On the Converse Theorem in Statistical Hypothesis Testing for Markov Chains

Kenji Nakagawa, *Member, IEEE,* and
Fumio Kanaya, *Member, IEEE*

***Abstract*—Discussion on the converse theorem in statistical hypothesis testing. Hypothesis testing for two Markov chains is considered. Under the constraint that the first-kind error probability is less than or equal to $\exp(-rn)$, the second-kind error probability is minimized. The geodesic that connects the two Markov chains is defined. By analyzing the geodesic, the power exponents are calculated and then represent in terms of Kullback–Leibler divergence.**

***Index Terms*— Hypothesis testing, Markov chain, randomized test, information geometry, geodesic, power exponent, converse theorem.**

I. Introduction

In this continuation of our work [8], we investigate simple statistical hypothesis testing for two Markov chains.

Let a finite set $\Omega = \{0, 1, \cdots, m\}$ be the state space of a Markov chain. Let S be the family of all subsets of Ω, Ω^∞ the set of all one-sided infinite sequences whose components are in Ω, and S^∞ the product σ-algebra of S. If a stationary probability measure on the measurable space $(\Omega^\infty, S^\infty)$ satisfies

$$P(\omega^n) = p(\omega_1)P(\omega_2|\omega_1)\cdots P(\omega_n|\omega_{n-1}), \tag{1}$$

for any sequence $\omega^n = \omega_1\omega_2\cdots\omega_n \in \Omega^n$, the triple (Ω, P, p), or simply P is called a finite-state Markov chain. Here, $P(i'|i)$, $i, i' \in \Omega$, is a conditional probability of i' given i, which is called a state transition probability of P, and $p(i)$, $i \in \Omega$, is a probability measure on Ω, which is called an initial probability measure. In (1), $P(\omega^n)$ should be written as $P([\omega^n])$, where $[\omega^n]$ is the cylinder set, i.e., $[\omega^n] = \{\omega^\infty \in \Omega^\infty |$ the ith component of ω^∞ is ω_i, $i = 1, \cdots, n\}$; however, we write $P(\omega^n)$ for the sake of simplicity.

Simple hypothesis testing for two Markov chains P_0 (null hypothesis) and P_1 (alternative hypothesis) whose state transition probabilities are all positive is considered. Let ϕ_n be a randomized test function of the hypothesis testing, i.e., ϕ_n is a mapping from Ω^n to the closed interval $[0, 1] = \{x | 0 \le x \le 1\}$. From now on, we assume that the term "test function" is always used as a randomized test function. For a test function ϕ_n, the first-kind error probability $\alpha(\phi_n)$ and the second-kind error probability $\beta(\phi_n)$ are defined as follows:

$$\alpha(\phi_n) \triangleq \sum_{\omega^n \in \Omega^n} (1 - \phi_n(\omega^n))P_0(\omega^n), \tag{2}$$

$$\beta(\phi_n) \triangleq \sum_{\omega^n \in \Omega^n} \phi_n(\omega^n)P_1(\omega^n). \tag{3}$$

The most powerful test function under the constraint $\alpha(\phi_n) \le \alpha$, $\alpha > 0$, is defined as a test function ϕ_n^* that satisfies $\alpha(\phi_n^*) \le \alpha$ and $\beta(\phi_n^*) \le \beta(\phi_n)$ for arbitrary ϕ_n with $\alpha(\phi_n) \le \alpha$. In this work, our goals are to determine the most powerful test function ϕ_n^* under the constraint $\alpha(\phi_n^*) \le \exp(-rn)$, where r is a given positive number, and calculate $\lim_{n\to\infty} -\frac{1}{n}\log\beta(\phi_n^*)$ or $\lim_{n\to\infty} -\frac{1}{n}\log(1-\beta(\phi_n^*))$, and represent them in terms of Kullback–Leibler divergence. These limiting values are called the power exponents of hypothesis testing.

Manuscript received March 6, 1990; revised June 1, 1992. This work was presented in part at the IEEE International Symposium on Information Theory, San Diego, CA, January 14–19, 1990.

K. Nakagawa is with Department of Planning & Management of Science, Nagaoka University of Technology, Nagaoka-shi, Niigata-ken, 940-21 Japan.
F. Kanaya is with NTT Transmission Systems Laboratories, 1-2356, Take, Yokosuka-shi, Kanagawa-ken, 238-03 Japan.

IEEE Log Number 9203870.

II. Preliminaries

For a finite sequence $\omega^n = \omega_1\omega_2\cdots\omega_n \in \Omega^n$, the frequency $N(i, i'|\omega^n)$, $i, i' \in \Omega$, is defined by

$$N(i, i'|\omega^n) \triangleq \sum_{k=1}^{n} \delta(ii', \omega_k\omega_{k+1}), \tag{4}$$

where $\delta(ii', \omega_k\omega_{k+1}) = 1$ if $\omega_k = i$ and $\omega_{k+1} = i'$, and $= 0$, otherwise. We regard $n + 1 = 1$. Let us define the second-order type [4] $P_{\omega^n}(i, i')$ of a sequence $\omega^n \in \Omega^n$ by

$$P_{\omega^n}(i, i') \triangleq \frac{1}{n} N(i, i'|\omega^n), \qquad i, i' \in \Omega. \tag{5}$$

It can be readily seen that the second-order type P_{ω^n} has the following properties [9]:

$$\sum_{i, i' \in \Omega} P_{\omega^n}(i, i') = 1. \tag{6}$$

$$\sum_{i' \in \Omega} P_{\omega^n}(i, i') = \sum_{i' \in \Omega} P_{\omega^n}(i', i), \qquad i \in \Omega. \tag{7}$$

We denote by $\overline{\Delta}^{(2)}$ the set of probability measures on $\Omega \times \Omega$ whose two marginal distributions are equal, i.e.,

$$\overline{\Delta}^{(2)} = \left\{ P = (P(i, i'))_{i, i' \in \Omega} | P(i, i') \ge 0, \sum_{i, i' \in \Omega} P(i, i') = 1, \sum_{i' \in \Omega} P(i, i') = \sum_{i' \in \Omega} P(i', i), \; i, i' \in \Omega \right\}. \tag{8}$$

We have $P_{\omega^n} \in \overline{\Delta}^{(2)}$ for any $\omega^n \in \Omega^n$. Let us denote by $\overline{\Delta}_n^{(2)}$ the set of second-order types of n-sequences, i.e., $\overline{\Delta}_n^{(2)} \triangleq \{P_{\omega^n} | \omega^n \in \Omega^n\}$. It is easily verified that $\bigcup_{n\ge 1} \overline{\Delta}_n^{(2)}$ is dense in $\overline{\Delta}^{(2)}$ (see [9]).

A subset $\Delta^{(2)}$ of $\overline{\Delta}^{(2)}$ is defined by

$$\Delta^{(2)} = \{P \in \overline{\Delta}^{(2)} | P(i, i') > 0, \; i, i' \in \Omega\}. \tag{9}$$

According to general theory of Markov chains, if the state transition probabilities $P(i'|i)$ of a Markov chain P are positive for all $i, i' \in \Omega$, there uniquely exists a probability measure $p = (p(i))_{i \in \Omega}$ on Ω which satisfies $\sum_{i \in \Omega} p(i)P(i'|i) = p(i')$, $i' \in \Omega$. We have $p(i) > 0$, $i \in \Omega$. This p is called the stationary probability measure of the Markov chain P.

We consider hypothesis testing for two Markov chains P_0, P_1 with $P_0(i'|i) > 0$, $P_1(i'|i) > 0$, $i, i' \in \Omega$. Denote by p_0, p_1 the stationary probability measures of P_0, P_1, respectively. We use p_0, p_1 as the initial probability of P_0, P_1, respectively.

The most powerful test function under the constraint $\alpha(\phi_n^*) \le \alpha$ is completely determined by the following Neyman–Pearson's lemma.

Lemma 1 (Neyman–Pearson, see [7]): For two Markov chains (Ω, P_0, p_0) and (Ω, P_1, p_1), and for any α, $0 < \alpha < 1$, there exists a test function ϕ_n^* with a constant λ_n that satisfies $\alpha(\phi_n^*) = \alpha$ and

$$\phi_n^*(\omega^n) = \begin{cases} 1, & \text{if } P_0(\omega^n) > \lambda_n P_1(\omega^n), \\ 0, & \text{if } P_0(\omega^n) < \lambda_n P_1(\omega^n). \end{cases} \tag{10}$$

0018–9448/93$03.00

This is the most powerful test function. Write $E = \{\omega^n \in \Omega^n | P_0(\omega^n) = \lambda_n P_1(\omega^n)\}$. If $P_0(E) \triangleq \sum_{\omega^n \in E} P_0(\omega^n) > 0$, we have $\phi_n^*(\omega^n) = \delta$ for $\omega^n \in E$, where

$$\delta = 1 - \frac{\alpha - P_0\{\omega^n | P_0(\omega^n) < \lambda_n P_1(\omega^n)\}}{P_0(E)}. \quad (11)$$

For $P \in \overline{\Delta}^{(2)}$ and a transition probability matrix $Q = Q(i'|i)_{i,\, i' \in \Omega}$, the Kullback–Leibler divergence $D(P\|Q)$ is defined by

$$D(P\|Q) = \sum_{i,\, i' \in \Omega} P(i,\, i') \log \frac{P(i'|i)}{Q(i'|i)}, \quad (12)$$

where $P(i'|i) = P(i,\, i')/p(i)$ if $p(i) > 0$, p is the marginal distribution of P. Particularly, for $P \in \overline{\Delta}^{(2)}$ and $Q \in \Delta^{(2)}$, $D(P\|Q)$ is defined by (12) with $Q(i'|i) = Q(i,\, i')/q(i)$, q is the marginal distribution of Q.

Here, we borrow a result from the large deviation theory for Markov chains. For $P \in \Delta^{(2)}$, the Markov chain with state transition probabilities $P(i'|i) = P(i,\, i')/p(i)$, $i,\, i' \in \Omega$, and p, the stationary probability of P, is stationary ergodic. So, by the Asymptotic Equipartition Property of a stationary ergodic Markov chain, the second-order types of almost all sequences $\omega^n \in \Omega^n$ are "close" to P. Hence, the probability of the set of ω^n's whose second-order types are "far" from P is small. The following lemma evaluates that small probability.

Lemma 2 (Natarajan [9]): Let A be a subset of $\overline{\Delta}^{(2)}$ and $A_n (\neq \emptyset)$ be the subset of A of the second-order types of n-sequences. Then for any $P \in \Delta^{(2)}$, we have

$$\left| \frac{1}{n} \log P\{\omega^n \mid P_{\omega^n} \in A\} + \min_{Q \in A_n} D(Q\|P) \right| \leq \frac{1}{n} \log \left(n^{m+1}(n+1)^{(m+1)^2}(m+1)/\gamma_1\gamma_2\right),$$

where $\gamma_1 = \min_{i \in \Omega} p(i)$, $\gamma_2 = \min_{i,\, i' \in \Omega} (i'|i)$ and p is the stationary probability measure of P. Particularly, if A is a closed set and $\bigcup_{n \geq 1} A_n$ is dense in A, we have

$$\lim_{n \to \infty} -\frac{1}{n} \log P\{\omega^n | P_{\omega^n} \in A\} = \min_{Q \in A} D(Q\|P). \quad (13)$$

According to Neyman–Pearson's lemma, in order to calculate power exponents, it is sufficient to look at only the second-order type P_{ω^n} of ω^n. So, it is more significant to consider the geometry of $\overline{\Delta}^{(2)}$ rather than that of Ω^n. From the general theory of information geometry of $\Delta^{(2)}$ (see [1], [2], [6]), we use here the notion of the $+1$ affine coordinate and the $+1$ geodesic.

The $+1$ affine coordinate of a point P in $\Delta^{(2)}$ denoted by $\theta = (\theta^{ii'})_{i,\, i \in \Omega,\, i' \neq 0}$ is defined [6] by

$$\theta^{ii'} = \log \frac{P(i'|i)P(0|i')}{P(0|i)P(0|0)}, \qquad i,\, i' \in \Omega, \qquad i' \neq 0. \quad (14)$$

The $+1$ geodesic L connecting two points $P_0, P_1 \in \Delta^{(2)}$ is defined, in terms of $+1$ affine coordinates, by

$$\theta_t^{ii'} = (1-t)\theta_0^{ii'} + t\theta_1^{ii'}, \qquad i,\, i' \in \Omega, \qquad i' \neq 0, \qquad t \in \boldsymbol{R}, \quad (15)$$

where $\theta_0 = (\theta_0^{ii'})$, $\theta_1 = (\theta_1^{ii'})$, and $\theta_t = (\theta_t^{ii'})$ are $+1$ affine coordinates of P_0, P_1, and P_t (a point on L), respectively.

By substituting (14) into (15), we have the following alternative expression of the geodesic equation:

$$P_t(i'|i) = \frac{v_t(i')}{\zeta_t v_t(i)} P_0(i'|i)^{1-t} P_1(i'|i)^t, \qquad i,\, i' \in \Omega, \quad (16)$$

where

$$v_t(i) = \frac{P_0(0|i)^{1-t} P_1(0|i)^t}{P_t(0|i)}, \qquad i \in \Omega, \quad (17)$$

and

$$\zeta_t = v_t(0). \quad (18)$$

By taking the summation of both sides of (16) with respect to $i' \in \Omega$, we have

$$\sum_{i' \in \Omega} P_0(i'|i)^{1-t} P_1(i'|i)^t v_t(i') = \zeta_t v_t(i), \qquad i \in \Omega. \quad (19)$$

Thus, we see that ζ_t is a right eigen value of the positive matrix $(P_0(i'|i)^{1-t} P_1(i'|i)^t)_{i,\, i' \in \Omega}$, and v_t is an eigen vector for ζ_t. From (17) and (18), we have $\zeta_t > 0$, $v_t(i) > 0$, $t \in \boldsymbol{R}$, $i \in \Omega$. Then by the positive matrix theory we see that ζ_t is the maximum eigen value with multiplicity 1. So, the geodesic equation (16) coincides with the equation which is given by [9, p. 363].

If the reader is not familiar with the information geometry of the space $\Delta^{(2)}$, he may consider (16) as a definition of the geodesic L that connects $P_0,\, P_1 \in \Delta^{(2)}$.

III. Lemmas

In this section, we provide lemmas on divergence geometry of $\overline{\Delta}^{(2)}$.

Let L be the geodesic connecting $P_0,\, P_1 \in \Delta^{(2)}$, whose equation is given by (16). We have the following lemma.

Lemma 3: $\log \zeta_t$ is a strictly convex function of t, $-\infty < t < \infty$.

Proof: We show

$$\frac{d^2}{dt^2} \log \zeta_t > 0, \qquad -\infty < t < \infty. \quad (20)$$

In fact, by (16), we have

$$\log \zeta_t = -\log P_t(0|0) + (1-t) \log P_0(0|0) + t \log P_1(0|0). \quad (21)$$

By (15), $\theta_t^{ii'}$ is a differentiable function of t, and (14) can be regarded as a differentiable coordinate transformation of $(P(i'|i))_{i,\, i' \in \Omega,\, i' \neq 0}$ and $(\theta^{ii'})_{i,\, i' \in \Omega,\, i' \neq 0}$, hence, we see that $P_t(i'|i)$, $i,\, i' \in \Omega$ are differentiable with respect to t. Therefore, by (21), $\log \zeta_t$ is also differentiable.

Now, by (16), we have

$$\sum_{i,\, i' \in \Omega} P_t(i,\, i') \left\{ \frac{d}{dt} \log P_t(i'|i) \right\}^2 = -\sum_{i,\, i' \in \Omega} P_t(i,\, i') \left\{ \frac{d^2}{dt^2} \log P_t(i'|i) \right\} = \frac{d^2}{dt^2} \log \zeta_t, \quad (22)$$

which leads $d^2/dt^2 \log \zeta_t \geq 0$, $-\infty < t < \infty$. Suppose $d^2/dt^2 \log \zeta_t = 0$ holds for some $t \in \boldsymbol{R}$. From (22), we have

$$\frac{d}{dt} \log P_t(i'|i) = 0, \qquad i,\, i' \in \Omega. \quad (23)$$

Then by (16) and (23), we have

$$\log \frac{P_1(i'|i)}{P_0(i'|i)} = \frac{d}{dt} \log \zeta_t + \frac{d}{dt} \log \frac{v_t(i)}{v_t(i')}, \qquad i,\, i' \in \Omega. \quad (24)$$

By taking the expectation of both sides of (24) with respect to P_0 and P_1, we have

$$D(P_0\|P_1) + D(P_1\|P_0) = 0, \quad (25)$$

a contradiction. Consequently, we have (20).

Lemma 4: Let us write $\sigma_0(t) \triangleq D(P_t\|P_0)$, $\sigma_1(t) \triangleq D(P_t\|P_1)$, and $\tau(t) \triangleq D(P_t\|P_0) - D(P_t\|P_1)$, $-\infty < t < \infty$. We have

a) $\sigma_0(t)$ is strictly increasing for $t > 0$, and strictly decreasing for $t < 0$;
b) $\sigma_1(t)$ is strictly increasing for $t > 1$, and strictly decreasing for $t < 1$;
c) $\tau(t)$ is strictly increasing for t, $-\infty < t < \infty$.

Proof: First, we show that the following relations hold:

$$\sigma_0(t) = -\log \zeta_t + t\frac{d}{dt}\log \zeta_t, \tag{26}$$

$$\sigma_1(t) = -\log \zeta_t + (t-1)\frac{d}{dt}\log \zeta_t, \tag{27}$$

and

$$\tau(t) = \frac{d}{dt}\log \zeta_t. \tag{28}$$

In fact, from (16) we have

$$\frac{d}{dt}\log P_t(i'|i) = \log\frac{P_1(i'|i)}{P_0(i'|i)} - \frac{d}{dt}\log \zeta_t + \frac{d}{dt}\log v_t(i') - \frac{d}{dt}\log v_t(i). \tag{29}$$

Taking the expectation of the both sides of (29) with respect to $P_t(i, i')$, we have

$$0 = \tau(t) - \frac{d}{dt}\log \zeta_t, \tag{30}$$

which implies (28). Again, from (16) we have

$$\log\frac{P_t(i'|i)}{P_0(i'|i)} = t\log\frac{P_1(i'|i)}{P_0(i'|i)} - \log \zeta_t + \log v_t(i') - \log v_t(i). \tag{31}$$

Taking the expectation of the both sides of (31) with respect to $P_t(i, i')$, we have

$$\sigma_0 = t\tau(t) - \log \zeta_t, \tag{32}$$

which proves (26) with the aid of (28). Equation (27) is immediately obtained from (26) and (28).

From (26), we have $(d\sigma_0(t)/dt) = t(d^2/dt^2)\log\zeta_t$, which implies a) by Lemma 3. Similarly, by differentiating both sides of (27) and (28), we have b) and c), respectively. □

For $P \in \overline{\Delta}^{(2)}$, let us define

$$\tau(P) \triangleq \sum_{i,i'\in\Omega} P(i,i')\log\frac{P_1(i'|i)}{P_0(i'|i)}. \tag{33}$$

Particularly, for $P_t \in L$, $\tau(P_t)$ is equal to $\tau(t)$, which was defined in Lemma 4.

The critical condition $P_0(\omega^n) = \lambda_n P_1(\omega^n)$ of the most powerful test function in Neyman–Pearson's lemma is rewritten as

$$\sum_{i,i'\in\Omega} P_{\omega^n}(i,i')\log\frac{P_1(i'|i)}{P_0(i'|i)} = -\frac{1}{n}\log\lambda_n + \frac{1}{n}\log\frac{p_0(\omega_1)P_1(\omega_1|\omega_n)}{p_1(\omega_1)P_0(\omega_1|\omega_n)}. \tag{34}$$

Therefore, we see that it is important to consider sets in $\overline{\Delta}^{(2)}$ of the form $\{P \in \overline{\Delta}^{(2)} | \tau(P) = \tau(\text{constant})\}$.

Since P_0 and $P_1 \in \Delta^{(2)}$, we find that $\tau(t)$ is bounded. By the monotonicity of τ, there exist limiting values of $\tau(t)$ as t tends to infinity or minus infinity. We write these limiting values by $\tau(\infty) = \lim_{t\to\infty}\tau(t)$, $\tau(-\infty) = \lim_{t\to-\infty}\tau(t)$. We have

$$\tau(-\infty) \le \tau(t) \le \tau(\infty), \qquad -\infty \le t \le \infty. \tag{35}$$

Let us write

$$E(t) \triangleq \{P \in \overline{\Delta}^{(2)} | \tau(P) = \tau(t)\}, \tag{36}$$

$-\infty \le t \le \infty$. Here, we have the following lemma.

Lemma 5: For any $P \in E(t)$, $-\infty < t < \infty$,

$$D(P\|P_0) = D(P\|P_t) + D(P_t\|P_0), \tag{37}$$

and

$$D(P\|P_1) = D(P\|P_t) + D(P_t\|P_1) \tag{38}$$

hold.

Proof: The assumption $P \in E(t)$ implies

$$D(P\|P_0) - D(P_t\|P_0) = D(P\|P_1) - D(P_t\|P_1). \tag{39}$$

Hence, from (39), it can be easily seen that for any s, $-\infty < s < \infty$,

$$D(P\|P_0) - D(P_t\|P_0) = D(P\|P_s) - D(P_t\|P_s) \tag{40}$$

holds. Substituting $s = t$ into (40), we have (37). Similarly, from (39) and (40), we have (38). □

Since the functions $\sigma_0(t)$ and $\sigma_1(t)$ are also bounded, there exist the limiting values $\sigma_0(\pm\infty) \triangleq \lim_{t\to\pm\infty}\sigma_0(t)$ and $\sigma_1(\pm\infty) \triangleq \lim_{t\to\pm\infty}\sigma_1(t)$.

Lemma 6: We have

$$\min_{P\in E(t)} D(P\|P_0) = \sigma_0(t), \tag{41}$$

and

$$\min_{P\in E(t)} D(P\|P_1) = \sigma_1(t), \tag{42}$$

for any t, $-\infty \le t \le \infty$.

Proof: For t, $-\infty < t < \infty$, (41) and (42) are readily obtained by (37) and (38), respectively. For $t = \infty$, we now show

$$\min_{P\in E(\infty)} D(P\|P_0) = \sigma_0(\infty). \tag{43}$$

Suppose there exists a $P' \in E(\infty)$ with

$$D(P'\|P_0) < \sigma_0(\infty). \tag{44}$$

Then, by the monotonicity of σ_0 and σ_1, there exist $t_0, t_1 < \infty$ that satisfies

$$D(P'\|P_0) = \sigma_0(t_0), \tag{45}$$

and

$$D(P'\|P_1) = \sigma_1(t_1). \tag{46}$$

If $t_0 \le t_1$, by subtracting (46) from (45), we have $\tau(P') = \sigma_0(t_0) - \sigma_1(t_1) \le \sigma_0(t_1) - \sigma_1(t_1) = \tau(t_1) < \tau(\infty)$, which contradicts the fact $P' \in E(\infty)$. Similarly, if $t_0 > t_1$, we have $\tau(P') \le \tau(t_0) < \tau(\infty)$, a contradiction. Thus, for any $P \in E(\infty)$,

$$\sigma_0(\infty) \le D(P\|P_0) \tag{47}$$

holds. Next, we show that the equality of (47) holds for some $P \in E(\infty)$. In fact, according to the compactness of $\overline{\Delta}^{(2)}$ and the monotonicity of σ_0 and σ_1, there exists a $P_\infty \in \overline{\Delta}^{(2)}$ with

$$D(P_\infty\|P_0) = \sigma_0(\infty), \tag{48}$$

and

$$D(P_\infty\|P_1) = \sigma_1(\infty). \tag{49}$$

Then, from (48) and (49), we have $P_\infty \in E(\infty)$. Hence, the equality of (47) holds for the P_∞. This completes the proof of (43). The other cases are proved in the same way. □

Let us write $G_0(t) \triangleq \{P \in \overline{\Delta}^{(2)} | \tau(P) \le \tau(t)\}$, $G_1(t) \triangleq \{P \in \overline{\Delta}^{(2)} | \tau(P) \ge \tau(t)\}$, for t, $-\infty \le t \le \infty$.

Lemma 7: For t, $-\infty \leq t \leq \infty$, if $P_0 \notin G_1(t)$, we have

$$\min_{P \in G_1(t)} D(P\|P_0) = \sigma_0(t). \tag{50}$$

Similarly, if $P_k \notin G_j(t)$, $k = 0, 1$, $j = 0, 1$, we have

$$\min_{P \in G_j(t)} D(P\|P_k) = \sigma_k(t). \tag{51}$$

Proof: We prove only (50). The other cases are proved similarly. We claim

$$\min_{P \in G_1(t)} D(P\|P_0) = \min_{P \in E(t)} D(P\|P_0), \tag{52}$$

which together with (41) implies (50). First, $E(t) \subset G_1(t)$ leads to $\min_{P \in G_1(t)} D(P\|P_0) \leq \min_{P \in E(t)} D(P\|P_0)$. Suppose there exists a $P' \in G_1(t) - E(t)$ that satisfies $D(P'\|P_0) < \min_{P \in E(t)} D(P\|P_0)$. Since $D(P\|P_0)$ is a continuous function of $P \in \overline{\Delta}^{(2)}$, there must exists a neighborhood $U \subset G_1(t) - E(t)$ of P such that for every $Q \in U \cap \Delta^{(2)}$,

$$D(Q\|P_0) < \min_{P \in E(t)} D(P\|P_0) \tag{53}$$

holds. Connect the two points P_0, Q by the geodesic $L_{P_0 Q}$ and denote by Q_s a point on $L_{P_0 Q}$. The parameter s is determined to satisfy $Q_0 = P_0$ and $Q_1 = Q$. From Lemma 4, we see that the divergence $D(Q_s\|P_0)$ is a strictly increasing function of $s > 0$, hence by (53), we have

$$D(Q_s\|P_0) < \min_{P \in E(t)} D(P\|P_0), \qquad 0 < s < 1. \tag{54}$$

By the assumption of the lemma and $Q \in G_1(t) - E(t)$, we have $\tau(P_0) < \tau(t)$ and $\tau(Q) > \tau(t)$, respectively. Therefore, by the intermediate value theorem, we see that there exists an s, $0 < s < 1$ that satisfies $\tau(Q_s) = \tau(t)$ or $Q_s \in E(t)$, but, which contradicts (54). □

IV. Direct Theorem

Let us denote by ϕ_n^* the most powerful test function under the constraint $\alpha(\phi_n) \leq \exp(-rn)$, and by $\rho_1(r)$ the power exponent, i.e.,

$$\rho_1(r) \triangleq \lim_{n \to \infty} -\frac{1}{n} \log \beta(\phi_n^*). \tag{55}$$

We have the following theorem.

Theorem 1: (see Theorem 2 in [9]): If $0 < r < D(P_1\|P_0)$,

$$\rho_1(r) = D(P_{t^*}\|P_1) \tag{56}$$

holds, where $P_{t^*} = (P_{t^*}(i, i'))_{i, i' \in \Omega}$ is a point on the geodesic connecting P_0 and P_1 that satisfies $D(P_{t^*}\|P_0) = r$, $t^* > 0$.

Proof: The critical condition of Neyman–Pearson's lemma is rewritten as

$$\tau(P_{\omega^n}) = -\frac{1}{n} \log \lambda_n + O\left(\frac{1}{n}\right). \tag{57}$$

First, we claim that

$$\lim_{n \to \infty} -\frac{1}{n} \log \lambda_n = \tau(t^*). \tag{58}$$

The assumption $0 < r < D(P_1\|P_0)$ implies $0 < t^* < 1$ by Lemma 4. Suppose $\tau(t^*) > \liminf_{n \to \infty} -\frac{1}{n} \log \lambda_n$. Then, there must exist t_1, t_2, $0 < t_1 < t_2 < t^*$, such that $\tau(t^*) > \tau(t_2) > \tau(t_1) > -\frac{1}{n} \log \lambda_n$ hold for infinitely many n's. Hence, by Neyman–Pearson's lemma, for sufficiently large n's, we have

$$\begin{aligned} \alpha(\phi_n^*) &\geq P_0\left\{\omega^n | \tau(P_{\omega^n}) > -\frac{1}{n} \log \lambda_n + O\left(\frac{1}{n}\right)\right\} \\ &\geq P_0\{\omega^n | \tau(P_{\omega^n}) \geq \tau(t_2)\} \\ &= P_0\{\omega^n | P_{\omega^n} \in G_1(t_2)\}. \end{aligned} \tag{59}$$

Notice that $t_2 > 0$ implies $P_0 \notin G_1(t_2)$. Thus, by Lemmas 2 and 7, for any ϵ with $0 < \epsilon < D(P_{t^*}\|P_0) - D(P_{t_2}\|P_0)$ and n's sufficiently large, we have by (59),

$$\begin{aligned} r &= -\frac{1}{n} \log \alpha(\phi_n^*) \\ &\leq -\frac{1}{n} \log P_0\{\omega^n | P_{\omega^n} \in G_1(t_2)\} \\ &\leq \min_{P \in G_1(t_2)} D(P\|P_0) + \epsilon \\ &= D(P_{t_2}\|P_0) + \epsilon \\ &< D(P_{t^*}\|P_0) \\ &= r, \end{aligned} \tag{60}$$

which is a contradiction. Hence, we have $\tau(t^*) \leq \lim|rminf_{n \to \infty} - \frac{1}{n} \log \lambda_n$. This argument is almost the same as the proof of Theorem 1 in [8], both of which are based on the monotonicity of functions σ_0, σ_1, and τ. So, similarly, we can show the inequality $\tau(t^*) \geq \limsup_{n \to \infty} -\frac{1}{n} \log \lambda_n$, and (56). □

V. Converse Theorem I

In the case $r > D(P_1\|P_0)$, we calculate the power exponent $\lim_{n \to \infty} -\frac{1}{n} \log(1 - \beta(\phi_n^*))$ instead of (55). We see that the condition $r > D(P_1\|P_0)$ is divided into two separate cases, i.e., $D(P_1\|P_0) < r < \sigma_0(\infty)$ and $r > \sigma_0(\infty)$. Although in the former case (and also in Theorem 1), the randomization of a test is not significant, in the latter case the randomization is essential to guarantee the finiteness of the power exponent.

Let us denote by $\rho_2(r)$ the power exponent in the case $D(P_1\|P_0) < r < \sigma_0(\infty)$, i.e.,

$$\rho_2(r) \triangleq \lim_{n \to \infty} -\frac{1}{n} \log(1 - \beta(\phi_n^*)). \tag{61}$$

Theorem 2: If $D(P_1\|P_0) < r < \sigma_0(\infty)$, we have

$$\rho_2(r) = D(P_{t^*}\|P_1), \tag{62}$$

where $P_{t^*} = (P_{t^*}(i, i'))_{i, i' \in \Omega}$ is a point on the geodesic connecting P_0 and P_1 that satisfies $D(P_{t^*}\|P_0) = r$, $t^* > 0$.

Proof: Let us consider the most powerful test function ϕ_n^* given in Neyman–Pearson's lemma. According to the monotonicity of σ_0, σ_1, and τ, we can show $\lim_{n \to \infty} -\frac{1}{n} \log \lambda_n = \tau(t^*)$ in the same way as the proof of Theorem 1. Furthermore, (62) is proved in a way similar to Theorem 2 in [8]. (It is evident here to see that (62) is obtained by the use of a nonrandomized test function.)

VI. Converse Theorem II

If $r > \sigma_0(\infty)$, as previously mentioned, the randomization of a test guarantees the finiteness of the power exponent. Here, we provide a slightly weak form of converse theorem. Let $\overline{\rho}_3(r)$ be the superior limit of the error exponent, i.e.,

$$\overline{\rho}_3(r) \triangleq \limsup_{n \to \infty} -\frac{1}{n} \log(1 - \beta(\phi_n^*)). \tag{63}$$

Theorem 3: If $r > \sigma_0(\infty)$, we have

$$\overline{\rho}_3(r) \leq r - \tau(\infty). \tag{64}$$

In order to prove the theorem, we first provide a lemma.

Lemma 8: For any $P \in \overline{\Delta}^{(2)}$, we have

$$\tau(-\infty) \leq \tau(P) \leq \tau(\infty). \tag{65}$$

Proof: For $P \in \Delta^{(2)}$, let us define a function $\xi(t)$ by

$$\xi(t) = D(P\|P_t), -\infty < t < \infty. \tag{66}$$

We have, by (28),

$$\frac{d\xi(t)}{dt} = \tau(t) - \tau(P) \tag{67}$$

and

$$\frac{d^2\xi(t)}{dt^2} = \frac{d^2}{dt^2}\log \zeta_t. \tag{68}$$

Hence, by Lemma 3, $\xi(t)$ is a convex function of t, $-\infty < t < \infty$. From (14) and (15), we see that some of $P_t(i'|i)$, $i, i' \in \Omega$, tend to zero as t goes to infinity or minus infinity. Thus, by the definition of divergence, we have

$$\lim_{t\to\infty}\xi(t) = \lim_{t\to-\infty}\xi(t) = \infty. \tag{69}$$

Therefore, from (69) and the convexity of $\xi(t)$, there exists a unique t with $(d\xi(t)/dt) = 0$, or, by (67), $\tau(t) = \tau(P)$. Consequently, from (35), we have (65) for $P \in \Delta^{(2)}$. By the continuity of $\tau(P)$, $P \in \overline{\Delta}^{(2)}$, (65) holds also for $P \in \overline{\Delta}^{(2)}$. □

Proof of Theorem 3: Let us define a test function $\tilde{\phi}_n$ by

$$\tilde{\phi}_n(\omega^n) = \begin{cases} 1, & \text{if } \tau(P_{\omega^n}) < \tau(\infty), \\ \tilde{\delta}, & \text{if } \tau(P_{\omega^n}) = \tau(\infty), \\ 0, & \text{if } \tau(P_{\omega^n}) > \tau(\infty), \end{cases} \tag{70}$$

with

$$\tilde{\delta} = 1 - \frac{\exp(-rn)}{P_0\{\omega^n|\tau(P_{\omega^n}) = \tau(\infty)\}}. \tag{71}$$

By Lemma 8, we see that the set $\{\omega^n|\tau(P_{\omega^n}) > \tau(\infty)\}$ is empty. Hence, we have $\alpha(\tilde{\phi}_n) = \exp(-rn)$. Since $-\frac{1}{n}\log P_0\{\omega^n|\tau(P_{\omega^n}) = \tau(\infty)\}$ converges to $\sigma_0(\infty)\ (< r)$ by Lemma 6, we have $0 < \tilde{\delta} < 1$ for sufficiently large n's. Therefore, for the most powerful test function ϕ_n^*, we have

$$\begin{aligned}\beta(\phi_n^*) &\leq \beta(\tilde{\phi}_n) \\ &= 1 - \exp(-rn)\frac{P_1\{\omega^n|\tau(P_{\omega^n}) = \tau(\infty)\}}{P_0\{\omega^n|\tau(P_{\omega^n}) = \tau(\infty)\}},\end{aligned}$$

which implies $\overline{\rho}_3(r) \leq r - \tau(\infty)$ by Lemmas 2 and 6.

Acknowledgment

The authors are grateful to the reviewers for suggestions on revising the original versions of this work and [8], and for several helpful comments.

References

[1] S. Amari, *Differential-Geometrical Methods in Statistics* (Lecture Notes in Statistics). New York: Springer-Verlag, 1985.

[2] S. Amari and T. S. Han, "Statistical inference under multiterminal rate restriction: A differential geometric approach," *IEEE Trans. Inform. Theory*, vol. 35, pp. 217–227, Mar. 1989.

[3] R. Ash, *Information Theory*. New York: John Wiley, 1965.

[4] I. Csiszár, T. M. Cover, and B.-S. Choi, "Conditional limit theorems and Markov conditioning," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 788–801, Nov. 1987.

[5] T. S. Han and K. Kobayashi, "The strong converse theorem for hypothesis testing," *IEEE Trans. Inform. Theory*, vol. 35, pp. 178–180, Jan. 1989.

[6] H. Ito and S. Amari, "The geometry of information sources," in *Proc. 11th SITA*, 1988, pp. 57–60, (in Japanese).

[7] E. L. Lehmann, *Testing Statistical Hypotheses*. New York: Wiley, 1959.

[8] K. Nakagawa and F. Kanaya, "On the converse theorem in statistical hypothesis testing," *IEEE Trans. Inform. Theory*, vol. 39, pp. 623–628, Mar. 1993.

[9] S. Natarajan, "Large deviations, hypothesis testing, and source coding for finite Markov chains," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 360–365, May 1985.

[10] K. Vašek, "On the error exponent for ergodic Markov sources," *Kybernetica*, vol. 16, pp. 318–329, 1980.

A Note on the Asymptotics of Distributed Detection with Feedback

Hossam M. H. Shalaby, *Member, IEEE,* and
Adrian Papamarcou, *Member, IEEE*

***Abstract*—The effect of feedback on the performance of a distributed Neyman–Pearson detection system consisting of n sensors, two-stage local quantizers, and a global detector is investigated. Two feedback schemes are discussed, only one of which yields an asymptotic gain in performance, as measured by an appropriate error exponent.**

***Index Terms*— Distributed detection, error exponents, feedback, hypothesis testing, Kullback–Leibler divergence, quantization.**

I. Introduction

A distributed (or decentralized) detection system is a network of n sensors which, together with a global detector (or data fusion center), cooperatively undertake the task of identifying a random signal source. Typically, the sensors compress their observations into low-rate data streams, which are then transmitted to the global detector for processing and decision making. Compared to a conventional detection system with passive sensors, the distributed setup offers the advantages of reduced communication bandwidth, shared processing and increased reliability, albeit at some expense of performance.

Most distributed detection models employed in the literature are feedforward systems, in which the information flow from the sensors to the global detector is unidirectional. Recently, there has been some interest [1], [2], [9] in models with bidirectional flow, in which feedback from the global detector to the sensors is allowed. In such systems, the process of local data compression and transmission

Manuscript received March 15, 1992; revised July 8, 1992. This work was supported by the Institute for Systems Research (a National Science Foundation Engineering Research Center) at the University of Maryland and by the Minta Martin Fund for Aeronautical Research administered by the College of Engineering, University of Maryland. This work was presented in part at the 1992 IEEE Symposium on Information Theory, Budapest, Hungary, June 24–28, 1991.

H. M. H. Shalaby was with the Electrical Engineering Department and the Institute for Systems Research, University of Maryland, College Park, MD 20742. He is now with the Department of Electrical Engineering, Faculty of Engineering, University of Alexandria, Alexandria 21544, Egypt.

A. Papamarcou is with the Electrical Engineering Department and the Institute for Systems Research, University of Maryland, College Park, MD 20742.

IEEE Log Number 9204211.