6

Quantized Identification and Asymptotic Efficiency

Up to this point, we have been treating binary-valued observations. The fundamental principles and basic algorithms for binary-valued observations can be modified to handle quantized observations as well. One way to understand the connection is to view a quantized observation as a vector-valued binary observation in which each vector component represents the output of one threshold, which is a binary-valued sensor. The dimension of the vector is the number of the thresholds in the quantized sensor.

However, since a binary-valued sensor is already sufficient for achieving strong convergence and mean-square convergence for the parameter estimates, and weak convergence of the centered and scaled estimation errors, it is natural to ask: Why do we need quantized observations? What benefits can be gained from using more complicated sensors? To answer these questions, we study the efficiency issue that is characterized by the Cramér–Rao (CR) lower bounds. We are seeking algorithms that utilize all statistical information in the data about the unknown parameters, in the sense that they achieve asymptotically the best convergence rates given by the CR lower bound. Consequently, the CR lower bounds become a characterization of the system complexity in this problem. Comparisons of such complexities among sensors of different thresholds permit us to answer the above questions rigorously and completely.

Section 6.1 begins with basic identification algorithms and their convergence properties. To utilize information from all thresholds collaboratively, Section 6.2 introduces an algorithm named by the authors as the quasi-convex combination estimator (QCCE). Expressions for identification errors are derived. Section 6.3 develops some important expressions of

L.Y. Wang et al., *System Identification with Quantized Observations*, Systems & Control: Foundations & Applications, DOI 10.1007/978-0-8176-4956-2_6,

identification error covariances that are essential for deriving the efficiency of the optimal QCCE algorithms. The main results of this chapter are contained in Section 6.4, in which the asymptotic efficiency of the optimal QCCE is established.

6.1 Basic Algorithms and Convergence

Consider a single-input–single-output, linear, time-invariant, stable, discrete-time system G given by

$$y_k = Gu_k + d_k,\ \ k = 1, 2, \dots, \tag{6.1}$$

where u_k is the input, d_k is the disturbance, and G is either a rational transfer function or an FIR (finite impulse response) system, or in the simplest case, a gain. The output y_k is measured by a sensor of m_0 thresholds $-\infty < C_1 < \cdots < C_{m_0} < \infty$. The sensor is represented by a set of m_0 indicator functions $s_k = [s_k^{\{1\}}, \dots, s_k^{\{m_0\}}]'$, where $s_k^{\{i\}} = I_{\{-\infty < y_k \le C_i\}}$, $i = 1, \dots, m_0$.

First, consider the simplest case of identifying a constant θ:

$$y_k = \theta + d_k.$$

Under Assumption (A3.1), $\{d_k\}$ is a sequence of i.i.d. random variables with distribution function $F(\cdot)$. Thus, for each threshold C_i, $\{s_k^{\{i\}}\}$ is also an i.i.d. sequence. Then

$$p_i = E(s_k^{\{i\}}) = F(C_i - \theta) := F_i(\theta).$$

Since $F_i(\theta)$ is invertible, we denote its inverse by $G_i(\cdot)$, and hence $G_i(p_i) = \theta$. Define

$$\xi_N^{\{i\}} = \frac{1}{N}\sum_{k=0}^{N-1} s_k^{\{i\}}; \quad \theta_N^{\{i\}} = G_i(\xi_N^{\{i\}}).$$

By virtue of the results in Chapter 3, $\theta_N^{\{i\}}$ is asymptotically unbiased. Let $\theta_N^{\{i\}}$, $i = 1, \dots, m_0$, be m_0 asymptotically unbiased estimators of θ based on samples of size N. Denote

$$\begin{aligned}
\Theta_N &= [\theta_N^{\{1\}}, \dots, \theta_N^{\{m_0\}}]', \\
e_N^{\{i\}} &= \theta_N^{\{i\}} - \theta, \\
e_N &= [e_N^{\{1\}}, \dots, e_N^{\{m_0\}}]', \\
\mathbb{1} &= [1, 1, \dots, 1]' \in \mathbb{R}^{m_0}.
\end{aligned}$$

Then $e_N = \Theta_N - \theta\mathbb{1}$. Define

$$V_N(\theta) = Ee_N e_N'. \tag{6.2}$$

Note that $Ee_N \to 0$ as $N \to \infty$, and that $V_N(\theta)$ is a covariance matrix of e_N that is positive semidefinite. Although the calculation of $V_N(\theta)$ may be cumbersome, the following asymptotic result shows that $V_N(\theta)$ can be approximated by a computable function.

From $p_i = F_i(\theta)$, define

$$h_i(\theta) = \partial F_i(\theta)/\partial\theta.$$

Then

$$\partial G_i(p_i)/\partial p_i = 1/h_i(\theta).$$

Denote

$$\begin{aligned} p &= [p_1, \dots, p_{m_0}]', \\ h(\theta) &= [h_1(\theta), \dots, h_{m_0}(\theta)]', \\ G(p) &= [G_1(p_1), \dots, G_{m_0}(p_{m_0})]', \end{aligned} \tag{6.3}$$

and

$$\begin{aligned} M &= \begin{bmatrix} p_1 & p_1 & \cdots & p_1 \\ p_1 & p_2 & \cdots & p_2 \\ \vdots & & & \vdots \\ p_1 & p_2 & \cdots & p_{m_0} \end{bmatrix}, \\ U &= \mathrm{diag}\left(\frac{1}{h_1(\theta)}, \dots, \frac{1}{h_{m_0}(\theta)}\right). \end{aligned} \tag{6.4}$$

Theorem 6.1. *As $N \to \infty$,*

$$NV_N(\theta) \to U(M - pp')U := \Psi(\theta). \tag{6.5}$$

Proof. As mentioned in Remark 3.11, the weak convergence of $B_N(\cdot)$ and the Skorohod representation (without changing notations) enable us to assume that $B_N(\cdot) \to B^0(\cdot)$ w.p.1. Let $\xi_N = [\xi_N^{\{1\}}, \dots, \xi_N^{\{m_0\}}]'$. Consider

$$e_N = \Theta_N - \theta\mathbb{1} = G(\xi_N) - G(p).$$

By Theorem 3.12,

$$\Upsilon_N^{\{i\}} = \sqrt{N} e_N^{\{i\}} = \sqrt{n}(G_i(\xi_N^{\{i\}}) - G_i(p_i))$$

converges to

$$\Upsilon_N^{\{i\}} \to \Upsilon^{\{i\}} = \frac{\partial G_i(p_i)}{\partial p_i} B^0(v^{\{i\}}) = \frac{B^0(v^{\{i\}})}{h_i(\theta)} \quad \text{w.p.1 as} \quad N \to \infty. \tag{6.6}$$

As a result, $\Upsilon = U(B^0)'$ with

$$\Upsilon = [\Upsilon^{\{1\}}, \dots, \Upsilon^{\{m_0\}}]',$$
$$B^0 = [B^0(v^{\{1\}}), \dots, B^0(v^{\{m_0\}}]',$$
$$U = [\frac{\partial G_1(p_1)}{\partial p_1}, \dots, \frac{\partial G_{m_0}(p_{m_0})}{\partial p_{m_0}}]'.$$

It follows from Theorem 3.12 that

$$E\Upsilon_N \Upsilon_N' \to EUB^0(B^0)'U$$

and

$$\begin{aligned} EB^0(B^0)' &= \begin{bmatrix} p_1 - p_1^2 & p_1 - p_1 p_2 & \dots & p_1 - p_1 p_{m_0} \\ p_1 - p_1 p_2 & p_2 - p_2^2 & \dots & p_2 - p_2 p_{m_0} \\ \vdots & & & \vdots \\ p_1 - p_{m_0} p_1 & p_2 - p_{m_0} p_2 & \dots & p_{m_0} - p_{m_0}^2 \end{bmatrix} \\ &= M - pp^T. \end{aligned}$$

Therefore,

$$E\Upsilon_N \Upsilon_N' \to U(M - pp')U.$$

The proof of the theorem is thus completed. □

The covariance in Theorem 6.1 reflects identification errors from each threshold and their correlations. However, it does not consider the unique feature here that all estimates are for the same parameter θ. Combining these estimates will eventually lead to an efficient algorithm.

6.2 Quasi-Convex Combination Estimators (QCCE)

Define $\beta = [\beta_1, \dots, \beta_{m_0}]'$ such that $\beta_1 + \dots + \beta_{m_0} = 1$. One can construct an estimator $\widehat{\theta}_N$ of θ by

$$\widehat{\theta}_N = \sum_{i=1}^{m_0} \beta_i \theta_N^{\{i\}} = \beta' \Theta_N. \tag{6.7}$$

$\widehat{\theta}_N$ is called a *quasi-convex combination estimator* (*QCCE*). The term "quasi-convex" is used since β_i need not be nonnegative. Since $\theta_N^{\{i\}}$ is asymptotically unbiased,

$$E\widehat{\theta}_N = \beta' E\Theta_N \to \beta' \theta \mathbb{1} = \theta \quad \text{as} \quad N \to \infty.$$

Hence, $\widehat{\theta}_N$ is an asymptotically unbiased estimate of θ. Moreover, the variance of the estimation error $\widehat{\theta}_N - \theta$ is given by

$$\begin{aligned}\overline{\sigma}_N^2 : &= E(\beta'\Theta_N - \theta)^2 = E(\beta'\Theta_N - \beta'\theta 1\!\!1)^2 \\ &= \beta' E e_N e_N' \beta = \beta' V_N(\theta)\beta.\end{aligned}$$

That is, the variance is in a quadratic form with respect to the vector β.

The estimator that minimizes $\overline{\sigma}_N^2$ is called *the optimal quasi-convex combination estimator* (*optimal QCCE*), which is obtained from

$$\sigma_N^2 = \min_{\beta, \beta' 1\!\!1 = 1} \overline{\sigma}_N^2 = \min_{\beta, \beta' 1\!\!1 = 1} \beta' V_N(\theta)\beta. \tag{6.8}$$

Theorem 6.2. *Under Assumption* (A3.1) *and assuming $V_N(\theta)$ is positive definite, the optimal QCCE can be obtained by choosing*

$$\beta^* = \frac{V_N^{-1}(\theta) 1\!\!1}{1\!\!1' V_N^{-1}(\theta) 1\!\!1}, \quad \widehat{\theta}_N = (\beta^*)'\Theta_N, \tag{6.9}$$

and the minimal variance is

$$\sigma_N^2 = \frac{1}{1\!\!1' V_N^{-1}(\theta) 1\!\!1}. \tag{6.10}$$

Proof. The estimator that solves (6.8) is in fact the Gauss–Markov estimator [64, p. 84] (linear minimum variance unbiased estimator) and (6.9) follows directly. For an elementary derivation, one defines the Hamiltonian

$$H(\beta, \lambda) = \beta' V_N(\theta)\beta + \lambda(1 - \beta' 1\!\!1),$$

where λ is a Lagrange multiplier. Using standard techniques in optimization (see [64, Chapter 10]) yields the stationary point (λ^*, β^*) of $H(\beta, \lambda)$ with $\lambda^* = 2/(1\!\!1' V_N^{-1}(\theta) 1\!\!1)$ and β^* given in (6.9). It can be verified that the stationary point is indeed a minimum. Substituting the above solutions into $\overline{\sigma}_N^2$, we obtain the optimal variance as in (6.10). □

Remark 6.3. The optimal QCCE naturally gives more weights on the thresholds that provide more accurate information. To gain insights, suppose hypothetically that the estimators $\theta_N^{\{i\}}$, $i = 1, \ldots, m_0$, are independent. Then V_N is diagonal: $V_N = \operatorname{diag}(v_N^1, \ldots, v_N^{m_0})$. It follows that the optimal weighting is

$$\beta_N = \frac{[(v_N^1)^{-1}, \ldots, (v_N^{m_0})^{-1}]'}{(v_N^1)^{-1} + \cdots + (v_N^{m_0})^{-1}}. \tag{6.11}$$

In other words, the estimators with smaller variances will be more heavily weighted.

6.3 Alternative Covariance Expressions of Optimal QCCEs

Recall from Section 2.2 that $\widetilde{s}_k^{\{i\}} = I_{\{C_{i-1}<y_k\le C_i\}}$, $i = 1, \ldots, m_0 + 1$, with $C_0 = -\infty$ and $\widetilde{s}_k(m_0 + 1) = I_{\{C_{m_0}<y_k<\infty\}}$. Let

$$\begin{aligned}\widetilde{p}_i &= P\{\widetilde{s}_k^{\{i\}} = 1\} = P\{C_{i-1} < y_k \le C_i\}\\ &= F(C_i - \theta) - F(C_{i-1} - \theta) := \widetilde{F}_i(\theta).\end{aligned}$$

Define

$$\widetilde{h}_i(\theta) = \frac{\partial \widetilde{F}_i(\theta)}{\partial \theta} = -f(C_i - \theta) + f(C_{i-1} - \theta).$$

The following relationships between $\widetilde{p}_i, \widetilde{h}_i$ and p_i, h_i in (6.3) can be easily established:

$$\begin{aligned}&\sum_{i=1}^{m_0+1} \widetilde{p}_i = 1, \quad \sum_{i=1}^{m_0+1} \widetilde{h}_i = 0,\\ &p_j = \sum_{i=1}^{j} \widetilde{p}_i, \quad h_j = \sum_{i=1}^{j} \widetilde{h}_i, j = 1, \ldots, m_0,\\ &\widetilde{p}_1 = p_1, \quad \widetilde{p}_i = p_i - p_{i-1}, i = 2, \ldots, m_0,\\ &\widetilde{p}_{m_0+1} = 1 - p_{m_0}.\end{aligned}$$

Denote

$$\begin{aligned}\widetilde{p}(\theta) &= [\widetilde{p}_1, \ldots, \widetilde{p}_{m_0}]',\\ \widetilde{U}(\theta) &= \operatorname{diag}(1/\widetilde{h}_1, \ldots, 1/\widetilde{h}_{m_0}),\\ \widetilde{M}(\theta) &= \operatorname{diag}(\widetilde{p}_1, \ldots, \widetilde{p}_{m_0}),\end{aligned}$$

and

$$\widetilde{\Psi}(\theta) = \widetilde{U}(\widetilde{M} - \widetilde{p}\widetilde{p}')\widetilde{U}. \tag{6.12}$$

For notational simplicity, we suppress the θ dependence in the expressions. Assume that $\widetilde{p}_i \neq 0$, $i = 1, \ldots, m_0 + 1$. If $\widetilde{p}_j = 0$, the threshold C_j can be eliminated since the interval $(C_{i-1}, C_j]$ contains no useful information on θ. Then, $\widetilde{M}$ is invertible. Let

$$\widetilde{W} = \widetilde{M}^{1/2} = \operatorname{diag}(\sqrt{\widetilde{p}_1}, \ldots, \sqrt{\widetilde{p}_{m_0}}).$$

We now prove an important lemma that will be essential for establishing the efficiency of the optimal QCCE algorithms in the next section.

Lemma 6.4. $\mathbb{1}'\widetilde{\Psi}^{-1}\mathbb{1} = \sum_{i=1}^{m_0+1} \widetilde{h}_i^2/\widetilde{p}_i.$

Proof. Note that

$$\begin{aligned}\widetilde{\Psi}^{-1} &= (\widetilde{U}(\widetilde{W}^2 - \widetilde{p}\widetilde{p}')\widetilde{U})^{-1} \\ &= (\widetilde{W}\widetilde{U})^{-1}(I - \widetilde{W}^{-1}\widetilde{p}\widetilde{p}'\widetilde{W}^{-1})^{-1}(\widetilde{U}\widetilde{W})^{-1}.\end{aligned}$$

By the well-known matrix inversion lemma ([62, p. 306, eq. (11.10)]),

$$(I - \widetilde{W}^{-1}\widetilde{p}\widetilde{p}'\widetilde{W}^{-1})^{-1} = I + \frac{\widetilde{W}^{-1}\widetilde{p}\widetilde{p}'\widetilde{W}^{-1}}{1 - \widetilde{p}'\widetilde{M}^{-1}\widetilde{p}}.$$

Observe that

$$\begin{aligned}&1 - \widetilde{p}'\widetilde{M}^{-1}\widetilde{p} = 1 - \sum_{i=1}^{m_0}\widetilde{p}_i = \widetilde{p}_{m_0+1}, \\ &\widetilde{p}'\widetilde{W}^{-1}(\widetilde{U}\widetilde{W})^{-1}\mathbb{1} = [\sqrt{\widetilde{p}_1}, \ldots, \sqrt{\widetilde{p}_{m_0}}]\Big[\frac{\widetilde{h}_1}{\sqrt{\widetilde{p}_1}}, \ldots, \frac{\widetilde{h}_{m_0}}{\sqrt{\widetilde{p}_{m_0}}}\Big]' = \sum_{i=1}^{m_0}\widetilde{h}_i,\end{aligned}$$

and

$$\mathbb{1}'(\widetilde{W}\widetilde{U})^{-1}(\widetilde{U}\widetilde{W})^{-1}\mathbb{1} = \sum_{i=1}^{m_0}\frac{\widetilde{h}_i^2}{\widetilde{p}_i}.$$

Consequently,

$$\begin{aligned}\mathbb{1}'\widetilde{\Psi}^{-1}\mathbb{1} &= \mathbb{1}'(\widetilde{U}(\widetilde{W}^2 - \widetilde{p}\widetilde{p}')\widetilde{U})^{-1}\mathbb{1} \\ &= \mathbb{1}'(\widetilde{W}\widetilde{U})^{-1}(I - \widetilde{W}^{-1}\widetilde{p}\widetilde{p}'\widetilde{W}^{-1})^{-1}(\widetilde{U}\widetilde{W})^{-1}\mathbb{1} \\ &= \sum_{i=1}^{m_0}\frac{\widetilde{h}_i^2}{\widetilde{p}_i} + \frac{(\sum_{i=1}^{m_0}\widetilde{h}_i)^2}{\widetilde{p}_{m_0+1}}.\end{aligned}$$

However, from $\sum_{i=1}^{m_0+1}\widetilde{p}_i = 1$, we have

$$\sum_{i=1}^{m_0+1}\widetilde{h}_i = \sum_{i=1}^{m_0+1}\frac{\partial\widetilde{p}_i}{\partial\theta} = 0,$$

or $\widetilde{h}_{m_0+1} = -\sum_{i=1}^{m_0}\widetilde{h}_i$. It follows that

$$\frac{(\sum_{i=1}^{m_0}\widetilde{h}_i)^2}{\widetilde{p}_{m_0+1}} = \frac{\widetilde{h}_{m_0+1}^2}{\widetilde{p}_{m_0+1}}.$$

Therefore,

$$\mathbb{1}'\widetilde{\Psi}^{-1}\mathbb{1} = \sum_{i=1}^{m_0+1}\frac{\widetilde{h}_i^2}{\widetilde{p}_i}, \tag{6.13}$$

which completes the proof. □

We have introduced $\widetilde{\Psi}$ to provide a convenient way of expressing the CR lower bound in the next section. Next, we establish the connection between $\widetilde{\Psi}$ in (6.12) and Ψ in (6.5). Denote $\widetilde{h} = [\widetilde{h}_1, \ldots, \widetilde{h}_{m_0}]'$, and refer to (6.3)–(6.5) for related expressions.

Lemma 6.5. $\mathbb{1}'\Psi^{-1}\mathbb{1} = \mathbb{1}'\widetilde{\Psi}^{-1}\mathbb{1}$.

Proof. First, note that

$$\mathbb{1}'\Psi^{-1}\mathbb{1} = h'(M - pp')^{-1}h,$$
$$\mathbb{1}'\widetilde{\Psi}^{-1}\mathbb{1} = \widetilde{h}'(\widetilde{M} - \widetilde{p}\widetilde{p}')^{-1}\widetilde{h}.$$

Let V_1 be the matrix

$$V_1 = \begin{bmatrix} 1 & -1 & \dots & -1 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

which is formed by subtracting columns $i = 2, \dots, m_0$ from the first column of the identity matrix. Then,

$$h'(M - pp')^{-1}h = h'V_1(V_1'MV_1 - V_1'pp'V_1)^{-1}V_1'h.$$

It is easy to verify that

$$V_1'MV_1 = \begin{bmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 - p_1 & \dots & p_2 - p_1 \\ \vdots & & & \vdots \\ 0 & p_2 - p_1 & \dots & p_{m_0} - p_1 \end{bmatrix} := M_1,$$

$$V_1'p = [p_1, p_2 - p_1, \dots, p_{m_0} - p_1]' := W_1,$$
$$V_1'h = [h_1, h_2 - h_1, \dots, h_{m_0} - h_1]' := H_1.$$

Hence,

$$\mathbb{1}'\Psi^{-1}\mathbb{1} = H_1'(M_1 - W_1W_1')H_1.$$

This process can be repeated, but restricting to the lower right $(m_0 - 1) \times (m_0 - 1)$ submatrix of M_1. In other words, using

$$V_2 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & -1 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix},$$

we obtain

$$\begin{aligned}\mathbb{1}'\Psi^{-1}\mathbb{1} &= H_1'V_2(V_2'M_1V_2 - V_2'W_1W_1'V_2)^{-1}V_2'H_1\\ &= H_2'(M_2 - W_2W_2')H_2.\end{aligned}$$

After $(m_0 - 1)$ such elementary operations, we obtain

$$M_{m_0-1} = \begin{bmatrix} p_1 & 0 & 0 & \dots & 0 \\ 0 & p_2 - p_1 & 0 & \dots & 0 \\ 0 & 0 & p_3 - p_2 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \dots & p_{m_0} - p_{m_0-1} \end{bmatrix} = \widetilde{M},$$

$$W_{m_0-1} = [p_1, p_2 - p_1, \dots, p_{m_0} - p_{m_0-1}]' = \widetilde{p},$$

$$H_{m_0-1} = [h_1, h_2 - h_1, \dots, h_{m_0} - h_{m_0-1}]' = \widetilde{h},$$

since $\widetilde{p}_1 = p_1$, $\widetilde{p}_i = p_i - p_{i-1}$, $i = 2, \dots, m_0$. Consequently,

$$\mathbb{1}'\Psi^{-1}\mathbb{1} = \widetilde{h}'(\widetilde{M} - \widetilde{p}\widetilde{p}')^{-1}\widetilde{h} = \mathbb{1}'\widetilde{\Psi}^{-1}\mathbb{1}.$$

□

6.4 Cramér–Rao Lower Bounds and Asymptotic Efficiency of the Optimal QCCE

We first recall the notion of efficiency from estimation theory. Suppose that $X_1, \dots, X_N$ is a random sample of size N from a distribution with probability density function $f(x; \vartheta)$, where ϑ is an unknown parameter. For two unbiased estimators $\widehat{\vartheta}_i$ of ϑ, with $E\widehat{\vartheta}_i^2 < \infty$ $(i = 1, 2)$, we say that $\widehat{\vartheta}_1$ is more efficient than $\widehat{\vartheta}_2$ if the relative efficiency $\mathrm{eff}_\vartheta(\widehat{\vartheta}_1|\widehat{\vartheta}_2) < 1$, where

$$\mathrm{eff}_\vartheta(\widehat{\vartheta}_1|\widehat{\vartheta}_2) = \frac{\mathrm{var}_\vartheta(\widehat{\vartheta}_1)}{\mathrm{var}_\vartheta(\widehat{\vartheta}_2)}.$$

Let $\widehat{\vartheta} = \vartheta(X_1, \dots, X_m)$ be an unbiased estimator of ϑ. Under certain regularity conditions (that are usually fulfilled for the identification problems we are working with), the well-known Cramér–Rao (CR) bound states that the variance $\sigma^2_{\widehat{\vartheta}}$ of the estimator $\widehat{\vartheta}$ is bounded below by

$$\sigma^2_{\widehat{\vartheta}} \geq \frac{1}{NE[(\partial/\partial\vartheta)f(X;\vartheta)]^2}. \tag{6.14}$$

We say that an estimator $\widehat{\vartheta}$ is efficient or most efficient if the CR bound is attained. For any unbiased estimator $\widehat{\vartheta}_1$, the efficiency of the estimator is defined as $\text{eff}_\vartheta(\widehat{\vartheta}_1|\widehat{\vartheta})$, where $\widehat{\vartheta}$ is an efficient estimator. An estimator $\widehat{\vartheta}_1$ with sample size N is asymptotically efficient if

(i) $\widehat{\vartheta}_1$ is at least asymptotically unbiased in the sense that $E\widehat{\vartheta}_1 \to \vartheta$ as $N \to \infty$, and

(ii) $\lim_N \text{eff}_\vartheta(\widehat{\vartheta}_1|\widehat{\vartheta}) = 1$.

For further discussion on related issues, we refer the reader to [78, Section 8.5] and the references therein.

Let $\widetilde{\xi}_N^{\{i\}} = \frac{1}{N}\sum_{k=1}^N \widetilde{s}_k^{\{i\}}$, which is the sample relative frequency of y_k taking values in $(C_{i-1}, C_i]$. In the statement of the following lemma, we note that information contained in $\{s_k\}$ is the same as that in $\{\widetilde{s}_k\}$.

Lemma 6.6. *The CR lower bound for estimating θ based on observations of $\{s_k\}$ is*

$$\sigma^2_{\text{CR}}(N, m_0) = \left(N \sum_{i=1}^{m_0+1} \frac{\widetilde{h}_i^2}{\widetilde{p}_i}\right)^{-1}. \tag{6.15}$$

Proof. Augment s_k to $s_{ak} = [s_k', 1]'$, where the added element represents $1 = P\{-\infty < y_k < \infty\}$. Let $x_k \in \mathbb{R}^{m_0+1}$ be some possible sample values of s_{ak}. Noting the i.i.d. assumption, the likelihood function of $s_{a1}, \ldots, s_{aN}$ taking values $x_1, \ldots, x_N$, conditioned on θ, is given by

$$\begin{aligned} \ell(x_1, \ldots, x_N; \theta) &= P\{s_{a1} = x_1, \ldots, s_{aN} = x_N; \theta\} \\ &= \prod_{k=1}^N P\{s_{ak} = x_k; \theta\}. \end{aligned}$$

Due to the sensor structure, x_k always takes the form of $[0, \ldots, 0, 1, 1, \ldots, 1]'$. Let $i_0(k)$ be the index of the first 1 in x_k. Then

$$P\{s_{ak} = x_k; \theta\} = P\{\widetilde{s}_k(i_0(k)) = 1; \theta\} = \widetilde{p}_{i_0(k)}.$$

Hence,

$$\ell(x_1, \ldots, x_N; \theta) = \prod_{k=1}^N \widetilde{p}_{i_0(k)}. \tag{6.16}$$

Replace the particular realizations x_k by their corresponding random elements v_k, and denote the resulting quantity by $\ell = \ell(v_1, \ldots, v_N; \theta)$. Note that ℓ is random by its definition. In (6.16), for a given i, $i_0(k) = i$ if and only if $\widetilde{s}_k^{\{i\}} = 1$. Consequently, for a given sample path, the number of occurrences of a particular $\widetilde{p}_i$ in (6.16) is $\sum_{k=1}^N \widetilde{s}_k^{\{i\}} = \widetilde{\xi}_N^{\{i\}} N$. As a result, by grouping all occurrences of $\widetilde{p}_i$ in (6.16), we have

$$\ell = \prod_{i=1}^{m_0+1} \widetilde{p}_i(\widetilde{\xi}_N^{\{i\}} N).$$

Consequently, $\log \ell = N \sum_{i=1}^{m_0+1} \widetilde{\xi}_N^{\{i\}} \log \widetilde{p}_i$, and

$$\frac{\partial \log \ell}{\partial \theta} = N \sum_{i=1}^{m_0+1} \widetilde{\xi}_N^{\{i\}} \frac{1}{\widetilde{p}_i} \widetilde{h}_i,$$
$$\frac{\partial^2 \log \ell}{\partial \theta^2} = N \sum_{i=1}^{m_0+1} \widetilde{\xi}_N^{\{i\}} \left[\frac{-1}{\widetilde{p}_i^2} \widetilde{h}_i^2 + \frac{1}{\widetilde{p}_i} \frac{\partial^2 \widetilde{p}_i}{\partial \theta^2} \right].$$

Since $\sum_{i=1}^{m_0+1} \widetilde{p}_i = 1$, we have

$$\sum_{i=1}^{m_0+1} \frac{\partial^2 \widetilde{p}_i}{\partial \theta^2} = 0.$$

As a result,

$$E \sum_{i=1}^{m_0+1} \frac{\widetilde{\xi}_N^{\{i\}}}{\widetilde{p}_i} \frac{\partial^2 \widetilde{p}_i}{\partial \theta^2} = \sum_{i=1}^{m_0+1} \frac{\partial^2 \widetilde{p}_i}{\partial \theta^2} = 0.$$

Hence,

$$E \frac{\partial^2 \log \ell}{\partial \theta^2} = -N \sum_{i=1}^{m_0+1} \frac{\widetilde{h}_i^2}{\widetilde{p}_i}.$$

The CR lower bound is then given by

$$\sigma_{\mathrm{CR}}^2(N, m_0) = -\left(E \frac{\partial^2 \log \ell}{\partial \theta^2} \right)^{-1} = \left(N \sum_{i=1}^{m_0+1} \frac{\widetilde{h}_i^2}{\widetilde{p}_i} \right)^{-1}. \qquad (6.17)$$

□

Recall that from Theorem 6.2, the variance of the optimal QCCE is

$$\sigma_N^2 = \frac{1}{\mathbb{1}' V_N^{-1}(\theta) \mathbb{1}}.$$

One of the main results of this chapter is the following theorem, which reveals that the optimal QCCE is asymptotically efficient.

Theorem 6.7. *The optimal QCCE is asymptotically efficient in the sense that*

$$N\sigma_N^2 - N\sigma_{\mathrm{CR}}^2(N, m_0) \to 0 \quad \text{as } N \to \infty.$$

Proof. By Theorems 6.1 and 6.2, the variance of the optimal QCCE satisfies

$$\begin{aligned} N\sigma_N^2 &= N\frac{1}{\mathbb{1}'V_N^{-1}(\theta)\mathbb{1}} \\ &= \frac{1}{\mathbb{1}'N^{-1}V_N^{-1}(\theta)\mathbb{1}} \\ &\to \frac{1}{\mathbb{1}'\Psi^{-1}(\theta)\mathbb{1}} \quad \text{as} \quad N \to \infty, \end{aligned}$$

where $\Psi^{-1}(\theta)$ is the limit of $N^{-1}V_N^{-1}(\theta)$. On the other hand, by Lemma 6.6,

$$N\sigma_{\mathrm{CR}}^2(N, m_0) = \left(\sum_{i=1}^{m_0+1} \frac{\widetilde{h}_i^2}{\widetilde{p}_i}\right)^{-1}. \tag{6.18}$$

Now, Lemmas 6.4 and 6.5 yield

$$\mathbb{1}'\Psi^{-1}\mathbb{1} = \sum_{i=1}^{m_0+1} \frac{\widetilde{h}_i^2}{\widetilde{p}_i},$$

which leads to the desired result. □

Remark 6.8. Expression (6.18) delineates the contribution of each sensor interval $(C_{i-1}, C_i]$ to the reduction of identification errors as $\widetilde{h}_i^2(\theta)/\widetilde{p}_i(\theta)$. The smaller the sensitivity $\widetilde{h}_i^2/\widetilde{p}_i$, the less useful is the threshold interval $(C_{i-1}, C_i]$ in error reduction. This may be used as a guide in deciding if increasing the quantization accuracy (that is, adding more thresholds) is worthwhile. Furthermore, the quantity $\sum_{i=1}^{m_0+1} \widetilde{h}_i^2/\widetilde{p}_i$ can be used for threshold selection.

A numerically less complex implementation of the optimal QCCE algorithm is to use the sample mean and sample covariance in place of θ and V_N. From the estimates $\{\Theta_j = G(\xi_j), j = 1, \ldots, N\}$, we compute its arithmetic average $\overline{\Theta}_N = \sum_{j=1}^N \Theta_j/N$. Since Θ_N is asymptotically unbiased, by elementary analysis, we also have $\sum_{j=1}^N E\Theta_j/N \to \theta\mathbb{1}$ as $N \to \infty$. This leads to the following algorithm:

$$\begin{aligned} \overline{\Theta}_N &= \sum_{j=1}^N \Theta_j/N, \\ \widehat{V}_N &= \frac{1}{N-1}\sum_{j=1}^N (\Theta_j - \overline{\Theta}_N)(\Theta_j - \overline{\Theta}_N)', \\ \beta_N &= \frac{\widehat{V}_N^{-1}\mathbb{1}}{\mathbb{1}'\widehat{V}_N^{-1}\mathbb{1}}, \\ \widehat{\theta}_N &= (\beta_N)'\Theta_N. \end{aligned} \tag{6.19}$$

This algorithm can be written recursively as

$$\overline{\Theta}_N = \overline{\Theta}_{N-1} - \frac{1}{N}\overline{\Theta}_{N-1} + \frac{\Theta_N}{N},$$
$$\widehat{V}_N = \widehat{V}_{N-1} - \frac{1}{N-1}\widehat{V}_{N-1} + \frac{(\Theta_N - \overline{\Theta}_N)(\Theta_N - \overline{\Theta}_N)'}{N-1}. \tag{6.20}$$

It can be shown that

$$\widehat{V}_N - V_N(\theta) \to 0, \quad \widehat{V}_N^{-1} - V_N^{-1}(\theta) \to 0, \text{ as } N \to \infty. \tag{6.21}$$

If we work with the implementable estimates defined in (6.19), we can define $\widehat{\sigma}_N^2 = 1/(\mathbb{1}'\widehat{V}_N^{-1}\mathbb{1})$, and obtain the following result.

Corollary 6.9. *For the estimators given in* (6.19)*,*

$$N\widehat{\sigma}_N^2 - N\sigma_{\text{CR}}^2(N, m_0) \to 0 \quad \textit{as } N \to \infty.$$

Proof. By virtue of Theorem 6.7, it suffices to show that $N\widehat{\sigma}_N^2 - N\sigma_N^2 \to 0$ as $N \to \infty$. In fact, a simple calculation shows that

$$\frac{1}{\mathbb{1}'\widehat{V}_N^{-1}\mathbb{1}} - \frac{1}{\mathbb{1}'V_N^{-1}\mathbb{1}} = \frac{\mathbb{1}'(V_N^{-1} - \widehat{V}_N^{-1})\mathbb{1}}{(\mathbb{1}'\widehat{V}_N^{-1}\mathbb{1})(\mathbb{1}'V_N^{-1}\mathbb{1})} \to 0 \text{ as } N \to \infty. \tag{6.22}$$

Hence, the result follows. □

6.5 Notes

This chapter introduces the optimal quasi-convex combination estimators and establishes their asymptotic efficiency. These optimality results lay a foundation in which complexity issues in system identification with quantized observations can be rigorously investigated. This chapter follows [104].

Space complexity (the number of intervals in quantization, or the word length of each measurement) is a relatively new paradigm in system identification. Traditional quantization uses uniform quantization intervals and ubiquitously employs quantization errors in analysis. However, when the signal range is large or even unbounded, such uniform quantization suffers from high or infinite space complexity. The results of this chapter provide a foundation to evaluate if finite quantization levels are sufficient. Studies of the impact of quantization errors can also be conducted in a worst-case or probabilistic framework, depending on how quantization errors are modeled [1, 2, 39, 34, 80].