Int. Journal of Math. Analysis, Vol. 6, 2012, no. 24, 1153 - 1162

False-Alarm and Non-Detection Probabilities for On-line Quality Control via HMM

C.C.Y. Dorea[a,1], C.R. Gonçalves[a], P.G. Medeiros[b] and W.B. Santos[a]

[a]Universidade de Brasilia
70910-900 Brasilia-DF, Brazil

[b]Universidade Federal do Rio Grande do Norte
59072-970 Natal-RN, Brazil

Abstract

On-line quality control during production calls for monitoring produced items according to some prescribed strategy. It is reasonable to assume the existence of system internal non-observable variables so that the carried out monitoring is only partially reliable. In this note, under the setting of a Hidden Markov Model (HMM) and assuming that the evolution of the internal state changes are governed by a two-state Markov chain, we derive estimates for false-alarm and non-detection malfunctioning probabilities. Kernel density methods are used to approximate the stable regime density and the stationary probabilities. As a side result, alternative monitoring strategies are proposed.

Mathematics Subject Classification: 60J20, 60J10, 90B30

Keywords: Hidden Markov model; False-alarm; Kernel density methods

1. Introduction

Limited by the cost function a sampling interval m is selected and classical on-line quality control during production adopts the procedure of monitoring the sequence of items, independently produced, by examining a single item at every m produced items. Based on the quality requirements and on the distribution of the examined variable a control region C is pre-specified. If the examined item satisfies the control limits, the process is said to be in

[1]Corresponding author : changdorea@unb.br.
Research partially supported by CNPq, CAPES, FAPDF and FINATEC.

control and the production continues; otherwise, the process is declared out of control and the production is stopped for adjustments. After adjustments the production is resumed and it is in control again. At each stoppage a new production cycle is defined (see, for example, [4]). Suppose now the internal working status of the system is non-observable and may change from a good working condition (on control) to a deteriorating status (out of control). In [2] it is proposed a model where these changes are governed by a two-state Markov chain $\{\theta_n\}_{n\geq 0}$. When $\theta_n = 0$ the process is said to be in control at time n; and, if $\theta_n = 1$ the process is out of control. In their proposal, an observable random variable X_n, related to characteristics of interest, is examined at every m produced items. It is assumed that X_n has a Gaussian distribution, $\mathcal{N}(\mu, \sigma^2)$, with $\mu = \mu_{\theta_n}$. And values of m as well as the parameter d of the control region $C = (\mu_0 - d\sigma, \mu_0 + d\sigma)$ were determined by considering a given cost function to be minimized.

Here, in the framework of a Hidden Markov Model (HMM), we compute and propose estimation techniques for the false-alarm and non-detection probabilities. False-alarm occurs if the observed variable falls outside the control region but the non-observable internal system state is 0, a good working state. When the opposite occurs we have a non-detection situation. It is assumed that all the working (good) states are lumped together as state 0 and the deteriorating states are gathered as state 1. A Markov chain $\{\theta_n\}_{n\geq 0}$ describes the evolution of the state of the production system. Associated with this chain we observe a sequence of conditionally independent random variables $\{X_n\}_{n\geq 1}$, with the distribution of each X_n depending on the corresponding state θ_n. This process $\{\theta_n, X_n\}$ is generally referred to as a HMM. More specifically, we have

$$P(X_{n+1} \in A | X_1, \ldots, X_n, \theta_0, \ldots, \theta_n) = P(X_{n+1} \in A | \theta_n) \tag{1}$$

and

$$P(X_1 \in A_1, \ldots, X_n \in A_n | \theta_1, \ldots, \theta_n) = \prod_{j=1}^{n} P(X_j \in A_j | \theta_j). \tag{2}$$

In section 2, we treat the special case where the transition matrix of $\{\theta_n\}$ as well as the conditional densities of X_n given θ_n are known, but not necessarily Gaussian densities. For any initial distribution of θ_0, Proposition 1 and Proposition 2 give the false-alarm and the non-detection probabilities. As a side result, upper bounds for the sampling interval m are proposed.

In section 3, we assume that the transition probabilities of the hidden chain $\{\theta_n\}$ are unknown. Based on the observable sample $X_1, \ldots, X_n$ kernel density methods are used to approximate the stable regime density. Theorem

1 shows that the limiting stable density is a mixture of densities, which in the particular case treated in [2] reduces to NORMIX (mixture of normal densities). Corollary 1 provide an estimate for the stationary distribution of $\{\theta_n\}$ and leads to an alternative monitoring strategy that takes into account the false-alarm situations. And the estimates for false-alarm and non-detection probabilities can be found in Corollary 2.

2. False-alarm and Non-detection

For the HMM process $\{\theta_n, X_n\}$ assume that the chain $\{\theta_n\}$ has known transition probabilities given by

$$P = \begin{pmatrix} 1-p & p \\ \epsilon & 1-\epsilon \end{pmatrix}$$

$$0 < p < 1\ ,\ \epsilon > 0 \text{ and } p + \epsilon < 1. \tag{3}$$

The conditional distribution of X_n given θ_n are also known and based on this distribution a control region C is pre-selected. It is assumed that

$$P(X_n \in A|\theta_n = i) = \int_A f(x|i)dx,$$

$$0 < q_0 = \int_{C^c} f(x|0)dx < 1 \text{ and } 0 < q_1 = \int_{C^c} f(x|1)dx < 1. \tag{4}$$

If the sampling interval $m = 1$, the on-line quality monitoring adopts the following strategy : items $X_1, X_2, \ldots$ are inspected and verified whether $X_1 \in C, X_2 \in C, \ldots$; maintenance is required at time n if $X_1 \in C, \ldots, X_{n-1} \in C$ and $X_n \notin C$. Thus we can define the alert times by

$$\tau_X = \inf\{k : k \geq 1, X_k \notin C\}.$$

False-alarm occurs at time k if $\tau_X = k$ but the non-observable internal system state is 0, a good working state. Let τ_θ be defined as the first time, after time 0, the system reaches state 1,

$$\tau_\theta = \inf\{k : k \geq 1, \theta_k = 1\}. \tag{5}$$

Then false-alarm and non-detection correspond, respectively, to the events $(\tau_X < \tau_\theta)$ and $(\tau_X > \tau_\theta)$. Let P_ν denote the probability when the initial distribution of θ_0 is ν. Similarly, let P_0 and P_1 denote the probability when $\theta_0 = 0$ and $\theta_0 = 1$ respectively.

Proposition 1. Let ν be the initial distribution of θ_0. Then the false-alarm probability is given by

$$P_\nu(\tau_X < \tau_\theta) = \nu(0)\frac{(1-p)q_0}{1-(1-p)(1-q_0)} + \nu(1)\frac{\epsilon q_0}{1-(1-p)(1-q_0)} \tag{6}$$

and the non-detection probability is given by

$$P_\nu(\tau_X > \tau_\theta) = \nu(0)\frac{p(1-q_1)}{1-(1-p)(1-q_0)} + \nu(1)\frac{[q_0(1-\epsilon)+(1-q_0)p](1-q_1)}{1-(1-p)(1-q_0)}. \tag{7}$$

Proof. (i) For $\theta_0 = 0$ we have from (3) and (4)

$$P_0(\tau_X = 1, \tau_\theta > 1) = P_0(\theta_1 = 0, X_1 \notin C) = (1-p)q_0$$

and from (1) and (2) we have for $k \geq 2$

$$\begin{aligned} P_0(\tau_X = k, \tau_\theta > k) &= P_0(X_1 \in C, \cdots, X_{k-1} \in C, X_k \notin C, \theta_1 = \cdots = \theta_k = 0) \\ &= P_0(\theta_1 = \cdots = \theta_k = 0)[\prod_{j=1}^{k-1} P(X_j \in C|\theta_j = 0)]P(X_k \notin C|\theta_k = 0) \\ &= (1-p)[(1-p)(1-q_0)]^{k-1}q_0. \end{aligned}$$

It follows that

$$P_0(\tau_X < \tau_\theta) = \frac{(1-p)q_0}{1-(1-p)(1-q_0)}.$$

Similarly, if $\theta_0 = 1$ we have

$$P_1(\tau_X = 1, \tau_\theta > 1) = P(\theta_1 = 0|\theta_0 = 1)P(X_1 \notin C|\theta_1 = 0) = \epsilon q_0$$

and for $k \geq 2$

$$P_1(\tau_X = k, \tau_\theta > k) = \epsilon[(1-p)(1-q_0)]^{k-1}q_0.$$

And (6) follows.

(ii) To prove (7) note that

$$P_0(\tau_X = \tau_\theta = 1) = P(\theta_1 = 1|\theta_0 = 0)P(X_1 \notin C|\theta_1 = 1) = pq_1$$

and for $k \geq 2$

$$P_0(\tau_X = \tau_\theta = k) = [(1-p)(1-q_0)]^{k-1}pq_1.$$

Moreover,

$$P_1(\tau_X = \tau_\theta = 1) = (1-\epsilon)q_1$$

and for $k \geq 2$

$$P_1(\tau_X = \tau_\theta = k) = \epsilon(1-q_0)[(1-p)(1-q_0)]^{k-2}pq_1.$$

It follows that

$$P_\nu(\tau_X = \tau_\theta) = \nu(0)\frac{pq_1}{1-(1-p)(1-q_0)} + \nu(1)\frac{[q_0(1-\epsilon)+(1-q_0)p]q_1}{1-(1-p)(1-q_0)}.$$

This along with (6) give us (7). □

Now, assume that a single item is inspected at every m items produced. Then for $m \geq 2$ we can define

$$\tau_X^{(m)} = \inf\{km : k \geq 1, X_{km} \notin C\}. \quad (8)$$

Using the same type of arguments as above we get

$$\begin{aligned} P_0(\tau_X^{(m)} = km, \tau_\theta > km) &= (1-p)^{km}(1-q_0)^{k-1}q_0 \\ &= (1-p)^m[(1-p)^m(1-q_0)]^{k-1}q_0 \end{aligned}$$

and

$$P_1(\tau_X^{(m)} = km, \tau_\theta > km) = \epsilon(1-p)^{m-1}[(1-p)^m(1-q_0)]^{k-1}q_0.$$

Moreover,

$$P_0(\tau_X^{(m)} = \tau_\theta = km) = [(1-p)^m(1-q_0)]^{k-1}(1-p)^{m-1}pq_1$$

and

$$P_1(\tau_X^{(m)} = \tau_\theta = km) = [(1-p)^m(1-q_0)]^{k-1}\epsilon(1-p)^{m-2}pq_1.$$

Proposition 2. Let ν be the initial distribution of θ_0. Assume that the monitoring strategy has sampling interval m. Then the false-alarm probability is given by

$$P_\nu(\tau_X^{(m)} < \tau_\theta) = \nu(0)\frac{(1-p)^m q_0}{1-(1-p)^m(1-q_0)} + \nu(1)\frac{\epsilon(1-p)^{m-1}q_0}{1-(1-p)^m(1-q_0)} \quad (9)$$

and for $m \geq 2$ the non-detection probability is given by

$$\begin{aligned} P_\nu(\tau_X^{(m)} > \tau_\theta) &= \nu(0)\frac{1-(1-p)^{m-1}+(1-p)^{m-1}p(1-q_1)}{1-(1-p)^m(1-q_0)} \\ &+ \nu(1)\frac{1-(1-p)^m(1-q_0)-\epsilon(1-p)^{m-2}[(1-p)q_0+pq_1]}{1-(1-p)^m(1-q_0)}. \end{aligned} \tag{10}$$

Next, we propose upper bounds for the sampling interval m. We may assume that the process starts at good working conditions ($\theta_0 = 0, X_0 \in C$). Our aim is to detect whether the chain $\{\theta_n\}$ has changed to state 1. A trivial upper bound is given by the mean changing time from state 0 to state 1. From (3) and (5) we have

$$m \leq \sum_{k\geq 1} kP(\tau_\theta = k|\theta_0 = 0) = \sum_{k\geq 1} k(1-p)^{k-1}p = \frac{1}{p}.$$

Since $\{\theta_n\}$ is non-observable, the designed monitoring strategy relies on the observable process $\{X_n\}$. Taking this into account one should choose m smaller so that in average we will detect the first time the process $\{\theta_n, X_n\}$ reaches the alert zone. Alert occurs at time k if either $X_k \in C^c$ or $\theta_k = 1$. And this can be expressed by

$$\tau_A = \inf\{k : k \geq 1, (\theta_k, X_k) \in A\}.$$

where $A = (\{0\} \times C^c) \cup (\{1\} \times C) \cup (\{1\} \times C^c)$. Since

$$P((\theta_{k+1}, X_{k+1}) \in A^c|(\theta_k, X_k) \in A^c) = (1-p)(1-q_0),$$

the expected first hitting time in alert zone is given by

$$m \leq E(\tau_A|\theta_0 = 0, X_0 \in C) = \frac{1}{1-(1-p)(1-q_0)}.$$

3. Estimation Results

In this section we assume that the hidden Markov chain $\{\theta_n\}$ has transition matrix P of the form (3), but unknown. As for the process $\{X_n\}$, we assume that the conditional densities are known and satisfy condition (4).

Note that since all entries of P are strictly positive, $\{\theta_n\}$ is an ergodic chain and the stationary (limiting) distribution exists, $\pi P = \pi$,

$$\pi(0) = \frac{\epsilon}{p+\epsilon} \text{ and } \pi(1) = \frac{p}{p+\epsilon}. \tag{11}$$

Define the mixture density function

$$f(x) = \pi(0)f(x|0) + \pi(1)f(x|1). \tag{12}$$

Then, in some sense, $f(\cdot)$ represents the density of $\{X_n\}$ when the process reaches some "stable" regime. Mixture models have been well recognized as useful in many practical applications. It is a popular model-based approach to dealing with data in the presence of population heterogeneity, in the sense that, data consist of unlabelled observations, each of which is thought to belong to one of the distinct classes. For a comprehensive list of applications and literature survey in this area see, for example, [3]. Though we are assuming known conditional densities, the stable regime density (12) may indicate which type of distribution one should assume for the variables X_n as well as which control region C one should select.

Results from [1] show that, based on the observable sample $X_1, \ldots, X_n$, kernel density methods can be successfully used to estimate the stable regime density. For a probability density K on R define

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{k=1}^{n} K(\frac{X_k - x}{h})\,,\ h = h_n \downarrow 0\,,\ nh_n \to \infty \text{ as } n \to \infty. \tag{13}$$

Typically one takes $K(\cdot)$ either a Gaussian density or an uniform density centered at x. Using these results we have Theorem 1 which allows us to make inference on P. Its proof will be postponed to the end of this section.

Theorem 1. Assume that the process $\{\theta_n, X_n\}$ satisfies (3) and (4) and let ν be any initial distribution of θ_0. Then given $\delta > 0$ there exist $c_1 = c_1(\delta) > 0$ and $c_2 = c_2(\delta) > 0$ such that

$$P_\nu(\int |\hat{f}_n(y) - f(y)|dy \geq \delta) \leq c_1 \exp\{-c_2 n\} \tag{14}$$

and almost surely (a.s.)

$$\lim_{n\to\infty} \int |\hat{f}_n(y) - f(y)|dy = 0 \ \text{ a.s.} \tag{15}$$

Next, we estimate the stationary probabilities (11). Choose x_* so that $f(x_*|0) \neq f(x_*|1)$ and define

$$\hat{\pi}_n(0) = \left|\frac{\hat{f}_n(x_*) - f(x_*|1)}{f(x_*|0) - f(x_*|1)}\right| \ \text{ and } \ \hat{\pi}_n(1) = 1 - \hat{\pi}_n(0). \tag{16}$$

From (15) it follows that,

Corollary 1. For any initial distribution of θ_0 we have with probability 1,

$$\hat{\pi}_n(0) \to \pi(0) \text{ and } \hat{\pi}_n(1) \to \pi(1) \text{ as } n \to \infty. \tag{17}$$

This result suggests an alternative on-line control procedure that will take into account the false-alarm situations. Let

$$\hat{f}_n^{(m)}(x_*) = \frac{1}{nh} \sum_{k=1}^{n} K(\frac{X_{km} - x_*}{h}).$$

Then, from (15) we have

$$\lim_{n\to\infty} |\hat{f}_n^{(m)}(x_*) - f(x_*)| \text{ a.s.}$$

Thus, with $\hat{f}_n^{(m)}(x_*)$ in place of $\hat{f}_n(x_*)$ in (16) we get (17). It follows that the posterior distribution $P(\theta_{nm} = 0|X_m, \ldots, X_{nm})$ can be approximated by

$$\hat{P}(\theta_{nm} = 0|X_m, \ldots, X_{nm}) = \left| \frac{\hat{f}_n^{(m)}(x_*) - f(x_*|1)}{f(x_*|0) - f(x_*|1)} \right| \frac{f(x_*|0)}{\hat{f}_n^{(m)}(x_*)}.$$

Monitoring Strategy. Assume that the conditional densities $f(\cdot|0)$ and $f(\cdot|1)$ are both known. Suppose that the control region C and the sampling interval m have been pre-selected. Let $\tau_X^{(m)}$ be defined by (8) and let $0 < \eta < 1$ be a given significance level. Then intervention is carried out at time nm if $\tau_X^{(m)} = nm$ and

$$\hat{P}(\theta_{nm} = 0|X_m, \ldots, X_{nm}) < \eta.$$

From (16) we can also estimate false-alarm and non-detection probabilities. Observe that, in long-run, the false-alarm can be computed as

$$P(\theta_{km} = 0|X_{km} \notin C) = \frac{P(X_{km} \notin C|\theta_{km} = 0)P(\theta_{km} = 0)}{P(X_{km} \notin C)}.$$

But $P(X_{km} \notin C|\theta_{km} = 0) = q_0$ and by (11) we have $\lim_{n\to\infty} P(\theta_{km} = 0) = \pi(0)$. From Theorem 1 with $f(\cdot)$ given by (12) we have

$$\lim_{n\to\infty} P(X_{km} \notin C) = \int_{C^c} f(x)dx.$$

Similarly, we have for non-detection

$$P(\theta_{km} = 1|X_{km} \in C) = \frac{P(X_{km} \in C|\theta_{km} = 1)P(\theta_{km} = 1)}{P(X_{km} \in C)}.$$

These results can be summarized by:

Corollary 2. For any initial distribution of θ_0 we have with probability 1,

$$\frac{q_0\hat{\pi}_n(0)}{\int_{C^c}\hat{f}_n(x)dx} \to P(\text{false-alarm}) \text{ and } \frac{(1-q_1)\hat{\pi}_n(1)}{\int_C \hat{f}_n(x)dx} \to P(\text{non-detection}). \quad (18)$$

Proof of Theorem 1. Enough to verify that the hypotheses of Theorem 2 and Corollary 1 from [1] are satisfied. First, we shown that $\{\theta_n\}$ is is uniformly ergodic, that is, there exists $0 < \rho < 1$ such that

$$|P_{ij}^n - \pi(j)| \leq \rho^n \ i,j = 0,1. \quad (19)$$

From (3) we have,

$$P^n = \begin{pmatrix} \dfrac{\epsilon}{p+\epsilon} + p\dfrac{(1-p-\epsilon)^n}{p+\epsilon} & \dfrac{p}{p+\epsilon} - p\dfrac{(1-p-\epsilon)^n}{p+\epsilon} \\ \dfrac{\epsilon}{p+\epsilon} - \epsilon\dfrac{(1-p-\epsilon)^n}{p+\epsilon} & \dfrac{p}{p+\epsilon} + \epsilon\dfrac{(1-p-\epsilon)^n}{p+\epsilon} \end{pmatrix}$$

and from (11) we get (19) by taking $\rho = 1-p-\epsilon$.

Next, we show that $\{\theta_n\}$ satisfies the ϕ-mixing condition : $\forall A \in \mathcal{F}_k$ and $\forall B \in \mathcal{F}_{k+n}^\infty$

$$|P_\pi(A \cap B) - P_\pi(A)P_\pi(B)| \leq \phi(n)P_\pi(A) \quad (20)$$

with $\sum_{n\geq 1}\phi(n) < \infty$ and σ-algebras defined by

$$\mathcal{F}_n = \sigma(\theta_0,\ldots,\theta_n) \text{ and } \mathcal{F}_n^\infty = \sigma(\theta_n,\theta_{n+1},\ldots).$$

Note that for $A = (\theta_0 = a_0,\ldots,\theta_k = a_k)$, $B = (\theta_{k+n} = b_{k+n},\ldots)$ and $P_\pi(A) > 0$ we have

$$P_\pi(B|A) = P_\pi(B|\theta_k = a_k) = P_{a_k b_{k+n}}^n P_\pi(\theta_{k+n+1} = b_{k+n+1},\ldots|\theta_{k+n} = b_{k+n})$$

and

$$P_\pi(B) = \pi(b_{k+n})P_\pi(\theta_{k+n+1} = b_{k+n+1},\ldots|\theta_{k+n} = b_{k+n}).$$

It follows that

$$|P_\pi(B|A) - P_\pi(B)| \leq |P_{a_k b_{k+n}}^n - \pi(b_{k+n})| \leq \rho^n = \phi(n)$$

and (20) is satisfied. $\square$

References :

[1] Dorea, C.C.Y. and Zhao, L.C. (2002) - *Nonparametric density estimation in Hidden Markov Models*, Statis. Inference for Stoch. Processes, vol. 5, 55-64.

[2] Ho, L.L.; Medeiros, P.G. and Borges, W. (2007) - *An alternative model for on-line quality monitoring for variables*, Intern. Jour. of Production Economics, vol. 107, 202-222.

[3] McLachlan, G.L. and Peel, D. (2000) - *Finite Mixture Models*,John Wiley, New York.

[4] Taguchi, G.; Elsayed, E.D. and Hsiang, T. (1989) - *Quality Engineering in Production Systems*, McGraw-Hill, New York.

Received: November, 2011