p do not contribute to the partial derivative. Standard differentiation then results in (for $1 \le j \le m_i - 1$)

$$\frac{\partial p}{\partial a_{ij}} = \sum_{j_n=1}^{m_n} \cdots \Big(\sum_{j_i=1}^{m_i}\Big)^* \cdots \sum_{j_1=1}^{m_1} \prod_{\substack{k=1\\k\neq i}}^{n} (a_{kj_k} - a_{k(j_k-1)}) \cdot \big[Q(\theta - g(x_{1j_1}, x_{2j_2}, \cdots, x_{ij_i}^{**}, \cdots, x_{nj_n})) - Q(\theta - g(x_{1j_1}, x_{2j_2}, \cdots, x_{ij_i}^{***}, \cdots, x_{nj_n}))\big],$$

where $*$ denotes deletion, $**$ denotes replacing x_{ij_i} with x_{ij}, and $***$ denotes replacing x_{ij_i} with $x_{i(j+1)}$.

By substituting the previous expression for the partial derivatives into (A.2), we may write ϕ as

$$\phi = \lim_{\max_i |\mathcal{P}_i| \to 0} \inf_{\mathcal{P}_i} \left[\frac{1}{1 + \sum_{i=1}^{n} \eta_i}\right],$$

where η_i is given by

$$\eta_i = \sum_{j=1}^{m_i} \Big| \sum_{j_n=1}^{m_n} \cdots \Big(\sum_{j_i=1}^{m_i}\Big)^* \cdots \sum_{j_1=1}^{m_1} \prod_{\substack{k=1\\k\neq i}}^{n} (a_{kj_k} - a_{k(j_k-1)}) \cdot \big[Q(\theta - g(x_{1j_1}, x_{2j_2}, \cdots, x_{ij_i}^{***}, \cdots, x_{nj_n})) - Q(\theta - g(x_{1j_1}, x_{2j_2}, \cdots, x_{ij_i}^{**}, \cdots, x_{nj_n}))\big] \Big|.$$

Rearranging η_i, we obtain

$$\eta_i = \sum_{j=1}^{m_i} \Big| \frac{1}{x_{i(j+1)} - x_{ij}} \Big\{ \sum_{j_n=1}^{m_n} \cdots \Big(\sum_{j_i=1}^{m_i}\Big)^* \cdots \sum_{j_1=1}^{m_1} \cdot \prod_{\substack{k=1\\k\neq i}}^{n} (a_{kj_k} - a_{k(j_k-1)}) Q(\theta - g(x_{1j_1}, x_{2j_2}, \cdots, x_{ij_i}^{***}, \cdots, x_{nj_n})) - \sum_{j_n=1}^{m_n} \cdots \Big(\sum_{j_i=1}^{m_i}\Big)^* \cdots \sum_{j_1=1}^{m_1} \cdot \prod_{\substack{k=1\\k\neq i}}^{n} (a_{kj_k} - a_{k(j_k-1)}) Q(\theta - g(x_{1j_1}, x_{2j_2}, \cdots, x_{ij_i}^{**}, \cdots, x_{nj_n})) \Big\} \Big| (x_{i(j+1)} - x_{ij}).$$

Now letting

$$\Delta = \lim_{\max_i |\mathcal{P}_i| \to 0} \inf_{\mathcal{P}_i} \eta_i,$$

we have that

$$\Delta = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \Big| \frac{\partial}{\partial x_i} \int_{R^{n-1}} Q(\theta - g(x_1, \cdots, x_n)) \cdot dF_{X_1}(x_1) \cdots dF_{X_i}^*(x_i) \cdots dF_{X_n}(x_n) \Big| dx_i,$$

when the integrals exist, and where once again $*$ denotes deletion. To arrive at this expression we note that the two inner summation terms which are enclosed by the absolute value symbols each corresponds to an iterated Stieltjes integral (when it exists) of the function $Q(\theta - g(\cdot, \cdot, \cdots, \cdot))$. The derivative term that appears immediately before the inner integral arises due to the $(x_{i(j+1)} - x_i)$ difference term that appears in the denominator of the η_i expression, coupled with the fact that the first inner summation of η_i is evaluated at $x_{i(j+1)}$ while the second inner summation term of η_i is evaluated x_{ij}. The outer summation term of η_i is of the form of a Riemann sum and hence results in the outer integral in expression for Δ.

References

[1] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, pp. 433–480, Mar. 1985.

[2] P. J. Huber, "Robust estimation of a local parameter," *Ann. Math. Stat.*, vol. 35, pp. 73–101, March 1964.

[3] ——, "A robust version of the probability ratio test," *Ann. Math. Statist.*, vol. 36, pp. 1753–1758, Dec. 1965.

[4] ——, *Robust Statistics*. New York: Wiley, 1981.

[5] E. B. Hall and G. L. Wise, "Simultaneous optimal estimation over a family of fidelity criteria," in *Proc. 1987 Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 25–27, 1987, pp. 65–70.

[6] D. R. Halverson, "Robust estimation and signal detection in the presence of dependent data," in *Proc. 1989 Conf. Inform. Sci. Syst.* , Baltimore, MD, Mar. 22–24, 1989, pp. 283–288.

[7] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* Reading, MA: Addison-Wesley, 1968.

[8] M. W. Thompson and D. R. Halverson, "A novel differential geometric approach toward robust signal detection," in *Proc. 1986 Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 19–21, 1986, pp. 11–16.

[9] M. W. Thompson and D. R. Halverson, "A differential geometric approach toward robust signal detection in nonstationary and/or dependent noise," in *Proc. Twenty-Fifth Ann. Allerton Conf. Commun., Contr., and Computing*, Monticello, IL, Sept. 30–Oct. 2, 1987, pp. 163–170.

[10] M. W. Thompson, "Application of a robustness measure for parameter estimation," in *Proc. 32nd Midwest Symp. Circuits Syst.*, Champaign, IL, Aug. 14–16, 1989, pp. 1154–1157.

[11] M. W. Thompson and D. R. Halverson, "A nonstandard family of signal detectors with enhanced robustness," in *Proc. 1989 Conf.Inform. Sci. Syst.*, Baltimore, MD, Mar. 22–24, 1989, pp. 277–281.

On the Converse Theorem in Statistical Hypothesis Testing

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

***Abstract*—Simple statistical hypothesis testing is investigated by making use of divergence geometric method. The asymptotic behavior of the minimum value of the second-kind error probability under the constraint that the first-kind error probability is bounded above by exp($-rn$) is looked for, where r is a given positive number. If r is greater than the divergence of the two probability measures, the so-called converse theorem holds. It is shown that the condition under which the converse theorem holds can be divided into two separate cases by analyzing the geodesic connecting the two probability measures and, as a result, a lucid explanation is given for Han-Kobayashi's linear function $f_r(\bar{X})$.**

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

I. Introduction

In this investigation of simple statistical hypothesis testing, let Ω be a set consisting of $m+1$ natural numbers, i.e., $\Omega = \{0, 1, \cdots, m\}$, $m \ge 1$. Define two sets $\overline{\Delta}$ and Δ of probability measures on Ω as

$$\overline{\Delta} \triangleq \left\{ p = (p(i))_{i\in\Omega} \Big| \sum_{i\in\Omega} p(i) = 1,\ p(i) \ge 0,\ i \in \Omega \right\}. \quad (1)$$

Manuscript received March 6, 1990; revised June 1, 1992. This work was presented 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 9203869.

$$\Delta \triangleq \{p \in \overline{\Delta} | p(i) > 0, \ i \in \Omega\}. \tag{2}$$

For two probability measures $p, q \in \overline{\Delta}$, we say that p is dominated by q and denote it by $p \ll q$ if $q(i) = 0$ always implies $p(i) = 0$, $i \in \Omega$.

Let us consider hypothesis testing for two probability measures $p_0, p_1 \in \Delta$. Let H_0 be the null hypothesis that p_0 gives the probability of an i.i.d. n-sequence $\omega^n \in \Omega^n$, where Ω^n is the Cartesian n-product of Ω, and let H_1 be the alternative hypothesis that p_1 gives the probability. A randomized test function of hypothesis testing is defined as a mapping from Ω^n to the closed unit interval $[0, 1] = \{x | 0 \le x \le 1\}$. We assume in this correspondence that the term "test function" always indicates randomized test function.

For a test function $\phi_n : \Omega^n \to [0, 1]$, the first-kind error probability $\alpha(\phi_n)$ and the second-kind error probability $\beta(\phi_n)$ are defined by

$$\alpha(\phi_n) \triangleq \sum_{\omega^n \in \Omega^n} (1 - \phi_n(\omega^n)) p_0(\omega^n), \tag{3}$$

$$\beta(\phi_n) \triangleq \sum_{\omega^n \in \Omega^n} \phi_n(\omega^n) p_1(\omega^n), \tag{4}$$

where $p_0(\omega^n) \triangleq p_0(\omega_1) \cdots p_0(\omega_n)$ for $\omega^n = \omega_1 \cdots \omega_n \in \Omega^n$, $\omega_j \in \Omega$, $j = 1, \cdots, n$.

Under the constraint $\alpha(\phi_n) \le \alpha$, we say that ϕ_n^* is the most powerful test function if ϕ_n^* satisfies $\alpha(\phi_n^*) \le \alpha$ and $\beta(\phi_n^*) \le \beta(\phi_n)$ holds for any ϕ_n with $\alpha(\phi_n) \le \alpha$. Our main problems 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 the Kullback–Leibler divergence. The above limiting values are called the power exponents of hypothesis testing.

The power exponent for a nonrandomized test under the constraint that the first-kind error probability is bounded above by a positive constant is calculated by Stein's lemma [4], which asserts

$$\lim_{n\to\infty} -\frac{1}{n} \log \beta(\phi_n^*) = D(p_0 \| p_1). \tag{5}$$

Blahut [3] and Han–Kobayashi [5] investigated the case where $\alpha(\phi_n) \le \exp(-rn)$ and obtained the following results. Both of them considered nonrandomized tests.

Theorem (Blahut [3]): If $r < D(p_1 \| p_0)$,

$$\lim_{n\to\infty} -\frac{1}{n} \log \beta(\phi_n^*) = D(p_t \| p_1), \tag{6}$$

where $p_t(i) = C_t \{p_0(i)\}^{1-t} \{p_1(i)\}^t$, $i \in \Omega$, $C_t^{-1} = \sum_{i\in\Omega} \{p_0(i)\}^{1-t}\{p_1(i)\}^t$, and $t \in \boldsymbol{R}$ is determined by the equation $D(p_t \| p_0) = r$, $t > 0$.

If $r > D(p_1 \| p_0) (\triangleq c)$,

$$\beta(\phi_n^*) > 1 - \frac{4\sigma^2}{n(r-c)^2} - e^{-n(r-c)/2}, \tag{7}$$

where σ^2 is the variance of a random variable $\log(p_1(i)/p_0(i))$ with respect to p_1.

Han–Kobayashi [5], with reference to Blahut's results, defined an r-divergent sequence, which is an extension of typical sequence, and then approached this problem using an information-theoretic method.

Theorem (Han–Kobayashi's Strong Converse Theorem [5]): If $r > D(p_1 \| p_0)$,

$$\limsup_{n\to\infty} -\frac{1}{n} \log (1 - \beta(\phi_n^*)) = \min_{D(p\|p_0)\le r} f_r(p), \tag{8}$$

where $f_r(p) = r + D(p \| p_1) - D(p \| p_0)$.

As we can see from these theorems, if $r < D(p_1 \| p_0)$, $\beta(\phi_n^*) \to 0$, while if $r > D(p_1 \| p_0)$, $\beta(\phi_n^*) \to 1$. The theorem for the latter case is called the converse theorem. In this work, we study these problems according to Neyman–Pearson's lemma and analyze the geodesic in the space Δ that connects the two probability measures p_0 and p_1. We show that the condition $r > D(p_1 \| p_0)$ is divided into two separate cases and a lucid explanation is given for the introduction of Han–Kobayashi's linear function f_r [5]. We also mention general upper bounds of $\alpha(\phi_n)$.

II. Preliminaries

We present some geometric preliminaries which are used to calculate power exponents of the hypothesis testing for $p_0, p_1 \in \Delta$. The most powerful test function under the constraint $\alpha(\phi_n) \le \alpha$ is completely determined by Neyman–Pearson's lemma.

Lemma 1 (Neyman–Pearson, see [6]): For any α, $0 < \alpha < 1$, there exists a test function ϕ_n^*, with a constant λ_n, which satisfies $\alpha(\phi_n^*) = \alpha$ and

$$\phi_n^*(\omega^n) = \begin{cases} 1, & \text{if } p_0(\omega^n) > \lambda_n p_1(\omega^n), \\ 0, & \text{if } p_0(\omega^n) < \lambda_n p_1(\omega^n). \end{cases} \tag{9}$$

This is the most powerful test function. Write $E = \{\omega^n | p_0(\omega^n) = \lambda_n p_1(\omega^n)\}$. If $p_0(E) \ne 0$, we have $\phi_n^*(\omega^n) = \delta$ for $\omega^n \in E$, where

$$\delta = 1 - \frac{\alpha - p_0\{\omega^n | p_0(\omega^n) < \lambda_n p_1(\omega^n)\}}{p_0(E)}. \tag{10}$$

For an n-sequence $\omega^n \in \Omega^n$, denote by $N(i|\omega^n)$ the number of $i \in \Omega$ appearing in ω^n. We have $N(i|\omega^n) \ge 0$, $\frac{1}{n}\sum_{i\in\Omega} N(i|\omega^n) = 1$. Then, putting $p_{\omega^n} \triangleq (\frac{1}{n}N(i|\omega^n))_{i\in\Omega}$, we have $p_{\omega^n} \in \overline{\Delta}$. The probability measure p_{ω^n} is called the type of ω^n. The following lemma is applicable to our problem of calculating power exponents.

Lemma 2 (Sanov, see [4]): Let A be a subset of $\overline{\Delta}$ and A_n be the set of types of n-sequences in A. Then for any $p \in \overline{\Delta}$ the following inequality holds:

$$\left| \frac{1}{n} \log p\{\omega^n | p_{\omega^n} \in A\} + \min_{q\in A_n} D(q \| p) \right| \le \frac{\log(n+1)}{n}(m+1). \tag{11}$$

If A is closed and $\bigcup_{n\ge 1} A_n$ is dense in A (with respect to the usual topology of $\boldsymbol{R}^{m+1}$), we obtain

$$\lim_{n\to\infty} -\frac{1}{n} \log p\{\omega^n | p_{\omega^n} \in A\} = \min_{q\in A} D(q \| p). \tag{12}$$

According to Neyman–Pearson's lemma, if two sequences $\omega^n, \tilde\omega^n \in \Omega^n$ have the same probability (with respect to p_0 and p_1), $\phi_n^*(\omega^n) = \phi_n^*(\tilde\omega^n)$ holds. Thus, it is sufficient to look at only the type of a sequence rather than each sequence itself. So, it is more significant to consider the geometry of $\overline{\Delta}$ rather than that of Ω^n. The general theory of information geometry [1], [7] shows that differential geometric structures are defined on the space Δ, such as Riemannian metric, α-connections, and dual coordinate systems, which have many beautiful properties. Particularly, the $+1$ affine coordinates θ^i, $i = 1, \cdots, m$, at $p \in \Delta$ is given by

$$\theta^i = \log \frac{p(i)}{p(0)}, \qquad i = 1, \cdots, m, \tag{13}$$

and the equation of the geodesic L connecting p_0 and p_1 is given by

$$\theta_t^i = (1-t)\theta_0^i + t\theta_1^i, \qquad i = 1, \cdots, m, \qquad t \in \boldsymbol{R}, \tag{14}$$

where, θ_0^i, θ_1^i, and θ_t^i are ith $+1$ affine coordinates of p_0, p_1, and p_t (a point on the geodesic), respectively. Then substituting (13) into (14), we have the following alternative form of the geodesic equation:

$$p_t(i) = C_t \{p_0(i)\}^{1-t} \{p_1(i)\}^t, \qquad i \in \Omega, \qquad t \in \boldsymbol{R}, \tag{15}$$

where C_t is the normalizing constant, i.e., $C_t^{-1} = \sum_{i\in\Omega}\{p_0(i)\}^{1-t} \cdot\{p_1(i)\}^t$.

Readers not familiar with information geometric notions may skip the previous manner of introducing the geodesic equation because an information geometric back ground is not required in the following discussion. They may regard (15) as a definition of the geodesic.

The Kullback-Leibler divergence, or simply, the divergence $D(p\|q)$ of two probability measures $p, q \in \overline{\Delta}$ is defined by

$$D(p\|q) \triangleq \sum_{i\in\Omega} p(i) \log \frac{p(i)}{q(i)}. \tag{16}$$

We have $0 \le D(p\|q) \le \infty$, $p, q \in \overline{\Delta}$, $D(p\|q) = 0$, if and only if $p = q$ and $D(p\|q) < \infty$, if and only if $p \ll q$ (see [4]).

III. Limiting Point of Geodesic

We study the limiting points of the geodesic L that connects two points p_0 and p_1 in Δ. Denote $M^+ = \max_{i\in\Omega}\{(p_1(i)/p_0(i))\} > 1$ and $\Omega^+ = \{i \in \Omega | (p_1(i)/p_0(i)) = M^+\}$. Since (15) of the geodesic L is rewritten as

$$p_t(i) = \frac{p_0(i)\left\{\frac{p_1(i)}{p_0(i)}\right\}^t}{\sum_{j\in\Omega} p_0(j)\left\{\frac{p_1(j)}{p_0(j)}\right\}^t}, \qquad i \in \Omega, \qquad t \in \boldsymbol{R}, \tag{17}$$

we have

$$\lim_{t\to\infty} p_t(i) = \begin{cases} \frac{p_0(i)}{\sum_{j\in\Omega^+} p_0(j)}, & \text{if } i \in \Omega^+, \\ 0, & \text{otherwise.}\end{cases} \tag{18}$$

We write $p_\infty = (p_\infty(i))_{i\in\Omega}$, where $p_\infty(i) = \lim_{t\to\infty} p_t(i)$ and call it *the limiting point* of L.

Let E_∞ be the boundary simplex of $\overline{\Delta}$ that includes p_∞ in its relative interior (with respect to the induced topology from $\boldsymbol{R}^{m+1}$), i.e.,

$$E_\infty = \{p \in \overline{\Delta} | p(i) = 0, \ i \in \Omega - \Omega^+\}. \tag{19}$$

In a similar way, we define $p_{-\infty} \in \overline{\Delta}$, i.e., let $M^- = \min_{i\in\Omega}\{(p_1(i)/p_0(i))\}$ and $\Omega^- = \{i \in \Omega | (p_1(i)/p_0(i)) = M^-\}$. Then, we have

$$\lim_{t\to-\infty} p_t(i) = \begin{cases} \frac{p_0(i)}{\sum_{j\in\Omega^-} p_0(j)}, & \text{if } i \in \Omega^-, \\ 0, & \text{otherwise.}\end{cases} \tag{20}$$

Write $p_{-\infty} = (p_{-\infty}(i))_{i\in\Omega}$, where $p_{-\infty}(i) = \lim_{t\to-\infty} p_t(i)$, and define

$$E_{-\infty} = \{p \in \overline{\Delta} | p(i) = 0, \ i \in \Omega - \Omega^-\}. \tag{21}$$

We notice here that E_∞ and $E_{-\infty}$ are disjoint sets, or,

$$E_\infty \cap E_{-\infty} = \emptyset \tag{22}$$

because $\Omega^+ \cap \Omega^- = \emptyset$ holds. Thus, putting $\overline{\Delta}_f \triangleq \overline{\Delta} - (E_\infty \cup E_{-\infty})$, we have a decomposition of $\overline{\Delta}$;

$$\overline{\Delta} = \overline{\Delta}_f \cup E_\infty \cup E_{-\infty}. \tag{23}$$

We find that E_∞ and $E_{-\infty}$ are parallel (in the Euclidean sense) to sets of the following form:

$$\left\{p \in \overline{\Delta} \Big| \sum_{i\in\Omega} p(i) \log \frac{p_1(i)}{p_0(i)} = \text{constant}\right\}. \tag{24}$$

This is because

$$\begin{aligned}\sum_{i\in\Omega} p(i) \log \frac{p_1(i)}{p_0(i)} &= \sum_{i\in\Omega^+ (\text{or, } i\in\Omega^-)} p(i) \log \frac{p_1(i)}{p_0(i)} \\ &= \log M^+ (\text{or, } \log M^-)\end{aligned} \tag{25}$$

holds for arbitrary $p \in E_\infty$ (or, $p \in E_{-\infty}$), respectively.

IV. Lemmas

First, we provide a lemma which is used in the proof of Lemma 4.

Lemma 3: For $p \in \overline{\Delta}_f$, write $\xi(t) = D(p\|p_t)$, $-\infty < t < \infty$. Then there exists a unique p_t that satisfies

$$\frac{d\xi(t)}{dt} = 0. \tag{26}$$

Proof: Since $p \in \overline{\Delta}_f$, p is not dominated by p_∞ nor $p_{-\infty}$. Thus, we have

$$\lim_{t\to\infty} \xi(t) = \lim_{t\to-\infty} \xi(t) = \infty. \tag{27}$$

From (15), we have

$$\frac{d \log p_t(i)}{dt} = \frac{d \log C_t}{dt} + \log \frac{p_1(i)}{p_0(i)}, \qquad i \in \Omega. \tag{28}$$

Since $\sum_{i\in\Omega}(dp_t(i)/dt) = (d/dt)\sum_{i\in\Omega} p_t(i) = 0$, from (28), we have

$$\frac{d \log C_t}{dt} = -\sum_{i\in\Omega} p_t(i) \log \frac{p_1(i)}{p_0(i)}. \tag{29}$$

Then, from (28) and (29) we have

$$\begin{aligned}\frac{d^2\xi(t)}{dt^2} &= \sum_{i\in\Omega} \frac{dp_t(i)}{dt} \log \frac{p_1(i)}{p_0(i)} \\ &= \sum_{i\in\Omega} p_t(i)\left\{\log \frac{p_1(i)}{p_0(i)}\right\}^2 \\ &\quad - \left\{\sum_{i\in\Omega} p_t(i) \log \frac{p_1(i)}{p_0(i)}\right\}^2 \\ &= \text{variance of } \log\frac{p_1(i)}{p_0(i)} \text{ w.r.t. } p_t \\ &> 0.\end{aligned} \tag{30}$$

Consequently, by (27) and (30), the existence and the uniqueness of p_t with (26) are guaranteed. □

We give the following lemma for later investigation of the critical condition in Neyman-Pearson's lemma.

Lemma 4: For any $p \in \overline{\Delta}$, there exists a unique p_t, $-\infty \le t \le \infty$, that satisfies both

$$D(p\|p_0) = D(p\|p_t) + D(p_t\|p_0). \tag{31}$$

and

$$D(p\|p_1) = D(p\|p_t) + D(p_t\|p_1). \tag{32}$$

Proof: Remember the decomposition (23) of $\overline{\Delta}$. First, we see that for $p \in E_\infty$, p_∞ is the unique p_t that satisfies (31) and (32). For, since $p(i) = 0$ and $p_\infty(i) = 0$ hold for $i \in \Omega - \Omega^+$, from (18) we have

$$\begin{aligned}&D(p\|p_0) - D(p\|p_\infty) - D(p_\infty\|p_0) \\ &\quad = -\sum_{i\in\Omega^+}\{p_\infty(i) - p(i)\} \log \frac{p_\infty(i)}{p_0(i)} \\ &\quad = \log\left(\sum_{i\in\Omega^+} p_0(i)\right) \sum_{i\in\Omega^+}\{p_\infty(i) - p(i)\} \\ &\quad = 0.\end{aligned}$$

In a similar way, we also have (32). Here, suppose there exist another p_t, $-\infty \le t < \infty$, which satisfies both (31) and (32). Then a simple calculation leads to

$$D(p\|p_s) = D(p\|p_t) + D(p_t\|p_s) \tag{33}$$

holding for any s, $-\infty < s < \infty$. Since $p \in E_\infty$, i.e., $p \ll p_\infty$, the left-hand side of (33) converges to $D(p\|p_\infty) < \infty$ as s tends to infinity, while the right-hand side goes to infinity because $p_t \not\ll_\infty$. This is a contradiction. Therefore, for $p \in E_\infty$, the existence and the uniqueness of p_t with (31) and (32) are proved. In exactly the same way, it is proved that for $p \in E_{-\infty}$, $p_{-\infty}$ is the unique p_t with (31) and (32).

Next, for $p \in \overline{\Delta}_f$, we show that there exists a unique p_t, $-\infty < t < \infty$, which satisfies (31) and (32). From Lemma 3, for $p \in \overline{\Delta}_f$ there uniquely exists a p_t, $-\infty < t < \infty$, with $(d\xi(t)/dt) = 0$, where $\xi(t) = D(p\|p_t)$. Since

$$\frac{d\xi(t)}{dt} = \sum_{i\in\Omega} p_t(i) \log \frac{p_1(i)}{p_0(i)} - \sum_{i\in\Omega} p(i) \log \frac{p_1(i)}{p_0(i)}, \tag{34}$$

by calculation we have

$$D(p\|p_0) - D(p\|p_t) - D(p_t\|p_0) = -t\frac{d\xi(t)}{dt}. \tag{35}$$

Thus, from (35), (31) is equivalent to $(d\xi(t)/dt) = 0$ or $t = 0$. In the same way, we have

$$D(p\|p_1) - D(p\|p_t) - D(p_t\|p_1) = (t-1)\frac{d\xi(t)}{dt}. \tag{36}$$

Then, (32) is equivalent to $d\xi(t)/dt = 0$, or $t = 1$. Consequently, the t that satisfies both (31) and (32) is given by the root of $d\xi(t)/dt = 0$ whose existence and uniqueness are guaranteed by Lemma 3. This completes the proof of the lemma. □

Next, we present a lemma on the growth of divergence along the geodesic.

Lemma 5: Let p_t be a point on the geodesic defined by (15). Write $\sigma_0(t) \triangleq D(p_t\|p_0)$, $\sigma_1(t) \triangleq D(p_t\|p_1)$, and $\tau(t) \triangleq D(p_t\|p_0) - D(p_t\|p_1)$, $-\infty \le t \le \infty$. We have the following.

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, $-\infty < t < \infty$.

d) σ_0, σ_1, τ are bounded functions. Especially,

$$\tau(-\infty) \le \tau(t) \le \tau(\infty). \tag{37}$$

Proof: From (28) and (29), we have

$$\begin{aligned}\frac{d\tau(t)}{dt} &= \sum_{i\in\Omega} p_t(i)\left\{\log\frac{p_1(i)}{p_0(i)}\right\}^2 \\ &\quad - \left\{\sum_{i\in\Omega} p_t(i)\log\frac{p_1(i)}{p_0(i)}\right\}^2 \\ &= \text{variance of } \log\frac{p_1(i)}{p_0(i)} \text{w.r.t. } p_t \\ &> 0,\end{aligned}$$

and $(d\sigma_0(t)/dt) = t(d\tau(t)/dt)$, $(d\sigma_1(t)/dt) = (1-t)(d\tau(t)/dt)$, which show a), b), and c). Part d) is readily seen.

The critical condition $p_0(\omega^n) = \lambda_n p_1(\omega^n)$ in Neyman–Pearson's lemma is rewritten as

$$\sum_{i\in\Omega} p_{\omega^n}(i) \log \frac{p_1(i)}{p_0(i)} = -\frac{1}{n}\log \lambda_n. \tag{38}$$

So, it is important to consider such subsets of $\overline{\Delta}$ as $\{p \in \overline{\Delta} | \sum_{i\in\Omega} p(i)\log(p_1(i)/p_0(i)) = \tau\}$, where τ is a constant.

Write $\overline{L} \triangleq L \cup \{p_\infty\} \cup \{p_{-\infty}\}$. According to Lemma 4, for any $p \in \overline{\Delta}$, there uniquely exists a p_t on $\overline{L}$ with (31) and (32). We call this p_t the projection of p onto $\overline{L}$. By subtracting (32) from (31), we have

$$D(p\|p_0) - D(p\|p_1) = D(p_t\|p_0) - D(p_t\|p_1) = \tau(t) \tag{39}$$

or

$$\sum_{i\in\Omega} p(i) \log \frac{p_1(i)}{p_0(i)} = \sum_{i\in\Omega} p_t(i) \log \frac{p_1(i)}{p_0(i)} = \tau(t). \tag{40}$$

Thus, we find that if p and $q \in \overline{\Delta}$ have the same projection p_t onto $\overline{L}$, we have

$$\sum_{i\in\Omega} p(i) \log \frac{p_1(i)}{p_0(i)} = \sum_{i\in\Omega} q(i) \log \frac{p_1(i)}{p_0(i)} = \tau(t). \tag{41}$$

Here, we say that p and $q \in \overline{\Delta}$ are equivalent if p and q have the same projection, say p_t, onto $\overline{L}$. From (41), it is seen that the mapping $p \mapsto \tau(t)$ is well defined on the equivalence classes. Therefore, for $p \in \overline{\Delta}$, we can define

$$\tau(p) = \sum_{i\in\Omega} p(i) \log \frac{p_1(i)}{p_0(i)}. \tag{42}$$

If $p = p_t \in L$, $\tau(p_t) = \tau(t)$. Then, the critical condition (38) in Neyman–Pearson's lemma is expressed as

$$\tau(p_{\omega^n}) = -\frac{1}{n}\log \lambda_n. \tag{43}$$

We provide the following lemma which is used for evaluating the first- and second-error kind probabilities.

Lemma 6: Let us write $E(\tau) \triangleq \{p \in \overline{\Delta} | \tau(p) = \tau\}$. If $E(\tau)$ is not empty, there exists a unique p_t such that $E(\tau) \cap \overline{L} = \{p_t\}$. For this p_t, we have

$$\min_{p\in E(\tau)} D(p\|p_0) = D(p_t\|p_0) \tag{44}$$

and

$$\min_{p\in E(\tau)} D(p\|p_1) = D(p_t\|p_1). \tag{45}$$

Proof: From Lemma 4, for $p \in E(\tau)$, there exists a p_t with (31), (32). Since $\tau(p_t) = \tau(p) = \tau$, we have $p_t \in E(\tau) \cap \overline{L}$. The uniqueness of p_t is obtained by Lemma 5c). (44) and (45) are alternative expressions of (31) and (32). □

Remark: From Lemma 4, we find that $E_\infty = E(\tau(\infty))$ and $E_{-\infty} = E(\tau(-\infty))$.

Lemma 7: Let us write $G_0(\tau) \triangleq \{p \in \overline{\Delta} | \tau(p) \le \tau\}$ and $G_1(\tau) \triangleq \{p \in \overline{\Delta} | \tau(p) \ge \tau\}$. If $p_0 \notin G_1(\tau)$, we have

$$\min_{p\in G_1(\tau)} D(p\|p_0) = \min_{p\in E(\tau)} D(p\|p_0). \tag{46}$$

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

$$\min_{p\in G_j(\tau)} D(p\|p_k) = \min_{p\in E(\tau)} D(p\|p_k). \tag{47}$$

Proof: We prove only (46) because (47) is proved in the same way. First, $E(\tau) \subset G_1(\tau)$ implies $\min_{p\in G_1(\tau)} D(p\|p_0) \le \min_{p\in E(\tau)} D(p\|p_0)$. We show the opposite inequality. For any $p' \in G_1(\tau)$, we denote by $p_{t'}$ the projection of p' onto $\overline{L}$. By (31), we have

$$D(p'\|p_0) \ge D(p_{t'}\|p_0). \tag{48}$$

Let p_t denote the unique point in $E(\tau) \cap \overline{L}$, then p_t attains $\min_{p\in E(\tau)} D(p\|p_0)$. Since $p_0 \notin G_1(\tau)$, $p_t \in E(\tau)$, and $p_{t'} \in G(\tau)$, we have

$$\tau(p_0) < \tau(p_t) \le \tau(p_{t'}). \tag{49}$$

Then, by the monotonicity of $\tau(t)$ and $\sigma_0(t)$ in Lemma 5, we obtain $0 < t \le t'$ and hence,

$$0 < D(p_t\|p_0) \le D(p_{t'}\|p_0). \tag{50}$$

Consequently, from (48) and (50), we have $D(p'\|p_0) \ge D(p_t\|p_0)$, which implies $\min_{p\in G_1(\tau)} D(p\|p_0) \ge \min_{p\in E(\tau)} D(p\|p_0)$.

V. Direct Theorem

Denote by ϕ_n^* the most powerful test function under the constraint $\alpha(\phi_n) \le \exp(-rn)$. Let $\rho_1(r)$ be the power exponent, i.e., $\rho_1(r) \triangleq \lim_{n\to\infty} -(1/n)\log \beta(\phi_n^*)$. We present the following theorem.

Theorem 1: If $0 < r < D(p_1\|p_0)$, we have

$$\rho_1(r) = D(p_{t^*}\|p_1). \tag{51}$$

where $p_{t^*} = (p_{t^*}(i))_{i\in\Omega}$ is given by $p_{t^*}(i) = C_{t^*}\{p_0(i)\}^{1-t^*}\{p_1(i)\}^{t^*}$, $i \in \Omega$, $C_{t^*}^{-1} = \sum_{i\in\Omega}\{p_0(i)\}^{1-t^*}\{p_1(i)\}^{t^*}$, and $D(p_{t^*}\|p_0) = r$, $t^* > 0$.

Proof: According to Neyman–Pearson's lemma, the most powerful test function ϕ_n^* is

$$\phi_n^*(\omega^n) = \begin{cases} 1. & \text{if } p_0(\omega^n) > \lambda_n p_1(\omega^n), \\ \delta. & \text{if } p_0(\omega^n) = \lambda_n p_1(\omega^n), \\ 0. & \text{if } p_0(\omega^n) < \lambda_n p_1(\omega^n), \end{cases} \tag{52}$$

where λ_n is a positive number and $0 \le \delta \le 1$.

Now, we claim

$$\lim_{n\to\infty} -\frac{1}{n}\log \lambda_n = \tau(t^*). \tag{53}$$

Remember $\tau(t^*) = D(p_{t^*}\|p_0) - D(p_{t^*}\|p_1)$. From the assumption $0 < r < D(p_1\|p_0)$, we have $0 < t^* < 1$ by Lemma 5. We show that a chain of inequalities

$$\tau(t^*) \le \liminf_{n\to\infty} -\frac{1}{n}\log \lambda_n \le \limsup_{n\to\infty} -\frac{1}{n}\log \lambda_n \le \tau(t^*) \tag{54}$$

hold. Suppose $\tau(t^*) > \liminf_{n\to\infty} -1/n\log\lambda_n$. Then, by Lemma 5, there exists a t_1, $0 < t_1 < t^*$, such that $\tau(t^*) > \tau(t_1) > -1/n\log\lambda_n$ hold for infinitely many n's. Then, by Neyman–Pearson's lemma, we have

$$\begin{aligned} \alpha(\phi_n^*) &\ge p_0\left\{\omega^n | \tau(p_{\omega^n}) > -\frac{1}{n}\log \lambda_n\right\} \\ &\ge p_0\{\omega^n|\tau(p_{\omega^n}) \ge \tau(t_1)\} \\ &= p_0\{\omega^n|p_{\omega^n} \in G_1(\tau(t_1))\} \end{aligned} \tag{55}$$

for infinitely many n's. Since $t_1 > 0$, we have $p_0 \notin G_1(\tau(t_1))$. Therefore, by Lemmas 2 and 7, for any ϵ with $0 < \epsilon < D(p_{t^*}\|p_0) - D(p_{t_1}\|p_0)$, and for some n sufficiently large, we have by (55),

$$\begin{aligned} r &= -\frac{1}{n}\log \alpha(\phi_n^*) \\ &\le -\frac{1}{n}\log p_0\{\omega^n|p_{\omega^n} \in G_1(\tau(t_1))\} \\ &\le \min_{p\in G_1(\tau(t_1))} D(p\|p_0) + \epsilon \\ &= D(p_{t_1}\|p_0) + \epsilon \\ &< D(p_{t^*}\|p_0) \\ &= r. \end{aligned} \tag{56}$$

This is a contradiction, so we have $\tau(t^*) \le \liminf_{n\to\infty} -1/n\log\lambda_n$. We can show $\limsup_{n\to\infty} -1/n\log\lambda_n \le \tau(t^*)$; however, the proof is almost the same as before, so we omit it.

Now, we prove $\rho_1(r) = D(p_{t^*}\|p_1)$. From (53) and Lemma 5, for any $\epsilon_1 > 0$, there exists a t_1, $0 < t_1 < t^* < 1$ such that

$$\tau(t_1) < -\frac{1}{n}\log \lambda_n \tag{57}$$

and

$$D(p_{t_1}\|p_1) < D(p_{t^*}\|p_1) + \epsilon_1 \tag{58}$$

hold for sufficiently large n's. Then, we have

$$\begin{aligned} \beta(\phi_n^*) &\ge p_1\left\{\omega^n|\tau(p_{\omega^n}) < -\frac{1}{n}\log\lambda_n\right\} \\ &\ge p_1\{\omega^n|\tau(p_{\omega^n}) \le \tau(t_1)\} \\ &= p_1\{\omega^n|p_{\omega^n} \in G_0(\tau(t_1))\}. \end{aligned}$$

Since $t_1 < 1$, we have $p_1 \notin G_0(\tau(t_1))$. Thus, by Lemmas 2 and 7, for any $\epsilon_2 > 0$ and for all sufficiently large n's, we have

$$\begin{aligned} -\frac{1}{n}\log\beta(\phi_n^*) &\le -\frac{1}{n}\log p_1\{\omega^n|p_{\omega^n} \in G_0(\tau(t_1))\} \\ &\le \min_{p\in G_0(\tau(t_1))} D(p\|p_1) + \epsilon_2 \\ &= D(p_{t_1}\|p_1) + \epsilon_2 \\ &< D(p_{t^*}\|p_1) + \epsilon_1 + \epsilon_2. \end{aligned}$$

Because ϵ_1 and ϵ_2 are arbitrary positive numbers, we have $\limsup_{n\to\infty} -1/n\log\beta(\phi_n^*) \le D(p_{t^*}\|p_1)$. The inequality $D(p_{t^*}\|p_1) \le \liminf_{n\to\infty} -1/n\log\beta(\phi_n^*)$ can be proved in the same way. □

VI. Converse Theorem I

We see from Theorem 1 that if $0 < r < D(p_1\|p_0)$, $\beta(\phi_n^*)$ converges to 0 as n tends to infinity. On the contrary, if $r > D(p_1\|p_0)$, $\beta(\phi_n^*)$ converges to 1. So, in this case, we calculate the power exponent $\lim_{n\to\infty} -1/n\log(1-\beta(\phi_n^*))$. We will show that the case $r > D(p_1\|p_0)$ is divided into two separate cases, i.e., $D(p_1\|p_0) < r < D(p_\infty\|p_0)$ and $r > D(p_\infty\|p_0)$. If $r < D(p_\infty\|p_0)$, including Theorem 1, the randomization of testing is not essential as we can see in the proof of Theorem 1; however, if $r > D(p_\infty\|p_0)$, the finiteness of the power exponent is guaranteed by randomization.

Let $\rho_2(r)$ be the power exponent in the case where $D(p_1\|p_0) < r < D(p_\infty\|p_0)$, i.e., $\rho_2(r) \triangleq \lim_{n\to\infty} -1/n\log(1-\beta(\phi_n^*))$. We present the following theorem.

Theorem 2: If $D(p_1\|p_0) < r < D(p_\infty\|p_0)$, we have

$$\rho_2(r) = D(p_{t^*}\|p_1). \tag{59}$$

where $p_{t^*} = (p_{t^*}(i))_{i\in\Omega}$ is given by $p_{t^*}(i) = C_{t^*}\{p_0(i)\}^{1-t^*}\{p_1(i)\}^{t^*}$, $i \in \Omega$, $C_{t^*}^{-1} = \sum_{i\in\Omega}\{p_0(i)\}^{1-t^*}\{p_1(i)\}^{t^*}$, and $D(p_{t^*}\|p_0) = r$, $t^* > 0$.

Proof: In the same way as the proof of Theorem 1, for the most powerful test function ϕ_n^*, we consider the critical condition in Neyman–Pearson's lemma, i.e., $\tau(p_{\omega^n}) = -1/n\log\lambda_n$. We claim

$$\lim_{n\to\infty} -\frac{1}{n}\log\lambda_n = \tau(t^*). \tag{60}$$

Suppose $\tau(t^*) > \liminf_{n\to\infty} -1/n\log\lambda_n$. By the assumption $D(p_1\|p_0) < r < D(p_\infty\|p_0)$, we notice that $1 < t^* < \infty$. Then, by Lemma 5, there exists a t_1, $1 < t_1 < t^*$, such that

$$\tau(t^*) > \tau(t_1) > -\frac{1}{n}\log\lambda_n \tag{61}$$

hold for infinitely many n's. By Neyman–Pearson's lemma, we have

$$\begin{aligned}\alpha(\phi_n^*) &\geq p_0\left\{\omega^n \middle| \tau(p_{\omega^n}) > -\frac{1}{n}\log\lambda_n\right\}\\ &\geq p_0\{\omega^n|\tau(p_{\omega^n}) \geq \tau(t_1)\}\\ &= p_0\{\omega^n|p_{\omega^n} \in G_1(\tau(t_1))\}, \end{aligned} \tag{62}$$

for infinitely many n's. Since $t_1 > 1$, we have $p_0 \notin G_1(\tau(t_1))$. Therefore, by Lemmas 2 and 7, for any ϵ with $0 < \epsilon < D(p_{t^*}\|p_0) - D(p_{t_1}\|p_0)$, and for all sufficiently large n's, we have

$$\begin{aligned} r &= -\frac{1}{n}\log\alpha(\phi_n^*)\\ &\leq \min_{p\in G_1(\tau(t_1))} D(p\|p_0) + \epsilon\\ &= D(p_{t_1}\|p_0) + \epsilon\\ &< D(p_{t^*}\|p_0)\\ &= r, \end{aligned}$$

which is a contradiction. Thus, we have $\tau(t^*) \leq \liminf_{n\to\infty} -1/n\log\lambda_n$. Similarly, we can show $\limsup_{n\to\infty} -1/n\log\lambda_n \leq \tau(t^*)$, which together with the previous inequality implies (60).

By Neyman–Pearson's lemma, we have

$$1-\beta(\phi_n^*) \geq p_1\left\{\omega^n \middle| \tau(p_{\omega^n}) \geq -\frac{1}{n}\log\lambda_n\right\}.$$

Then, through an argument similar to the proof of Theorem 1, we consequently obtain (59). (Here, it is easy to check that (59) can be achieved also by using a nonrandomized test function.) □

VII. Converse Theorem II

We notice that the proof of Theorem 2 is not available if $r > D(p_\infty\|p_0)$. As we see later, if $r > D(p_\infty\|p_0)$, the set $\{\omega^n|\tau(p_{\omega^n}) > -1/n\log\lambda_n\}$ is empty for all sufficiently large n's. So the number δ in Neyman–Pearson's lemma determines the power exponent. The randomization is essential in this case.

Let $\rho_3(r)$ be the power exponent in the case $r > D(p_\infty\|p_0)$, i.e., $\rho_3(r) \triangleq \lim_{n\to\infty} -1/n\log(1-\beta(\phi_n^*))$.

Theorem 3: If $r > D(p_\infty\|p_0)$, we have

$$\rho_3(r) = r - D(p_\infty\|p_0) + D(p_\infty\|p_1). \tag{63}$$

Proof: First, for the most powerful test function ϕ_n^*, we see that

$$-\frac{1}{n}\log\lambda_n = \tau(\infty) \tag{64}$$

holds for all sufficiently large n's. For, by Lemma 2, we have

$$\lim_{n\to\infty} -\frac{1}{n}\log p_0\{\omega^n|\tau(p_{\omega^n}) = \tau(\infty)\} = D(p_\infty\|p_0) < \infty, \tag{65}$$

thus, we find that $p_0\{\omega^n|\tau(p_{\omega^n}) = \tau(\infty)\} > 0$ for all sufficiently large n's. While the set $\{\omega^n|\tau(p_{\omega^n}) > \tau(\infty)\}$ is empty because $\tau(p_{\omega^n}) \leq \tau(\infty)$ holds for all n's by (37), (41), and (42). Therefore, it is readily seen that a test function $\tilde{\phi}_n$ defined by

$$\tilde{\phi}_n(\omega^n) = \begin{cases} 1, & \text{if } p_0(\omega^n) > \lambda_n p_1(\omega^n),\\ \delta, & \text{if } p_0(\omega^n) = \lambda_n p_1(\omega^n),\\ 0, & \text{if } p_0(\omega^n) < \lambda_n p_1(\omega^n), \end{cases} \tag{66}$$

with $\lambda_n = \exp(-n\tau(\infty))$, and

$$\delta = 1 - \frac{\exp(-rn)}{p_0\{\omega^n|\tau(p_{\omega^n}) = \tau(\infty)\}}$$

satisfies

$$\alpha(\tilde{\phi}_n) = \exp(-rn) \text{ and } 0 \leq \delta \leq 1, \tag{67}$$

for all sufficiently large n's, by using the assumption $r > D(p_\infty\|p_0)$ and (65). Hence, this $\tilde{\phi}_n$ is the most powerful test function, i.e., $\tilde{\phi}_n = \phi_n^*$, and $\lambda_n = \exp(-n\tau(\infty))$ or (64) holds for all sufficiently large n's.

Now, we show (63). For n, which satisfies (64), we have, by (66),

$$\begin{aligned} 1-\beta(\phi_n^*) &= 1 - p_1\{\omega^n|p_0(\omega^n) > \lambda_n p_1(\omega^n)\}\\ &\quad - \delta p_1\{\omega^n|p_0(\omega^n) = \lambda_n p_1(\omega^n)\}\\ &= (1-\delta)p_1\{\omega^n|\tau(p_{\omega^n}) = \tau(\infty)\}\\ &= \frac{\exp(-rn)p_1\{\omega^n|\tau(p_{\omega^n}) = \tau(\infty)\}}{p_0\{\omega^n|\tau(p_{\omega^n}) = \tau(\infty)\}}. \end{aligned}$$

Thus, from Lemmas 2 and 6, we have (63), i.e.,

$$\lim_{n\to\infty} -\frac{1}{n}\log(1-\beta(\phi_n^*)) = r - D(p_\infty\|p_0) + D(p_\infty\|p_1).$$

□

VIII. Relation with Han–Kobayashi's Result

We now show that the power exponents obtained in Han–Kobayashi's paper [5] coincide with those obtained in our Theorems 2 and 3. For, if $D(p_1\|p_0) < r < D(p_\infty\|p_0)$, by Lemmas 4 and 5, we have

$$\begin{aligned}\min_{p:\ D(p\|p_0)\leq r} f_r(p) &= r - \max_{t:\ D(p_t\|p)\leq r}\tau(t)\\ &= r - \tau(t^*)\\ &= D(p_{t^*}\|p_1), \end{aligned}$$

where t^* is determined by $D(p_p^*\|p_0) = r$, $t^* > 0$. While, if $r > D(p_\infty\|p_0)$, we have

$$\begin{aligned}\min_{p:\ D(p\|p_0)\leq r} f_r(p) &= r - \max_{t:\ D(p_t\|p)\leq r}\tau(t)\\ &= r - \tau(\infty)\\ &= r - D(p_\infty\|p_0) + D(p_\infty\|p_1). \end{aligned}$$

Incidentally, the proof of Theorem 2 in [5] includes a small mistake. In our notation, if $N = 1$, $E_\infty = E(\tau(\infty))$ is a vertex of $\overline{\Delta}$ (we consider $\overline{\Delta}$ as a simplex in $\boldsymbol{R}^{m+1}$). Remember N is the number of indexes that attain the maximum of $\log(p_1(i)/p_0(i))$, $i \in \Omega$. So, $N = 1$ is the generic case. In this case, in the notation of [5], $X^{(n)}$ coincides with $\tilde{X}_0$, and $\tilde{X}_0$ coincides with our E_∞, a vertex. They considered proper subsets of $S_0^n(X^{(n)})$ (where $S_0^n(X^{(n)})$ is the set of n-sequences whose type is $X^{(n)}$, in the notation of [5]); however, this is impossible because $|S_0^n(X^{(n)})| = |S_0^n(\tilde{X}_0)| = 1$, $S_0^n(X^{(n)})$ does not have a proper subset. This is also the reason we must use randomized test functions.

IX. General Constraint

Now, we consider a more general constraint $\alpha_n \leq c_n$. If the sequence $\{-1/n\log c_n\}_{n=1}^\infty$ is decreasing, write $r_0 = \lim_{n\to\infty} -1/n\log c_n$. If $0 < r_0 < D(p_1\|p_0)$, we obtain the power exponent as

$$\lim_{n\to\infty}\frac{1}{n}\log\beta(\phi_n^*) = \rho_1(r_0).$$

Then, for example, under the constraint of the type $\alpha_n \leq 1/n^k$, $k > 0$, we see that the power exponent equals $D(p_1\|p_0)$, which is the same result as under the constraint $\alpha_n \leq \epsilon$, $\epsilon > 0$.

References

[1] S. Amari, *Differential-Geometrical Methods in Statistics* (Lecture Notes in Statistics). New York: Springer-Verlage, 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. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 405–417, July 1974.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

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

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

[7] H. Nagaoka and S. Amari, "Differential geometry of smooth families of probability distributions," Tech. Rep. METR 82-7, Dept. Math. Eng., Univ. of Tokyo, 1982.