption of independent and identically distributed (i.i.d.) samples. In practice, this assumption does not always hold and needs to be relaxed. In [2, 3], authors have shown that the optimum sequential detector for dependent samples is in the form of a generalized sequential probability ratio test (GSPRT). In [4], the SPRT and the sequential linear detector (SLD) are compared in an autoregressive noise. The author has shown that when correlation is positive, the SLD outperforms the SPRT.

In all the works mentioned above, authors have assumed that the data have already been discretized via uniform temporal sampling. We will tackle this problem from a system-design point of view. To reduce the negative effect of or even take advantage of the correlation between samples, we explore several sampling schemes. Two of them use uniform sampling procedures. The other two are based on the group sampling ideas. For all the schemes, the samples are manipulated such that they can still be processed by the simple framework of a SPRT. We will show that the scheme using group sampling with an optimal signal is much more efficient than other schemes.

2. SYSTEM MODEL AND SPRT

We consider a binary hypothesis problem:

$$\begin{aligned} H_1: \quad y(t) &= \theta + \omega(t) \\ H_0: \quad y(t) &= \omega(t) \end{aligned} \tag{1}$$

where $y(t)$ is the received observation, θ is the amplitude of the signal, and $\omega(t)$ is a wide-sense stationary Gaussian random noise, with zero mean and autocorrelation function

$$R_\omega(\tau) = \sigma^2 e^{-\frac{|\tau|}{\mu}} \tag{2}$$

where μ indicates the rate at which the correlation decays.

In this section, Wald's SPRT is introduced briefly. The successive observations $\{y_1,\ y_2, \cdots y_n\}$ are assumed to be i.i.d. The lower and upper thresholds of the SPRT, A and B, can be determined by the desired type I and type II error probabilities (α and β), as follows

$$\begin{aligned} a = lnA &= ln\frac{\beta}{1-\alpha} \\ b = lnB &= ln\frac{1-\beta}{\alpha} \end{aligned} \tag{3}$$

At each stage n, the summation of the log-likelihood ratios is computed:

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} \ln \frac{f(y_i|H_1)}{f(y_i|H_0)} \tag{4}$$

and then compared with the thresholds as follows:

$$\sum_{i=1}^{n} \lambda_i \begin{cases} \geq b & \text{stop and decide } H_1 \\ \leq a & \text{stop and decide } H_0 \\ \text{otherwise} & \text{continue} \end{cases} \tag{5}$$

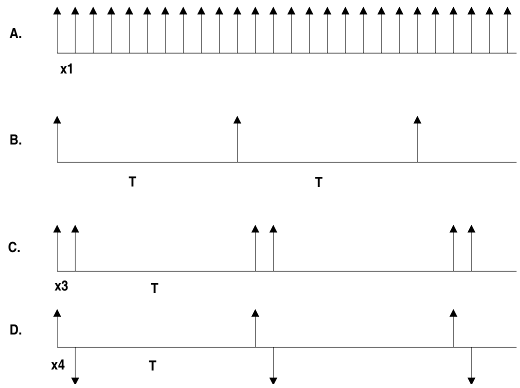

Fig. 1. Four sampling schemes in colored noise.

3. SAMPLING SCHEMES

The four sampling schemes are shown in Fig. 1. We assume that $x_1 = x_3 = x_4 = x$.

3.1. Uniform Sampling at High-Rate

The most common procedure is to uniformly sample $y(t)$, as shown in Fig. 1. After sampling, we have a digital detection problem:

$$\begin{aligned} H_1: \quad \mathbf{y} &= \theta_1 + \vec{\omega} \\ H_0: \quad \mathbf{y} &= \vec{\omega} \end{aligned} \tag{6}$$

where $\mathbf{y} = [y(x) \cdots y(nx)]$, $\vec{\omega} = [\omega(x) \cdots \omega(nx)]$, and θ_1 is the constant amplitude of the signal.

With $\omega(t)$ being stationary, it is easy to show that $\vec{\omega}$ is also stationary with autocorrelation

$$R(m) = \sigma^2 \rho^{|m|} \tag{7}$$

where $\rho = e^{-\frac{x}{\mu}}$ is the correlation coefficient between adjacent samples. The noise with this kind of autocorrelation function can be modeled as an autoregressive sequence:

$$\omega_i = \nu_i + \rho\omega_{i-1} \quad (i = 1,\ 2,\ ,\cdots) \tag{8}$$

where the sequence $\{\nu_i\}$ is i.i.d. and

$$\nu_i \ \sim \mathcal{N}\ (0,\ (1-\rho^2)\sigma^2) \tag{9}$$

In addition, $\omega_0 \ \sim \mathcal{N}\ (0,\ \sigma^2)$.

To employ the framework of the SPRT, we use a whitening filter to make the samples i.i.d., which is a differentiation process:

$$z_i \triangleq y_i - \rho y_{i-1} \tag{10}$$

Therefore, the new sequential detection problem becomes

$$\begin{aligned} H_1: \quad z_i &= (1-\rho)\theta_1 + \nu_i \\ H_0: \quad z_i &= \nu_i \end{aligned} \tag{11}$$

With (9), it is easy to show that the log-likelihood ratio for each sample is

$$\lambda_1 = \frac{z\theta_1}{(1+\rho)\sigma^2} - \frac{(1-\rho)\theta_1^2}{2(1+\rho)\sigma^2} \tag{12}$$

Therefore,

$$E[\lambda_1|H_1] = -E[\lambda_1|H_0] = \frac{(1-\rho)\theta_1^2}{2(1+\rho)\sigma^2} \tag{13}$$

For simplicity, we assume that $\alpha = \beta$. Thus, $a = -b$, and the average sample numbers (ASN) are the same under both hypotheses H_1 and H_0. According to Wald's first equation [1], the ASNs for this problem are

$$E[M_1|H_1] = E[M_1|H_0] = \frac{2(1-2\beta)b(1+\rho)\sigma^2}{(1-\rho)\theta_1^2} \tag{14}$$

To facilitate the comparison between different schemes, we define a new metric, the average termination time (ATT), which is the average time needed before either threshold (a or b) is crossed. The ATT for Scheme A is thus

$$E[t_1|H_1] = E[t_1|H_0] = \frac{2(1-2\beta)b(1+\rho)\sigma^2 x}{(1-\rho)\theta_1^2} \tag{15}$$

For a fair comparison, we assume a constant signal power P for all the sampling schemes. Therefore, the scheme with a smaller ATT consumes less energy and is more efficient. For Scheme A, the power is $P = \frac{\theta_1^2}{x}$ and (15) becomes

$$E[t_1] = \frac{2(1-2\beta)b\sigma^2}{P} \frac{1+e^{-\frac{x}{\mu}}}{1-e^{-\frac{x}{\mu}}} \tag{16}$$

Obviously, $E[t_1]$ is a monotone decreasing function of x. When $x \to \infty$ (or approximately $x \to T$), $E[t_1]$ tends to $E[t_2]$, which we will derive next. This implies that Scheme A is always less efficient than Scheme B.

3.2. Uniform Sampling at Low-Rate

In this scheme (Scheme B), as shown in Fig. 1, the sampling rate is low enough such that the correlation between adjacent samples are negligible, meaning that,

$$e^{-\frac{T}{\mu}} \leq \epsilon \tag{17}$$

where T is the sampling interval and ϵ is a very small constant. Therefore, samples can be taken as i.i.d. and a standard SPRT can be used to perform hypothesis testing

$$\begin{aligned} H_1: \quad y_i &= \theta_2 + \omega_i \\ H_0: \quad y_i &= \omega_i \end{aligned} \tag{18}$$

where θ_2 is the signal amplitude and $\{\omega_i\}$ is assumed to be an i.i.d. Gaussian sequence with zero mean and variance σ^2. It is easy to obtain the log-likelihood ratio for each sample:

$$\lambda_2 = \frac{\theta_2(2y - \theta_2)}{2\sigma^2} \tag{19}$$

Therefore,

$$E[\lambda_2|H_1] = -E[\lambda_2|H_0] = \frac{\theta_2^2}{2\sigma^2} \tag{20}$$

The ASN for this case is

$$E[M_2] = \frac{2(1-2\beta)b\sigma^2}{\theta_2^2} \tag{21}$$

The ATT is, therefore

$$E[t_2] = \frac{2(1-2\beta)b\sigma^2 T}{\theta_2^2} \tag{22}$$

Substituting the power $P = \frac{\theta_2^2}{T}$ into (22), we have

$$E[t_2] \quad = \quad \frac{2(1-2\beta)b\sigma^2}{P} \tag{23}$$

3.3. Group Sampling with Constant Amplitude

The scheme (Scheme C) is illustrated in Fig. 1. Within each individual group, uniformly spaced samples are collected. The inter-group gap T, identical to that defined in Section 3.2, is large enough so that the inter-group correlation is negligible. Note that a similar scheme based on grouping consecutive samples has been used by authors in [5].

The samples within each group are combined to form a super sample, these super samples can be taken as i.i.d. and processed by a SPRT. We denote the amplitude of each sample as θ_3, and the number of samples of each group as N. The hypothesis testing problem is

$$\begin{aligned} H_1: \quad \mathbf{y}_i &= \vec{\theta} + \vec{\omega}_i \\ H_0: \quad \mathbf{y}_i &= \vec{\omega}_i \end{aligned} \tag{24}$$

where $\mathbf{y}_i = [y_{i1}, \cdots, y_{iN}]'$, $\vec{\theta} = \theta_3[1, \cdots, 1]'$, and $\vec{\omega}_i = [\omega_{i1}, \cdots, \omega_{iN}]'$, which follows a Gaussian distribution:

$$\vec{\omega}_i \quad \sim \mathcal{N}\,(\vec{0},\ \Sigma) \tag{25}$$

where

$$\Sigma \quad = \quad \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix} \tag{26}$$

The log-likelihood ratio of each super sample ($\mathbf{y_i}$) is

$$\lambda_3 = \vec{\theta}'\Sigma^{-1}\mathbf{y} - \frac{1}{2}\vec{\theta}'\Sigma^{-1}\vec{\theta} \tag{27}$$

Therefore,

$$E[\lambda_3|H_1] = -E[\lambda_3|H_0] = \frac{1}{2}\vec{\theta}'\Sigma^{-1}\vec{\theta} \tag{28}$$

An explicit solution of Σ^{-1} is available in [6], with which and (28), we can derive that

$$E[\lambda_3|H_1] = \frac{\theta_3^2}{2\sigma^2(1+\rho)}[N - (N-2)\rho] \tag{29}$$

The ASN for this scheme is

$$E[M_3] = \frac{2(1-2\beta)b\sigma^2(1+\rho)}{\theta_3^2[N-(N-2)\rho]} \tag{30}$$

and the ATT is

$$E[t_3] = \frac{2(1-2\beta)b\sigma^2(1+\rho)}{\theta_3^2[N-(N-2)\rho]}[(N-1)x + T] \tag{31}$$

Plugging $P = \frac{N\theta_3^2}{(N-1)x+T}$ into (31), we have

$$E[t_3] = \frac{2(1-2\beta)b\sigma^2}{P}\frac{1+e^{-\frac{x}{\mu}}}{1-(1-\frac{2}{N})e^{-\frac{x}{\mu}}} \tag{32}$$

Again, $E[t_3]$ is a monotone decreasing function of x. When $x \to \infty$, $E[t_3] \to E[t_2]$. Therefore, Scheme C is always less efficient than Scheme B. It is easy to show that

$$\frac{1+e^{-\frac{x}{\mu}}}{1-e^{-\frac{x}{\mu}}} > \frac{1+e^{-\frac{x}{\mu}}}{1-(1-\frac{2}{N})e^{-\frac{x}{\mu}}} \tag{33}$$

Therefore, Scheme C is always more efficient than Scheme A. Also, $E[t_3]$ is a monotone increasing function of N and as $N \to \infty$,

$$\frac{1+e^{-\frac{x}{\mu}}}{1-(1-\frac{2}{N})e^{-\frac{x}{\mu}}} \to \frac{1+e^{-\frac{x}{\mu}}}{1-e^{-\frac{x}{\mu}}} \tag{34}$$

Therefore, when N is very large, the performance of Scheme C will converge to that of Scheme A. As we can see, Scheme C's efficiency lies somewhere between Schemes A and B.

3.4. Group Sampling with Optimal Signal Waveform

This scheme (Scheme D) is still based on the group sampling idea as in Scheme C. The difference is that the constant-amplitude signal within each group is replaced with an optimal signal. According to [7], for a FSS detection problem, under a fixed energy constraint, the optimal signal that maximizes the SNR at matched filter output is the eigenvector of the noise covariance matrix corresponding to the minimum eigenvalue. For example, if group size $N = 2$, we have

$$\Sigma \quad = \quad \sigma^2 \left[\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right] \tag{35}$$

The optimal signal, which is the eigenvector with minimum eigenvalue, is in the form of $\theta[1 \quad -1]'$, as shown in Fig. 1 for the case $N = 2$.

It is clear that the covariance matrix is the same as (26). We denote its minimum eigenvalue and corresponding eigenvector as λ_{min} and $\vec{v}_{min}$, respectively. Similar to the derivation of (27), we have the log-likelihood ratio of each super sample is

$$\lambda_4 = \vec{v}'_{min}\Sigma^{-1}\mathbf{y} - \frac{1}{2}\vec{v}'_{min}\Sigma^{-1}\vec{v}_{min} \tag{36}$$

Hence,

$$\begin{aligned} E[\lambda_4|H_1] = -E[\lambda_4|H_0] \quad &= \quad \frac{1}{2}\vec{v}'_{min}\Sigma^{-1}\vec{v}_{min} \\ &= \quad \frac{\vec{v}'_{min}\vec{v}_{min}}{2\lambda_{min}} \end{aligned} \tag{37}$$

The ASN for Scheme D is

$$E[M_4] = \frac{2(1-2\beta)b\lambda_{min}}{\vec{v}'_{min}\vec{v}_{min}} \tag{38}$$

and the ATT is

$$E[t_4] = \frac{2(1-2\beta)b\lambda_{min}}{\vec{v}'_{min}\vec{v}_{min}}[(N-1)x + T] \tag{39}$$

Plugging $P = \frac{\vec{v}'_{min}\vec{v}_{min}}{(N-1)x+T}$ into (39), we have

$$E[t_4] \quad = \quad \frac{2(1-2\beta)b\sigma^2}{P}\frac{\lambda_{min}}{\sigma^2} \tag{40}$$

From (40), it is clear that the ATT is proportional to λ_{min}. λ_{min} can be calculated numerically. However, we do not show the results here due to the limited space. Basically, λ_{min} is a monotone increasing function of x, meaning that the smaller the sampling interval, the better the performance is. This is because that smaller sampling interval gives rise to higher correlation between samples, with which the scheme can do a better noise cancelation. In addition, λ_{min} is a monotone decreasing function of N, meaning that the larger the group size, the better the performance.

4. NUMERICAL RESULTS AND DISCUSSION

Without loss of generality, we assume $\frac{2(1-2\beta)b\sigma^2}{P} = 1$ and $\mu = 1$. We take $\epsilon = 10^{-3}$ and hence $T = 6.91\mu$. The ATTs for different schemes, are plotted in Fig. 2. It is evident that

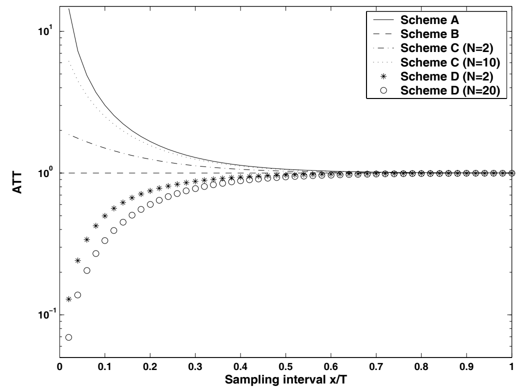

Fig. 2. Average termination times for different schemes.

Scheme A has the worst efficiency (largest ATT). Scheme D has the best performance, especially when sampling rate is high and group size N is large. It is also noteworthy that when $x \to T$, all the schemes converge to Scheme B.

With the identical signal power, Schemes B, C and D are more efficient than Scheme A, especially when the correlation between samples is strong, and Scheme D has the highest efficiency among all the schemes.

5. REFERENCES

[1] A. Wald, *Sequential Analysis*, John Wiley & Sons, Inc., 1947.

[2] J. Cochlar and I. Vrana, "On the Optimum Sequential Test of Two Hypotheses for Statistically Dependent Observations," *Kybernetika*, vol. 14, no. 1, pp. 57–69, 1978.

[3] B. Eisenberg, B.K. Ghosh, and G. Simons, "Properties of Generalized Sequential Probability Ratio Tests," *The Annals of Statistics*, vol. 4, no. 2, pp. 237–251, March 1976.

[4] S. Tantaratana, "Comparison of the SPRT and the Sequential Linear Detector in Autoregressive Noise," *IEEE Transactions on Information Theory*, vol. 31, no. 5, pp. 693–697, September 1985.

[5] S.A. Kassam and J.B. Thomas, "A Class of Nonparametric Detectors for Dependent Input Data," *IEEE Transactions on Information Theory*, vol. 21, no. 4, pp. 431–437, July 1975.

[6] J.K. Lindsey, *Models for Repeated Measurements*, Oxford University Press, 1999.

[7] S.M. Kay, *Fundamentals of Statistical Signal Processing II: Detection Theory*, Prentice Hall, Englewood Cliffs, NJ 07632, 1998.