# Proofs of Lemma III.2 and Proposition III.3 in "Electronic Component Parameter Recognition by a Nonparametric Approach with an Assignment Algorithm"

Shih-Feng Huang, Tzung-Pei Hong, *Senior Member, IEEE,* Hao-Tse Hsueh, and Yi-Ting Chen

**Lemma III.2.** *Let $p_{i,(k)} = \max_{j\in\{1,\dots,k\}} p_{i,j}$ and $\hat{p}_{i,(k)} = \max_{j\in\{1,\dots,k\}} \hat{p}_{i,j}$, where $\hat{p}_{i,j}$ is defined in (9). Let $\hat{p}_{i,h_i^0}$ been defined similar to $p_{i,h_i^0}$ in Proposition III.1 with replacing $p_{i,j}$ by $\hat{p}_{i,j}$. If (C.1) holds and $p_{i,(k)} = p_{i,h_i^0}$ for each $i = 1,\dots,k$, we have $P(\hat{p}_{i,(k)} = \hat{p}_{i,h_i^0}) = 1$ as $n_{i,\cdot} \to \infty$.*

*Proof.* Let $p_{i,h_i^0} = p_{i,1}$ without loss of generality for a fixed $i \in \{1,\dots,k\}$. It suffices to show that $P(\hat{p}_{i,h_i^0} = \hat{p}_{i,1}) = P(A\hat{\mathbf{p}}_{i,n} > 0) \to 1$ as $n \to \infty$, where

$$A = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 1 & \cdots & \cdots & 0 & -1 \end{pmatrix}_{(k-1)\times k}.$$

and $\hat{\mathbf{p}}_{i,n}$ defined in (9) is the MLE of $\mathbf{p}_i$ under (C.1). In addition, let $\widehat{\Sigma}_n$ be the estimator of $\Sigma$ with replacing the $p_{i,j}$ in (10) by $\hat{p}_{i,j}$. To simplify the notation, we skip the subscript $i$ in the following derivation since the following arguments hold for any fixed $i \in \{1,\dots,k\}$.

By (9), we have $\mathrm{E}(A\hat{\mathbf{p}}_n) = A\mathbf{p} > 0$, where the inequality holds by the assumption of $p_{h^0} = p_1$. In addition, $\mathrm{Var}(A\hat{\mathbf{p}}_n)$ can be approximated well by $B_n = n^{-1}A\Sigma A^\top$ for sufficiently large $n$. Furthermore, since there still exist unknown parameters in $\Sigma$, $B_n$ is usually estimated by $\hat{B}_n = n^{-1}A\widehat{\Sigma}_n A^\top$. By the Strong Law of Large Numbers, $||\hat{\mathbf{p}}_n - \mathbf{p}||_2 = o_p(1)$ and thereby $||\widehat{\Sigma}_n - \Sigma||_F = o_p(1)$, where $||\cdot||_2$ denotes the $L_2$ norm of a vector and $||\cdot||_F$ denotes the Frobenius norm of a matrix. Therefore, $||\hat{B}_n - B_n||_F = o_p(n^{-1})$ as $n \to \infty$.

Since $B_n$ is p.d. and by Cholesky decomposition, there exists a lower triangular matrix $L_n$ with size $(k-1)\times(k-1)$ such that $B_n = L_n L_n^\top$, which indicates that $||L_n L_n^\top - \hat{B}_n||_F = o_p(n^{-1})$. Therefore, we have $P(A\hat{\mathbf{p}}_n > 0) \approx P(\mathbf{Z} > -L_n^{-1}\boldsymbol{\mu})$, where $\mathbf{Z}$ is a $(k-1)$-dimensional standard normal random vector and $\boldsymbol{\mu} = A\mathbf{p}$ is a $(k-1)\times 1$ positive vector. In addition,

$$L_n^{-1} = (I_{k-1} - H_n)(\mathrm{diag}(L_n))^{-1}, \tag{III.1}$$

where $H_n$ is a $(k-1)\times(k-1)$ lower triangular matrix with $(i,j)$ element being $-\ell_n(i,j)/\ell_n(i,i)$ for $j < i$ and $i = 2,\dots,k-1$, and $\ell_n(i,j)$ is the $(i,j)$ component of $L_n$. Since $B_n = L_n L_n^\top$ and $||B_n||_F = O(n^{-1})$, we have $\ell_n(i,j) = O(n^{-1/2})$ for $i \le j$. This together with (III.1) lead to $\inf(L_n^{-1}\boldsymbol{\mu}) = O(n^{1/2})$. Consequently, we have $P(A\hat{\mathbf{p}}_n > 0) \approx P(\mathbf{Z} > -L_n^{-1}\boldsymbol{\mu}) \to 1$ as $n \to \infty$. The proof is complete. □

**Proposition III.3.** *Under (A.1), (A.2), (C.1), and (C.2), if $\hat{\sigma}^2_{c|d,i,h_i^0} \ge \hat{\sigma}^2_{c|d,i,j}$ and*

$$\Phi\Big(\frac{|\hat{\Delta}_{i,j}|}{2\hat{\sigma}_{c|d,i,h_i^0}}\Big) > \frac{\hat{r}_{i,j}}{\hat{r}_{i,j}+1}, \tag{III.2}$$

*or if $\hat{\sigma}^2_{c|d,i,h_i^0} < \hat{\sigma}^2_{c|d,i,j}$ and*

$$\Phi\Big(\frac{|\hat{\Delta}_{i,j}|}{2\hat{\sigma}_{c|d,i,j}}\Big) > \frac{\hat{r}_{i,j}+0.5}{\hat{r}_{i,j}+1}, \tag{III.3}$$

*we have*

$$P\Big(\min_{j\in\{1,\dots,k\}} \frac{\hat{p}_{i,h_i^0}}{\hat{p}_{i,j}} \times \frac{\hat{q}_{i,h_i^0}}{\hat{q}_{i,j}} > 1\Big) = 1, \text{ as } n_{i,j} \to \infty, \tag{III.4}$$

*where $\Phi(\cdot)$ is the distribution function of $N(0,1)$, $I_A$ denotes an indicator function of an event $A$, $\hat{p}_{i,h_i^0}$ and $\hat{p}_{i,j}$ are defined the same as in Lemma III.2, and $\hat{q}_{i,j}$, $\hat{\Delta}_{i,j} = \hat{\mu}_{c|d,i,h_i^0} - \hat{\mu}_{c|d,i,j}$, and $\hat{r}_{i,j} = \hat{p}_{i,j}/\hat{p}_{i,h_i^0}$ are the MLEs of $q_{i,j}$, $\Delta_{i,j} = \mu_{c|d,i,h_i^0} - \mu_{c|d,i,j}$, and $r_{i,j} = p_{i,j}/p_{i,h_i^0}$, respectively, for $i,j = 1,\dots,k$.*

*Proof.* Without loss of generality, consider the case of $i = 1$ and let $q_{1,h_1^0} = q_{1,1}$. To simplify the illustration, we skip the subscripts $c|d$ and $i$ in the following derivation since the following arguments hold for any fixed $i \in \{1,\dots,k\}$.

In the case of $n_c = 1$, the $d(\mathbf{X}_{c,i}, \boldsymbol{\mu}_{c|d,i,j})$ in (A.2) is simplified as $d(X_c,\mu_j) = |X_c - \mu_j|/\sigma_j$. If $\mathbf{X} = (\mathbf{X}_d^\top, X_c)^\top$ should be correctly assigned to $\theta_1$, (1) indicates that $\mathbf{X}_d$ should be assigned to $\theta_1$ and $X_c$ is $N(\mu_1,\sigma_1^2)$ distributed. In this case and by (A.2), it suffices to show that

$$q_1 = P(X_c \to \theta_1 \mid E_1) > \max_{j=2,\dots,k} r_j q_j, \tag{III.5}$$

where $r_j = p_j/p_1 > 0$.

On the LHS of (III.5), we need to consider the event of $\{d(X_c,\mu_1) < \min_{j=2,\dots,k} d(X_c,\mu_j)\}$. For any $j \in$

Shih-Feng Huang is with the Graduate Institute of Statistics, National Central University, Taoyuan 320, Taiwan (email: huangsf@ncu.edu.tw). Tzung-Pei Hong is with the Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan, and also with the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan. Hao-Tse Hsueh is with the Institute of Statistics, National University of Kaohsiung, Kaohsiung 811, Taiwan. Yi-Ting Chen is with the Footprintku Inc., Kaohsiung 806, Taiwan.

$\{2,\ldots,k\}$, since $\{d(X_c,\mu_1) < d(X_c,\mu_j)\} = \{a_j Z^2 + 2\sigma_1\Delta_j Z + \Delta_j^2 > 0\}$, where $a_j = \sigma_1^2 - \sigma_j^2$, $\Delta_j = \mu_1 - \mu_j$, and $Z$ is a standard normal random variable. If $a_j = 0$ (that is, $\sigma_1 = \sigma_j$), we have

$$P\Big(d(X_c,\mu_1) < d(X_c,\mu_j)\Big) = \Phi\Big(\frac{|\Delta_j|}{2\sigma_1}\Big) := \Phi_j, \quad \text{(III.6)}$$

where $\Phi(\cdot)$ is the distribution function of $N(0,1)$. If $a_j > 0$ (that is, $\sigma_1 > \sigma_j$), we have

$$P\Big(d(X_c,\mu_1) < d(X_c,\mu_j)\Big) > \Phi_j. \quad \text{(III.7)}$$

If $a_j < 0$ (that is, $\sigma_1 < \sigma_j$), we have

$$P\Big(d(X_c,\mu_1) < d(X_c,\mu_j)\Big) > 2\Phi^* - 1, \quad \text{(III.8)}$$

where $\Phi_j^* = \Phi(|\Delta_j|/(2\sigma_j))$. Therefore, (III.6)-(III.8) reveal that

$$\begin{aligned} &P\Big(d(X_c,\mu_1) < d(X_c,\mu_j)\Big) \\ &\qquad \geq \Phi_j I_{\{\sigma_1^2 \geq \sigma_j^2\}} + \Big(2\Phi_j^* - 1\Big) I_{\{\sigma_1^2 < \sigma_j^2\}}, \end{aligned} \quad \text{(III.9)}$$

which tends to 1 for sufficiently large $|\Delta_j|/\max(\sigma_1,\sigma_j)$, $j = 2,\ldots,k$. Consequently, if $\min_{\ell=1,\ldots,k} |\Delta_j|/\sigma_\ell$ is sufficiently large, $q_1 = P(d(X_c,\mu_1) < \min_{j=2,\ldots,k} d(X_c,\mu_j))$ also approaches to 1.

On the RHS of the inequality of (III.5), we need to consider the event of $\{\min_{\ell\in\{1,\ldots,k\},\ell\neq j} d(X_c,\mu_\ell) > d(X_c,\mu_j)\}$ under the consideration that $\mathbf{X}_d$ should be assigned to $\theta_1$ but now is incorrectly assigned to $\theta_j$, and $X_c$ is $N(\mu_1,\sigma_1^2)$ distributed. Note that $q_j \leq q_j^1 = P(d(X_c,\mu_1) > d(X_c,\mu_j))$ and (III.9) leads to

$$q_j^1 \leq \Big(1-\Phi_j\Big) I_{\{\sigma_1^2 \geq \sigma_j^2\}} + 2\Big(1-\Phi_j^*\Big) I_{\{\sigma_1^2 < \sigma_j^2\}}. \quad \text{(III.10)}$$

By (III.5), (III.9), and (III.10), the desired result $\min_{j=2,\ldots,k}\{p_1 q_1/(p_j q_j)\} > 1$ holds in the population perspective if $q_1 > r_j q_j^1$ for all $j = 2,\ldots,k$, which holds if

$$\Phi_j > \frac{r_j}{r_j+1} I_{\{\sigma_1^2 \geq \sigma_j^2\}} \ \text{ or } \ \Phi_j^* > \frac{r_j+0.5}{r_j+1} I_{\{\sigma_1^2 < \sigma_j^2\}}, \quad \text{(III.11)}$$

for all $j = 2,\ldots,k$, if the parameters $\mu_j$, $\sigma_j$, and $p_j$ for all $j = 1,\ldots,k$ are given.

The final step of the proof is to consider the estimation of $\mu_j$, $\sigma_j$, and $p_j$, $j = 1,\ldots,k$. By Lemma III.2 and (C.2), $\hat{p}_j$, $\hat{\mu}_j$, and $\hat{\sigma}_j$ are the MLEs of $p_j$, $\mu_j$, and $\sigma_j$, respectively, for $j = 1,\ldots,k$. Consequently, we have $|\hat{p}_j - p_j| = o_p(1)$, $|\hat{\mu}_j - \mu_j| = o_p(1)$, and $|\hat{\sigma}_j - \sigma_j| = o_p(1)$ for $j = 1,\ldots,k$. By the invariant property of MLEs, we also have $|\hat{d}(X_c,\hat{\mu}_j) - d(X_c,\mu_j)| = o_p(1)$, $|\hat{r}_j - r_j| = o_p(1)$, $|\hat{\Delta}_j - \Delta_j| = o_p(1)$, $|\hat{\Phi}_j - \Phi_j| = o_p(1)$, and $|\hat{\Phi}_j^* - \Phi_j^*| = o_p(1)$ as $n_j \to \infty$, where $\hat{d}(X_c,\hat{\mu}_j) = |X_c - \hat{\mu}_j|/\hat{\sigma}_j$, $\hat{\Phi}_j = \Phi\Big(|\hat{\Delta}_j|/(2\hat{\sigma}_1)\Big)$, and $\hat{\Phi}_j^* = \Phi\Big(|\hat{\Delta}_j|/(2\hat{\sigma}_j)\Big)$ for $j = 1,\ldots,k$. Therefore, if (III.11) holds, (III.2) and (III.3) follow with probability 1, and thereby $P\Big(\frac{\hat{p}_1}{\hat{p}_j} \times \frac{\hat{q}_1}{\hat{q}_j} > 1\Big) = 1$, as $n_j \to \infty$, for all $j = 2,\ldots,k$. $\square$