

Article

First- and Second-Order Hypothesis Testing for Mixed Memoryless Sources

Te Sun Han ¹ and Ryo Nomura ^{2,*}

¹ National Institute of Information and Communications Technology (NICT), Tokyo 184-8795, Japan; han@is.uec.ac.jp

² School of Network and Information, Senshu University, Kanagawa 214-8580, Japan

* Correspondence: nomu@isc.senshu-u.ac.jp; Tel.: +81-44-911-1261

Received: 23 January 2018; Accepted: 5 March 2018; Published: 6 March 2018

Abstract: The first- and second-order optimum achievable exponents in the simple hypothesis testing problem are investigated. The optimum achievable exponent for type II error probability, under the constraint that the type I error probability is allowed asymptotically up to ε , is called the ε -optimum exponent. In this paper, we first give the second-order ε -optimum exponent in the case where the null hypothesis and alternative hypothesis are a mixed memoryless source and a stationary memoryless source, respectively. We next generalize this setting to the case where the alternative hypothesis is also a mixed memoryless source. Secondly, we address the first-order ε -optimum exponent in this setting. In addition, an extension of our results to the more general setting such as hypothesis testing with mixed general source and a relationship with the general compound hypothesis testing problem are also discussed.

Keywords: general source; hypothesis testing; information spectrum; mixed source; optimum exponent

1. Introduction

Let $\mathbf{X} = \{X^n\}_{n=1}^\infty$ and $\bar{\mathbf{X}} = \{\bar{X}^n\}_{n=1}^\infty$ be two general sources (cf. Han [1]), where we use the term of general source to denote a sequence of random variables X^n (respectively, \bar{X}^n) indexed by block length n , where each component of X^n (respectively, \bar{X}^n) takes values in alphabet \mathcal{X} and may vary depending on n .

We consider the hypothesis testing problem with null hypothesis \mathbf{X} , alternative hypothesis $\bar{\mathbf{X}}$ and acceptance region $\mathcal{A}_n \subset \mathcal{X}^n$. The probabilities of type I error and type II error are defined, respectively, as

$$\mu_n := \Pr \{X^n \notin \mathcal{A}_n\}, \quad \lambda_n := \Pr \{\bar{X}^n \in \mathcal{A}_n\}. \quad (1)$$

We focus mainly on how to determine the ε -optimum exponent, defined as the supremum of achievable exponents R for the type II error probability $\lambda_n \simeq e^{-nR}$ under the constraint that the type I error probability is allowed asymptotically up to a constant ε ($0 \leq \varepsilon < 1$). The classical but fundamental result in this setting is so-called Stein's lemma [2], which gives the ε -optimum exponent in the case where both the null and alternative hypotheses are stationary memoryless sources. The lemma shows that the ε -optimum exponent is given by $D(P_X || P_{\bar{X}})$, the divergence between stationary memoryless sources X and \bar{X} . Chen [3] has generalized this lemma to the case where both of \mathbf{X} and $\bar{\mathbf{X}}$ are general sources, and established the general formula of ε -optimum exponent in terms of divergence spectra. The ε -optimum exponent derived in [3] is called in this paper the first-order ε -optimum exponent.

On the other hand, the second-order asymptotics have also been investigated in several contexts of information theory [4–9] to analyze the finer asymptotic behavior of the form $\lambda_n \simeq e^{-nR - \sqrt{n}S}$.

Strassen [4] has first introduced the notion of ε -optimum achievable exponent of the second-order in hypothesis testing problem in the case where both of \mathbf{X} and $\bar{\mathbf{X}}$ are stationary memoryless sources. The results in [4] have also revealed that the asymptotic normality of divergence density rate (or likelihood ratio rate) plays an important role in computing the second-order ε -optimum exponent.

In this paper, on the other hand, we investigate the hypothesis testing for mixed memoryless sources. The class of mixed sources is quite important, because all stationary sources can be regarded as mixed sources consisting of stationary ergodic sources. Therefore, the analysis for mixed sources is primitive but fundamental and thus we first focus on the case where the null hypothesis is a mixed memoryless source and the alternative hypothesis is a memoryless source. In this direction, Han [1] has first derived the single-letter formula for the first-order ε -optimum exponent in the case with mixed memoryless source \mathbf{X} and stationary memoryless source $\bar{\mathbf{X}}$. The first main result in this paper is to establish the single-letter second-order ε -optimum exponent in the same setting by invoking the relevant asymptotic normality. The result is a substantial generalization of that of Strassen [4]. Second, we generalize this setting to the case where both null and alternative hypotheses are mixed memoryless \mathbf{X} , $\bar{\mathbf{X}}$ to establish the single-letter first-order ε -optimum exponent.

It should be emphasized that our results described here are valid for mixed memoryless sources with general mixture in the sense that the mixing weight for component sources may be an arbitrary probability measure. For the case of mixed general sources with finite discrete mixture, we reveal the deep relationship with the compound hypothesis testing problem. We notice that the compound hypothesis testing problem is important from both of theoretical and practical points of view. We show that the first-order 0-optimum (respectively, exponentially r -optimum) exponent for the mixed general hypothesis testing coincides with that for the 0-optimum (respectively, exponentially r -optimum) exponent in the compound general hypothesis testing.

The present paper is organized as follows. In Section 2, we fix the problem setting and review the general formula (Theorem 1) for the first-order ε -optimum exponent. This is used to prove Theorem 5 to establish a first-order single-letter formula for hypothesis testing in the case where both the null and alternative hypotheses are mixed memoryless. Moreover, we give the general formula (Theorem 2) for the second-order ε -optimum exponent, which is used to prove Theorem 4 to establish a second-order single-letter formula for hypothesis testing in the case where the null hypothesis is mixed memoryless and the alternative hypothesis is stationary memoryless. In Section 3, we establish the single-letter second-order ε -optimum exponent in the case with mixed memoryless source \mathbf{X} and stationary memoryless source $\bar{\mathbf{X}}$ (cf. Theorem 4). Furthermore, in Section 4, we consider the case where both of null and alternative hypotheses are mixed memoryless sources, and derive the single-letter first-order ε -optimum exponent (cf. Theorem 5). Section 5 is devoted to an extension of mixed memoryless sources to mixed general sources. Finally, in Section 6, we define the optimum exponent for the compound general hypothesis testing problem to discuss a relevant relationship with the hypothesis testing with mixed general sources. We conclude the paper in Section 7.

2. General Formulas for ε -Hypothesis Testing

In this section, we first review the first-order general formula and then give the second-order general formula. Throughout in this paper, the following lemmas play the important role, where we use the notation that P_Z indicates the probability distribution of random variable Z .

Lemma 1 ([1] (Lemma 4.1.1)). *For any $t > 0$, define the acceptance region as*

$$\mathcal{A}_n = \left\{ \mathbf{x} \in \mathcal{X}^n \mid \frac{1}{n} \log \frac{P_{\mathbf{X}^n}(\mathbf{x})}{P_{\bar{\mathbf{X}}^n}(\mathbf{x})} \geq t \right\}, \quad (2)$$

then, it holds that

$$\Pr \left\{ \bar{X}^n \in \mathcal{A}_n \right\} \leq e^{-nt}. \quad (3)$$

Lemma 2 ([1] (Lemma 4.1.2)). For any $t > 0$ and any \mathcal{A}_n , it holds that

$$\Pr \{X^n \notin \mathcal{A}_n\} + e^{nt} \Pr \left\{ \bar{X}^n \in \mathcal{A}_n \right\} \geq \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(\bar{X}^n)} \leq t \right\}. \quad (4)$$

Proofs of these lemmas are found in [1].

We define the first and second-order ε -optimum exponents as follows.

Definition 1. Rate R is said to be ε -achievable, if there exists an acceptance region \mathcal{A}_n such that

$$\limsup_{n \rightarrow \infty} \mu_n \leq \varepsilon \text{ and } \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n} \geq R. \quad (5)$$

Definition 2 (First-order ε -optimum exponent).

$$B_\varepsilon(\mathbf{X}|\bar{\mathbf{X}}) := \sup \{R | R \text{ is } \varepsilon\text{-achievable}\}. \quad (6)$$

The right-hand side of Equation (5) specifies the asymptotic behavior of the form $\lambda_n \simeq e^{-nR}$. Chen [3] has derived the general limiting formula for $B_\varepsilon(\mathbf{X}|\bar{\mathbf{X}})$ as follows, which is utilized to establish Theorem 5 in Section 4.

Theorem 1 (Chen [3] (Theorem 1)).

$$B_\varepsilon(\mathbf{X}|\bar{\mathbf{X}}) = \sup \{R | K(R) \leq \varepsilon\} \quad (0 \leq \forall \varepsilon < 1), \quad (7)$$

where

$$K(R) = \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(\bar{X}^n)} \leq R \right\}. \quad (8)$$

Moreover, we consider the second-order (ε, R) -optimum exponent as follows.

Definition 3. Rate S is said to be (ε, R) -achievable, if there exists an acceptance region \mathcal{A}_n such that

$$\limsup_{n \rightarrow \infty} \mu_n \leq \varepsilon \text{ and } \liminf_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \log \frac{1}{\lambda_n e^{nR}} \geq S. \quad (9)$$

Definition 4 (Second-order (ε, R) -optimum exponent).

$$B_\varepsilon(R|\mathbf{X}|\bar{\mathbf{X}}) := \sup \{S | S \text{ is } (\varepsilon, R)\text{-achievable}\}. \quad (10)$$

The right-hand side of Equation (9) specifies the asymptotic behavior of the form $\lambda_n \simeq e^{-nR - \sqrt{n}S}$. The general limiting formula for $B_\varepsilon(R|\mathbf{X}|\bar{\mathbf{X}})$ is given as follows, which is the second-order counterpart of Theorem 1, and is utilized to establish Theorem 4 in the next Section 3.2 to give a second-order single-letter formula for hypothesis testing in the case where the null hypothesis is mixed memoryless and the alternative hypothesis is stationary memoryless.

Theorem 2.

$$B_\varepsilon(R|\mathbf{X}|\bar{\mathbf{X}}) = \sup \{S | K(R, S) \leq \varepsilon\} \quad (0 \leq \forall \varepsilon < 1), \quad (11)$$

where

$$K(R, S) = \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S}{\sqrt{n}} \right\}. \quad (12)$$

Proof. See Appendix A. \square

3. Mixed Memoryless Sources

3.1. First-Order ε -Optimum Exponent

In the previous section, we have demonstrated the “limiting” formulas for general hypothesis testing. In this and subsequent sections, we consider special but insightful cases and compute the optimum exponents in single-letter forms.

Let Θ be an arbitrary probability space with general probability measure $w(\theta)$ ($\theta \in \Theta$). Then, the hypothesis testing problem to be considered in this section is stated as follows:

- The null hypothesis is a mixed stationary memoryless source $\mathbf{X} = \{X^n\}_{n=1}^\infty$, that is, for $\forall \mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$

$$P_{X^n}(\mathbf{x}) = \int_{\Theta} P_{X_\theta^n}(\mathbf{x}) dw(\theta), \quad (13)$$

where X_θ^n is a stationary memoryless source for each $\theta \in \Theta$ and

$$P_{X_\theta^n}(\mathbf{x}) = \prod_{i=1}^n P_{X_\theta}(x_i) \quad (14)$$

with generic random variable X_θ ($\theta \in \Theta$) taking values in \mathcal{X} .

- The alternative hypothesis is a stationary memoryless source $\bar{\mathbf{X}} = \{\bar{X}^n\}_{n=1}^\infty$ with generic random variable \bar{X} taking values in \mathcal{X} , that is,

$$P_{\bar{X}^n}(\mathbf{x}) = \prod_{i=1}^n P_{\bar{X}}(x_i). \quad (15)$$

We assume \mathcal{X} to be a finite alphabet hereafter.

To investigate this special case, first we introduce an expurgated parameter set on the basis of types, where the type T of sequence $\mathbf{x} \in \mathcal{X}^n$ is the empirical distribution of \mathbf{x} , that is, $T = (N(x|\mathbf{x})/n)_{x \in \mathcal{X}}$ with the number $N(x|\mathbf{x})$ of i such that $x_i = x$ ($i = 1, 2, \dots, n$).

Let T_1, T_2, \dots, T_{N_n} denote all possible types of sequences of length n . Then, it is well-known that

$$N_n \leq (n+1)^{|\mathcal{X}|}. \quad (16)$$

Now, for each $\mathbf{x} \in \mathcal{X}^n$, we define the set

$$\Theta(\mathbf{x}) := \left\{ \theta \in \Theta \mid P_{X_\theta^n}(\mathbf{x}) \leq e^{\frac{4}{\sqrt{n}}} P_{X^n}(\mathbf{x}) \right\}. \quad (17)$$

Since $P_{X_\theta^n}$ is an i.i.d. source for each $\theta \in \Theta$, the set $\Theta(\mathbf{x})$ depends only on the type T_k of sequence \mathbf{x} , and therefore, we may write $\Theta(T_k)$ instead of $\Theta(\mathbf{x})$. Moreover, we define the “expurgated” set

$$\Theta_n^* := \bigcap_{k=1}^{N_n} \Theta(T_k). \quad (18)$$

Then, we have the following lemma:

Lemma 3 (Han [1]). Let $\mathbf{X} = \{X^n\}_{n=1}^\infty$ denote a mixed memoryless source defined in Equation (13), then we have

$$\int_{\Theta_n^*} dw(\theta) \geq 1 - (n+1)^{|\mathcal{X}|} e^{-\frac{4}{\sqrt{n}}}. \quad (19)$$

Next, we introduce two basic “decomposition” lemmas as follows.

Lemma 4 (Upper Decomposition Lemma). Let $\mathbf{X} = \{X^n\}_{n=1}^\infty$ be a mixed memoryless source and $\bar{\mathbf{X}} = \{\bar{X}^n\}_{n=1}^\infty$ be an arbitrary general source. Then, for any $\theta \in \Theta_n^*$ and any real z_n it holds that

$$\Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq z_n \right\} \leq \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq z_n + \frac{1}{\sqrt{n}^3} \right\}. \quad (20)$$

Proof. See Appendix B. \square

Lemma 5 (Lower Decomposition Lemma). Let $\mathbf{X} = \{X^n\}_{n=1}^\infty$ be a mixed memoryless source and $\bar{\mathbf{X}} = \{\bar{X}^n\}_{n=1}^\infty$ be an arbitrary general source. Then, for any $\theta \in \Theta$, z_n and $\gamma > 0$ it holds that

$$\Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq z_n \right\} \geq \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq z_n - \frac{\gamma}{\sqrt{n}} \right\} - e^{-\sqrt{n}\gamma}. \quad (21)$$

Proof. See Appendix C. \square

These Lemmas 3–5 are used later in order to establish Theorems 3–5. First, Theorem 3 concerning the first-order ε -optimum exponent for mixed memoryless sources has earlier been given as follows:

Theorem 3 (First-order ε -optimum exponent: Han [1]). For $0 \leq \varepsilon < 1$,

$$B_\varepsilon(\mathbf{X}||\bar{\mathbf{X}}) = \sup \left\{ R \left| \int_{\{\theta | D(P_{X_\theta}||P_{\bar{X}}) < R\}} dw(\theta) \leq \varepsilon \right. \right\} \quad (22)$$

where $D(P_X||P_{\bar{X}})$ denotes the Kullback–Leibler divergence between P_X and $P_{\bar{X}}$.

Remark 1. If Θ is a singleton, the above formula reduces to

$$B_\varepsilon(\mathbf{X}||\bar{\mathbf{X}}) = D(P_X||P_{\bar{X}}) \quad (0 \leq \forall \varepsilon < 1), \quad (23)$$

which is nothing but Stein’s lemma [2].

Remark 2. $B_\varepsilon(\mathbf{X}||\bar{\mathbf{X}})$ can be expressed also as

$$B_\varepsilon(\mathbf{X}||\bar{\mathbf{X}}) = \sup \left\{ R \left| \int_{\{\theta | D(P_{X_\theta}||P_{\bar{X}}) \leq R\}} dw(\theta) \leq \varepsilon \right. \right\}. \quad (24)$$

This can be verified as follows. Set

$$\beta_\varepsilon := \sup \left\{ R \left| \int_{\{\theta | D(P_{X_\theta}||P_{\bar{X}}) < R\}} dw(\theta) \leq \varepsilon \right. \right\}, \quad (25)$$

$$\tilde{\beta}_\varepsilon := \sup \left\{ R \left| \int_{\{\theta | D(P_{X_\theta}||P_{\bar{X}}) \leq R\}} dw(\theta) \leq \varepsilon \right. \right\}. \quad (26)$$

Then, clearly $\tilde{\beta}_\varepsilon \leq \beta_\varepsilon$. Here, we assume that $\tilde{\beta}_\varepsilon < \beta_\varepsilon$ to show a contradiction. From the assumption, there exists a constant $\gamma > 0$ satisfying $\tilde{\beta}_\varepsilon + 2\gamma < \beta_\varepsilon$. On the other hand, from the definition of β_ε , for any $\eta > 0$

$$\varepsilon \geq \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) < \beta_\varepsilon - \eta\}} dw(\theta) \quad (27)$$

holds. Thus, setting $\eta < \gamma$ leads to

$$\begin{aligned} \varepsilon &\geq \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) < \beta_\varepsilon - \eta\}} dw(\theta) \\ &\geq \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) < \beta_\varepsilon - \gamma\}} dw(\theta) \\ &\geq \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) \leq \tilde{\beta}_\varepsilon + \gamma\}} dw(\theta) \\ &> \varepsilon, \end{aligned} \quad (28)$$

which is a contradiction, where the last inequality is due to the definition of $\tilde{\beta}_\varepsilon$.

3.2. Second-Order ε -Optimum Exponent

Next, we establish the second-order ε -optimum exponent for mixed sources, which is the first main result in this paper.

Theorem 4 (Second-order ε -optimum exponent). For $0 \leq \varepsilon < 1$,

$$B_\varepsilon(R|\mathbf{X}||\bar{\mathbf{X}}) = \sup \left\{ S \left| \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) < R\}} dw(\theta) + \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) = R\}} \Phi_\theta(S) dw(\theta) \leq \varepsilon \right. \right\}, \quad (29)$$

where

$$\Phi_\theta(S) := G\left(\frac{S}{\sqrt{V_\theta}}\right), \quad (30)$$

$$G(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx, \quad (31)$$

$$V_\theta := \sum_{x \in \mathcal{X}} P_{X_\theta}(x) \left(\log \frac{P_{X_\theta}(x)}{P_{\bar{X}}(x)} - D(P_{X_\theta} || P_{\bar{X}}) \right)^2. \quad (32)$$

Proof. See Appendix D. \square

Remark 3. If Θ is a singleton ($\Theta = \{\theta_0\}$), Theorem 4 reduces to $B_\varepsilon(R|\mathbf{X}||\bar{\mathbf{X}}) = \sqrt{V_{\theta_0}} \Phi_{\theta_0}^{-1}(\varepsilon)$ for $R = B_\varepsilon(\mathbf{X}||\bar{\mathbf{X}})$, which is originally due to Strassen [4].

Remark 4. From Theorem 3 with $R = B_\varepsilon(\mathbf{X}||\bar{\mathbf{X}})$, it is not difficult to verify that

$$\int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) < R\}} dw(\theta) \leq \varepsilon \quad (33)$$

and

$$\int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) \leq R\}} dw(\theta) \geq \varepsilon. \quad (34)$$

Here, let us consider the following canonical equation for S

$$\int_{\Theta} dw(\theta) \lim_{n \rightarrow \infty} \Phi_\theta(\sqrt{n}(B_\varepsilon(\mathbf{X}||\bar{\mathbf{X}}) - D(P_{X_\theta} || P_{\bar{X}})) + S) = \varepsilon. \quad (35)$$

In view of Equations (33) and (34), this equation always has a solution $S = S(\varepsilon)$. It should be noted that if $\int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) = B_\varepsilon(\mathbf{X} || \bar{\mathbf{X}})\}} dw(\theta) = 0$ holds, the solution is not unique and so $S(\varepsilon) = +\infty$. By using the solution $S(\varepsilon)$, it is not difficult to check that Theorem 4 with $R = B_\varepsilon(\mathbf{X} || \bar{\mathbf{X}})$ can be expressed as

$$B_\varepsilon(R | \mathbf{X} || \bar{\mathbf{X}}) = S(\varepsilon). \quad (36)$$

The canonical equation is a useful expression for the second-order ε -optimum rate [7,10–12]. Equation (35) is the hypothesis testing counterpart of these results.

4. Mixed Memoryless Alternative Hypothesis

In this section, we consider the case where not only the null hypothesis but also the alternative hypothesis are mixed memoryless sources to establish the single-letter formula for the first-order ε -optimum exponent, by which we intend to generalize Theorem 3.

Let $\{P_{\bar{X}_\sigma}\}_{\sigma \in \Sigma}$ be a family of probability distributions on \mathcal{X} , where Σ is a probability space with probability measure $v(\sigma)$. We assume here that Σ is a compact space and $P_{\bar{X}_\sigma}$ is continuous as a function of $\sigma \in \Sigma$.

The hypothesis testing problem considered in this section is stated as follows:

- The null hypothesis is a mixed memoryless source $\mathbf{X} = \{X^n\}_{n=1}^\infty$ as defined by Equations (13) and (14) in Section 3.1.
- The alternative hypothesis is another mixed memoryless source $\bar{\mathbf{X}} = \{\bar{X}^n\}_{n=1}^\infty$, that is, for $\forall \mathbf{x} \in \mathcal{X}^n$

$$P_{\bar{\mathbf{X}}^n}(\mathbf{x}) = \int_{\Sigma} P_{\bar{X}_\sigma^n}(\mathbf{x}) dv(\sigma), \quad (37)$$

where

$$P_{\bar{X}_\sigma^n}(\mathbf{x}) = \prod_{i=1}^n P_{\bar{X}_\sigma}(x_i). \quad (38)$$

Let us now consider, for each $P \in \mathcal{P}(\mathcal{X})$ (the set of probability distributions on \mathcal{X}), the equation with respect to $\sigma' \in \Sigma$ as follows:

$$D(P || \bar{P}_{\sigma'}) = v\text{-ess. inf } D(P || \bar{P}_\sigma) \quad (\text{for each } P \in \mathcal{P}(\mathcal{X})) \quad (39)$$

with $v\text{-ess. inf } f_\sigma := \sup\{\beta | \Pr\{f_\sigma < \beta\} = 0\}$ (the essential infimum of f_σ with respect to $v(\sigma)$), where “ \Pr ” is measured with respect to the probability measure $v(\sigma)$.

Since the solution σ' of this equation depends on P , we may write as $\sigma' = \sigma(P)$ ($\sigma(\cdot) : \mathcal{P}(\mathcal{X}) \rightarrow \Sigma$). Notice here that $D(P || \bar{P}_\sigma)$ is continuous in (P, \bar{P}_σ) , and as we have assumed that Σ is compact and \bar{P}_σ is continuous in σ , there indeed exists such a function $\sigma(P)$. Now, to avoid technical subtleties, we assume here that the function $\sigma(P)$ may be chosen so as to be continuous. For example, if we consider a special case such that Σ is a closed convex subset of $\mathcal{P}(\mathcal{X})$, then it is not difficult to verify that the function $\sigma(P)$ is uniquely determined and continuous (or even differentiable), which follows from the strict convexity of $D(P || \bar{P})$ in (P, \bar{P}) . Another simple example will be the case that Σ is a countable set.

Hereafter, for simplicity, we write P_θ, P_θ^n (respectively, $\bar{P}_\sigma, \bar{P}_\sigma^n$) instead of $P_{X_\theta}, P_{X_\theta^n}$ (respectively, $P_{\bar{X}_\sigma}, P_{\bar{X}_\sigma^n}$), then we have the second main result in this paper as

Theorem 5 (First order ε -optimum exponent). For $0 \leq \varepsilon < 1$,

$$B_\varepsilon(\mathbf{X} || \bar{\mathbf{X}}) = \sup \left\{ R \left| \int_{\{\theta | D(P_\theta || \bar{P}_{\sigma(P_\theta)}) < R\}} dw(\theta) \leq \varepsilon \right. \right\}. \quad (40)$$

Remark 5. In the case that Σ is a singleton, the above theorem coincides with Theorem 3. Therefore, this theorem is a direct generalization of Theorem 3. This means also that both Θ and Σ are singletons, the theorem coincides with Stein's lemma (see Remark 1).

Remark 6. Remark 2 is also valid in this theorem. That is, $B_\varepsilon(\mathbf{X}|\overline{\mathbf{X}})$ can be expressed also as

$$B_\varepsilon(\mathbf{X}|\overline{\mathbf{X}}) = \sup \left\{ R \left| \int_{\{\theta | D(P_\theta || \overline{P}_{\sigma(P_\theta)}) \leq R\}} dw(\theta) \leq \varepsilon \right. \right\}. \quad (41)$$

Proof of Theorem 5. To show the theorem, let $\mathcal{T}_{\theta, \nu}^n \subseteq \mathcal{X}^n$ be the set of ν -typical sequence with respect to P_{X_θ} , that is, let $\mathcal{T}_{\theta, \nu}^n$ be the set of all $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ such that

$$|N(x|\mathbf{x})/n - P_{X_\theta}(x)| \leq \nu P_{X_\theta}(x) \quad (\forall x \in \mathcal{X}), \quad (42)$$

where $N(x|\mathbf{x})$ is the number of i such that $x_i = x$, and $\nu > 0$ is an arbitrary constant. Then, it is well known that

$$\Pr \{X_\theta^n \in \mathcal{T}_{\theta, \nu}^n\} \rightarrow 1 \quad (n \rightarrow \infty). \quad (43)$$

In the sequel, we use the upper and lower bounds of the probability

$$P_{\overline{\mathbf{X}}^n}(\mathbf{x}) = \int_{\Sigma} \overline{P}_{\sigma}^n(\mathbf{x}) dv(\sigma) \quad (44)$$

in the form

$$\frac{1}{n} \log \frac{1}{P_{\overline{\mathbf{X}}^n}(\mathbf{x})} \geq \frac{1}{n} \log \frac{1}{\overline{P}_{\sigma(P_\theta)}^n(\mathbf{x})} - \delta_\theta(\nu), \quad (45)$$

$$\frac{1}{n} \log \frac{1}{P_{\overline{\mathbf{X}}^n}(\mathbf{x})} \leq \frac{1}{n} \log \frac{1}{\overline{P}_{\sigma(P_\theta)}^n(\mathbf{x})} + \frac{1}{n} \log \frac{1}{c_{\tau_m}(P_\theta)} + (\tau - \delta_\theta(\nu)), \quad (46)$$

for each $\mathbf{x} \in \mathcal{T}_{\theta, \nu}^n$, where $\delta_\theta(\nu)$ satisfies $\delta_\theta(\nu) \rightarrow 0$ as $\nu \rightarrow 0$, and $\tau > 0$ and $c_{\tau_m}(P_\theta) > 0$ are some constants independent of n . Proofs of Equations (45) and (46) appear in Appendix E.

We then prove the theorem by using Equations (45) and (46) as follows. In view of Theorem 1 and Remark 6, it suffices to show two inequalities:

$$\limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\overline{\mathbf{X}}^n}(X^n)} \leq R \right\} \leq \int_{\{\theta | D(P_\theta || \overline{P}_{\sigma(P_\theta)}) \leq R\}} dw(\theta), \quad (47)$$

$$\limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\overline{\mathbf{X}}^n}(X^n)} \leq R \right\} \geq \int_{\{\theta | D(P_\theta || \overline{P}_{\sigma(P_\theta)}) < R\}} dw(\theta). \quad (48)$$

- *Proof of Equation (47):*

Similar to the derivation of Equation (A23) with Lemma 4, we have

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\overline{\mathbf{X}}^n}(X^n)} \leq R \right\} \\ &= \limsup_{n \rightarrow \infty} \int_{\Theta} dw(\theta) \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\overline{\mathbf{X}}_\theta^n}(X_\theta^n)} \leq R \right\} \\ &\leq \int_{\Theta} dw(\theta) \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\overline{\mathbf{X}}_\theta^n}(X_\theta^n)} \leq R + \frac{1}{\sqrt[4]{n^3}} \right\}. \end{aligned} \quad (49)$$

From the definition of the ν -typical set and Equation (45), we also have

$$\begin{aligned}
 & \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{1}{\sqrt[4]{n^3}} \right\} \\
 & \leq \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{1}{\sqrt[4]{n^3}}, X_\theta^n \in \mathcal{T}_{\theta, \nu}^n \right\} + \limsup_{n \rightarrow \infty} \Pr \{ X_\theta^n \notin \mathcal{T}_{\theta, \nu}^n \} \\
 & \leq \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq R + \frac{1}{\sqrt[4]{n^3}} + \delta_\theta(\nu), X_\theta^n \in \mathcal{T}_{\theta, \nu}^n \right\} \\
 & \leq \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq R + \frac{1}{\sqrt[4]{n^3}} + \delta_\theta(\nu) \right\}, \tag{50}
 \end{aligned}$$

for any $\theta \in \Theta$. Here, we define two sets:

$$\Theta_1 := \left\{ \theta \in \Theta \mid D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) \leq R \right\}, \tag{51}$$

$$\Theta_2 := \left\{ \theta \in \Theta \mid D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) > R \right\}. \tag{52}$$

Then, from the definition of Θ_2 there exists a small constant $\gamma > 0$ satisfying

$$D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) \geq R + 3\gamma \tag{53}$$

for $\theta \in \Theta_2$. Thus, it holds that

$$\begin{aligned}
 & \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq R + \frac{1}{\sqrt[4]{n^3}} + \delta_\theta(\nu) \right\} \\
 & \leq \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) + \frac{1}{\sqrt[4]{n^3}} + \delta_\theta(\nu) - 3\gamma \right\} \\
 & \leq \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) - \gamma \right\}, \tag{54}
 \end{aligned}$$

where we have used the relation $\frac{1}{\sqrt[4]{n^3}} < \gamma$, and $\delta_\theta(\nu) < \gamma$ for sufficiently large n and sufficiently small $\nu > 0$.

Therefore, noting that, with $X_\theta^n = (X_{\theta,1}, X_{\theta,2}, \dots, X_{\theta,n})$,

$$\frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} = \frac{1}{n} \sum_{i=1}^n \log \frac{P_{X_\theta}(X_{\theta,i})}{\bar{P}_{\sigma(P_\theta)}(X_{\theta,i})} \tag{55}$$

gives the arithmetic average of n i.i.d. variables with expectation

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \log \frac{P_{X_\theta}(X_{\theta,i})}{\bar{P}_{\sigma(P_\theta)}(X_{\theta,i})} \right] = D(P_\theta \| \bar{P}_{\sigma(P_\theta)}). \tag{56}$$

Then, the weak law of large numbers yields that for $\forall \theta \in \Theta_2$,

$$\limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) - \gamma \right\} = 0. \tag{57}$$

Thus, from Equations (54) and (57), the right-hand side of Equation (49) is upper bounded by

$$\begin{aligned}
 & \int_{\Theta} dw(\theta) \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_{\theta}^n}(X_{\theta}^n)}{\bar{P}_{\sigma(P_{\theta})}^n(X_{\theta}^n)} \leq R + \frac{1}{\sqrt[4]{n^3}} + \delta_{\theta}(\nu) \right\} \\
 & \leq \int_{\Theta_1} dw(\theta) \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_{\theta}^n}(X_{\theta}^n)}{\bar{P}_{\sigma(P_{\theta})}^n(X_{\theta}^n)} \leq D(P_{\theta} || \bar{P}_{\sigma(P_{\theta})}) - \gamma \right\} \\
 & \quad + \int_{\Theta_2} dw(\theta) \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_{\theta}^n}(X_{\theta}^n)}{\bar{P}_{\sigma(P_{\theta})}^n(X_{\theta}^n)} \leq D(P_{\theta} || \bar{P}_{\sigma(P_{\theta})}) - \gamma \right\} \\
 & \leq \int_{\Theta_1} dw(\theta) \\
 & = \int_{\{\theta | D(P_{\theta} || \bar{P}_{\sigma(P_{\theta})}) \leq R\}} dw(\theta), \tag{58}
 \end{aligned}$$

which completes the proof of (47).

- *Proof of Equation (48):*

Similar to the derivation of Equation (A32) with Lemma 5, we have

$$\begin{aligned}
 & \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R \right\} \\
 & \geq \int_{\Theta} dw(\theta) \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_{\theta}^n}(X_{\theta}^n)}{P_{\bar{X}^n}(X_{\theta}^n)} \leq R - \frac{\gamma}{\sqrt{n}} \right\}. \tag{59}
 \end{aligned}$$

From the definition of the ν -typical set and Equation (46), we also have

$$\begin{aligned}
 & \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_{\theta}^n}(X_{\theta}^n)}{P_{\bar{X}^n}(X_{\theta}^n)} \leq R - \frac{\gamma}{\sqrt{n}} \right\} \\
 & \geq \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_{\theta}^n}(X_{\theta}^n)}{P_{\bar{X}^n}(X_{\theta}^n)} \leq R - \frac{\gamma}{\sqrt{n}}, X_{\theta}^n \in \mathcal{T}_{\theta, \nu}^n \right\} \\
 & \geq \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_{\theta}^n}(X_{\theta}^n)}{\bar{P}_{\sigma(P_{\theta})}^n(X_{\theta}^n)} \leq R - \frac{\gamma}{\sqrt{n}} - \frac{1}{n} \log \frac{1}{c_{\tau_m}(P_{\theta})} - \tau + \delta_{\theta}(\nu) \right\} \\
 & \quad - \limsup_{n \rightarrow \infty} \Pr \{ X_{\theta}^n \notin \mathcal{T}_{\theta, \nu}^n \} \\
 & = \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_{\theta}^n}(X_{\theta}^n)}{\bar{P}_{\sigma(P_{\theta})}^n(X_{\theta}^n)} \leq R - \frac{\gamma}{\sqrt{n}} - \frac{1}{n} \log \frac{1}{c_{\tau_m}(P_{\theta})} - \tau + \delta_{\theta}(\nu) \right\} \tag{60}
 \end{aligned}$$

for any $\theta \in \Theta$.

We also partition the parameter space Θ into two sets.

$$\Theta'_1 := \left\{ \theta \in \Theta \mid D(P_{\theta} || \bar{P}_{\sigma(P_{\theta})}) < R \right\}, \tag{61}$$

$$\Theta'_2 := \left\{ \theta \in \Theta \mid D(P_{\theta} || \bar{P}_{\sigma(P_{\theta})}) \geq R \right\}. \tag{62}$$

Then, for $\theta \in \Theta'_1$, if we set $\nu > 0$ and $\tau > 0$ sufficiently small, then there exists a constant $\eta > 0$ satisfying

$$R - \frac{\gamma}{\sqrt{n}} - \frac{1}{n} \log \frac{1}{c_{\tau_m}(P_{\theta})} - \tau + \delta_{\theta}(\nu) > D(P_{\theta} || \bar{P}_{\sigma(P_{\theta})}) + \eta \quad (\forall n > n_0). \tag{63}$$

Thus, again by invoking the weak law of large numbers, we have for $\forall \theta \in \Theta'_1$

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq R - \frac{\gamma}{\sqrt{n}} - \frac{1}{n} \log \frac{1}{c_{\tau_m}(P_\theta)} - \tau + \delta_\theta(\nu) \right\} \\ & \geq \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) + \eta \right\} \\ & = 1. \end{aligned} \quad (64)$$

Summarizing up, we obtain

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{\bar{P}_{\bar{X}}(X^n)} \leq R \right\} \\ & \geq \int_{\Theta} dw(\theta) \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq R - \frac{\gamma}{\sqrt{n}} - \frac{1}{n} \log \frac{1}{c_{\tau_m}(P_\theta)} - \tau + \delta_\theta(\nu) \right\} \\ & \geq \int_{\Theta'_1} dw(\theta) \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{\bar{P}_{\sigma(P_\theta)}^n(X_\theta^n)} \leq D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) + \eta \right\} \\ & = \int_{\Theta'_1} dw(\theta) \\ & = \int_{\{\theta | D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) < R\}} dw(\theta). \end{aligned} \quad (65)$$

This completes the proof of Equation (48). \square

Remark 7. Theorem 3 is a special case of Theorem 5 when Σ is a singleton.

To illustrate a significance of Theorem 5, let us now consider the special case with $\varepsilon = 0$. Then, by virtue of Theorem 5, we have the following simplified result:

Corollary 1. In the special case of $\varepsilon = 0$, we have

$$B_0(\mathbf{X} \| \bar{\mathbf{X}}) = w\text{-ess.} \inf_{\theta \in \Theta} v\text{-ess.} \inf_{\sigma \in \Sigma} D(P_\theta \| \bar{P}_\sigma). \quad (66)$$

Proof. The formula (40) can be written in this case as

$$B_0(\mathbf{X} \| \bar{\mathbf{X}}) = \sup \left\{ R \left| \int_{\{\theta | D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) < R\}} dw(\theta) = 0 \right. \right\}. \quad (67)$$

Let

$$R_1 < \sup \left\{ R \left| \int_{\{\theta | D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) < R\}} dw(\theta) = 0 \right. \right\}, \quad (68)$$

then this means that

$$\begin{aligned} R_1 & \leq w\text{-ess.} \inf_{\theta \in \Theta} D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) \\ & = w\text{-ess.} \inf_{\theta \in \Theta} v\text{-ess.} \inf_{\sigma \in \Sigma} D(P_\theta \| \bar{P}_\sigma). \end{aligned} \quad (69)$$

Contrarily, let

$$R_2 > \sup \left\{ R \left| \int_{\{\theta | D(P_\theta \| \bar{P}_{\sigma(P_\theta)}) < R\}} dw(\theta) = 0 \right. \right\}, \quad (70)$$

then this means that

$$\begin{aligned} R_2 &\geq w\text{-ess.} \inf_{\theta \in \Theta} D(P_\theta || \bar{P}_{\sigma(P_\theta)}) \\ &= w\text{-ess.} \inf_{\theta \in \Theta} v\text{-ess.} \inf_{\sigma \in \Sigma} D(P_\theta || \bar{P}_\sigma). \end{aligned} \quad (71)$$

As a consequence, (66) follows from (67), (69) and (71). \square

Remark 8. One may wonder if it might be possible to deal with the second-order ε -optimum problem too using the arguments as developed in the above for the first-order ε -optimum problem with mixed memoryless sources \mathbf{X} and $\bar{\mathbf{X}}$. To do so, however, it seems that we need some novel techniques, which remain to be studied.

5. Hypothesis Testing with Mixed General Sources

We have so far investigated the ε -hypothesis testing for mixed memoryless sources. In this section, we deal with more general settings such as hypothesis testings with mixed general sources, which inherits the crux of that for mixed memoryless sources (cf. Theorem 5). This leads us to a primitive but insightful “general” observation.

To do so, we consider the case where both of null hypothesis \mathbf{X} and alternative hypothesis $\bar{\mathbf{X}}$ are finite mixtures of general sources as follows:

- The null hypothesis is a mixed general source $\mathbf{X} = \{X^n\}_{n=1}^\infty$ consisting of K general (not necessarily memoryless) sources $\mathbf{X}_i = \{X_i^n\}_{n=1}^\infty$ ($i = 1, 2, \dots, K$), that is, $\forall \mathbf{x} \in \mathcal{X}^n$,

$$P_{X^n}(\mathbf{x}) = \sum_{i=1}^K \alpha_i P_{X_i^n}(\mathbf{x}), \quad (72)$$

where $\alpha_i > 0$ ($i = 1, 2, \dots, K$) and $\sum_{i=1}^K \alpha_i = 1$.

- The alternative hypothesis is another mixed general source $\bar{\mathbf{X}} = \{\bar{X}^n\}_{n=1}^\infty$ consisting of L general (not necessarily memoryless) sources $\bar{\mathbf{X}}_j = \{\bar{X}_j^n\}_{n=1}^\infty$ ($j = 1, \dots, L$), that is, $\forall \mathbf{x} \in \mathcal{X}^n$,

$$P_{\bar{X}^n}(\mathbf{x}) = \sum_{j=1}^L \beta_j P_{\bar{X}_j^n}(\mathbf{x}), \quad (73)$$

where $\beta_j > 0$ ($j = 1, 2, \dots, L$) and $\sum_{j=1}^L \beta_j = 1$.

In this general setting, it is hard to derive a compact formula for the first-order ε -optimum exponent (with $0 < \varepsilon < 1$). Instead, we can obtain the following theorem in the special case of $\varepsilon = 0$.

Theorem 6.

$$B_0(\mathbf{X} || \bar{\mathbf{X}}) = \min_{1 \leq i \leq K, 1 \leq j \leq L} B_0(\mathbf{X}_i || \bar{\mathbf{X}}_j). \quad (74)$$

In particular, if \mathbf{X}_i and $\bar{\mathbf{X}}_j$ are all stationary memoryless sources specified by X_i ($i = 1, 2, \dots, K$) and \bar{X}_j ($j = 1, 2, \dots, L$), respectively, then

$$B_0(\mathbf{X} || \bar{\mathbf{X}}) = \min_{1 \leq i \leq K, 1 \leq j \leq L} D(P_{X_i} || P_{\bar{X}_j}), \quad (75)$$

which is a special case of Corollary 1.

Proof. See Appendix F. \square

Furthermore, we can also consider the following exponentially r -optimum exponent in hypothesis testing with two mixed general sources \mathbf{X} and $\bar{\mathbf{X}}$ as above.

Definition 5. Let $r > 0$ be any fixed constant. Rate R is said to be exponentially r -achievable if there exists an acceptance region \mathcal{A}_n such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\mu_n} \geq r, \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n} \geq R. \quad (76)$$

Definition 6 (First-order exponentially r -optimum exponent).

$$B_e(r|\mathbf{X}||\bar{\mathbf{X}}) := \sup\{R | R \text{ is exponentially } r\text{-achievable}\}. \quad (77)$$

Then, it is not difficult to verify that a result analogous to Theorem 6 holds, which is a generalization of [1] (Remark 4.4.3):

Theorem 7.

$$B_e(r|\mathbf{X}||\bar{\mathbf{X}}) = \min_{1 \leq i \leq K, 1 \leq j \leq L} B_e(r|\mathbf{X}_i||\bar{\mathbf{X}}_j). \quad (78)$$

In particular, if the null and alternative hypotheses consist of stationary memoryless sources X_i ($i = 1, 2, \dots, K$) and \bar{X}_j ($j = 1, 2, \dots, L$), respectively, then

$$B_e(r|\mathbf{X}||\bar{\mathbf{X}}) = \min_{1 \leq i \leq K, 1 \leq j \leq L} \inf_{P: D(P||P_{X_i}) < r} D(P||P_{\bar{X}_j}), \quad (79)$$

by virtue of Hoeffding's theorem.

6. Hypothesis Testing with Compound General Sources

In this section, let us consider the compound hypothesis testing problem with finite null hypotheses $\mathbf{X}_i = \{X_i^n\}_{n=1}^\infty$ ($i = 1, 2, \dots, K$) and finite alternative hypotheses $\bar{\mathbf{X}}_j = \{\bar{X}_j^n\}_{n=1}^\infty$ ($j = 1, 2, \dots, L$), where \mathbf{X}_i and $\bar{\mathbf{X}}_j$ are general sources. As is well-known, this problem is expected to have a primitive but “general” relationship to that of mixed hypothesis at the structural level.

Specifically, the compound hypothesis testing is the problem in which a pair of general sources $(\mathbf{X}_i, \bar{\mathbf{X}}_j)$ occurs as a pair (null hypothesis, alternative hypothesis), and the tester does not know which pair $(\mathbf{X}_i, \bar{\mathbf{X}}_j)$ is actually working. This means that the acceptance region \mathcal{A}_n cannot depend on i and j . The type I error probabilities of the compound hypothesis testing are given by

$$\mu_n^{(i)} := \Pr\{X_i^n \notin \mathcal{A}_n\}, \quad (80)$$

for each general null hypothesis \mathbf{X}_i . The type II error probabilities are also given by

$$\lambda_n^{(j)} := \Pr\{\bar{X}_j^n \in \mathcal{A}_n\}, \quad (81)$$

for each general alternative hypothesis $\bar{\mathbf{X}}_j$. Then, the following achievability is of our interest.

Definition 7. Rate R is said to be 0-achievable for the compound hypothesis testing, if there exists an acceptance region \mathcal{A}_n such that

$$\lim_{n \rightarrow \infty} \mu_n^{(i)} = 0 \quad \text{and} \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n^{(j)}} \geq R, \quad (82)$$

for all $i = 1, 2, \dots, K$ and $j = 1, 2, \dots, L$.

Definition 8 (First-order 0-optimum exponent).

$$B(\{\mathbf{X}_i\}_{i=1}^K || \{\bar{\mathbf{X}}_j\}_{j=1}^L) := \sup\{R | R \text{ is 0-achievable}\}. \quad (83)$$

Now, we have

Theorem 8. Assuming that $\alpha_i > 0$ and $\beta_j > 0$ hold for all $i = 1, 2, \dots, K$ and $j = 1, 2, \dots, L$, it holds that

$$B(\{\mathbf{X}_i\}_{i=1}^K || \{\bar{\mathbf{X}}_j\}_{j=1}^L) = B(\{\alpha_i, \mathbf{X}_i\}_{i=1}^K || \{\beta_j, \bar{\mathbf{X}}_j\}_{j=1}^L), \quad (84)$$

where with sources Equations (72) and (73), we use here the notation

$$B(\{\alpha_i, \mathbf{X}_i\}_{i=1}^K || \{\beta_j, \bar{\mathbf{X}}_j\}_{j=1}^L) \quad (85)$$

to denote $B_0(\mathbf{X} || \bar{\mathbf{X}})$ to make explicit dependence on α_i, β_j .

Proof. See Appendix G. \square

From Theorems 6 and 8, we immediately obtain the first-order 0-optimum exponent for the compound hypothesis testing as:

Corollary 2. Assuming that $\alpha_i > 0$ and $\beta_j > 0$ hold for all $i = 1, 2, \dots, K$ and $j = 1, 2, \dots, L$, we have

$$B(\{\mathbf{X}_i\}_{i=1}^K || \{\bar{\mathbf{X}}_j\}_{j=1}^L) = \min_{1 \leq i \leq K, 1 \leq j \leq L} B_0(\mathbf{X}_i || \bar{\mathbf{X}}_j). \quad (86)$$

In particular, if \mathbf{X}_i and $\bar{\mathbf{X}}_j$ are all stationary memoryless sources specified by X_i and \bar{X}_j , respectively, Equation (86) reduces to

$$B(\{\mathbf{X}_i\}_{i=1}^K || \{\bar{\mathbf{X}}_j\}_{j=1}^L) = \min_{1 \leq i \leq K, 1 \leq j \leq L} D(P_{X_i} || P_{\bar{X}_j}). \quad (87)$$

Remark 9. Similar to Definition 5, we can define the exponentially r -optimum exponent also for the compound hypothesis testing problem as follows.

Definition 9. Let $r > 0$ be any fixed constant. Rate R is said to be exponentially r -achievable for the compound hypothesis testing, if there exists an acceptance region \mathcal{A}_n such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\mu_n^{(i)}} \geq r, \quad (88)$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n^{(j)}} \geq R, \quad (89)$$

for all $i = 1, 2, \dots, K$ and $j = 1, 2, \dots, L$.

Definition 10 (First-order exponentially r -optimum exponent).

$$B_e(r | \{\mathbf{X}_i\}_{i=1}^K || \{\bar{\mathbf{X}}_j\}_{j=1}^L) := \sup\{R | R \text{ is exponentially } r\text{-achievable}\}. \quad (90)$$

Then, using an argument similar to the proof of Theorem 8, the following theorem can be shown:

Theorem 9. Let $\alpha_i > 0$ and $\beta_j > 0$ hold for all $i = 1, 2, \dots, K$ and $j = 1, 2, \dots, L$, then it holds that

$$B_e(r | \{\mathbf{X}_i\}_{i=1}^K || \{\bar{\mathbf{X}}_j\}_{j=1}^L) = B_e(r | \{\alpha_i, \mathbf{X}_i\}_{i=1}^K || \{\beta_j, \bar{\mathbf{X}}_j\}_{j=1}^L), \quad (91)$$

where with sources Equations (72) and (73) we use the notation

$$B_e(r|\{\alpha_i, \mathbf{X}_i\}_{i=1}^K|\{\beta_j, \bar{\mathbf{X}}_j\}_{j=1}^L) \quad (92)$$

to denote $B_e(r|\mathbf{X}||\bar{\mathbf{X}})$ (cf. Definitions 5 and 6).

Combining Theorems 7 and 9, we immediately obtain the following corollary:

Corollary 3. Let $\alpha_i > 0$ and $\beta_j > 0$ hold for all $i = 1, 2, \dots, K$ and $j = 1, 2, \dots, L$, then it holds that

$$B_e(r|\{\mathbf{X}_i\}_{i=1}^K|\{\bar{\mathbf{X}}_j\}_{j=1}^L) = \min_{1 \leq i \leq K, 1 \leq j \leq L} B_e(r|\mathbf{X}_i||\bar{\mathbf{X}}_j). \quad (93)$$

In particular, if the null and alternative hypotheses consist of stationary memoryless sources specified by X_i ($i = 1, 2, \dots, K$) and \bar{X}_j ($j = 1, 2, \dots, L$), respectively, as in Theorem 7, then

$$B_e(r|\{\mathbf{X}_i\}_{i=1}^K|\{\bar{\mathbf{X}}_j\}_{j=1}^L) = \min_{1 \leq i \leq K, 1 \leq j \leq L} \inf_{P: D(P||P_{\bar{X}_j}) < r} D(P||P_{X_i}), \quad (94)$$

which corresponds to Equation (79).

7. Concluding Remarks

Thus far, we have investigated the first- and second-order ε -optimum exponents in the hypothesis testing problem. First, we have studied the second-order ε -optimum problem with mixed memoryless null hypothesis and stationary memoryless alternative hypothesis. As we have shown in the analysis of the second-order ε -optimum exponent, we use, as a key property, the asymptotic normality of divergence density rate for each of the component sources. We also observe that the canonical representation, first introduced in [11], is still efficient to express the second-order ε -optimum exponent for mixed memoryless sources in the hypothesis testing problem.

The first-order ε -optimum exponent in the case with mixed memoryless null and alternative hypotheses has also been established. One may wonder whether we can apply the same approach in the derivation of the second-order ε -optimum exponent in this setting. Notice that one of our key techniques to derive the first-order ε -optimum exponent is an expansion $P_{\mathbf{X}}$ around P_{θ} . More careful evaluation of this expansion would be needed to compute the second-order ε -optimum exponent. This remains to be a future work. Our final goal is the problem of hypothesis testing in which both of null and alternative hypotheses are general stationary sources. This paper characterizes the first- and second-order performance of hypothesis testing for mixed memoryless sources as a simple but crucial step toward this goal.

Finally, the relationship between the first-order 0-optimum (respectively, exponentially r -optimum) exponent in the hypothesis testing with mixed general sources and the 0-optimum (respectively, exponentially r -optimum) exponent in the compound hypothesis testing has also been demonstrated.

Acknowledgments: We are grateful to the reviewers for useful comments. In particular, we greatly appreciate the unusually thorough and insightful comments by Reviewer 3, which indeed helped us enhance the quality of the paper. The second author of this work was supported in part by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP26420371.

Author Contributions: Te Sun Han first presented Theorem 4, part of the paper, at IEEE Information Theory Workshop, Jeju, 2015, and subsequently, Te Sun Han and Ryo Nomura discussed together and collaborated to establish Theorem 5.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 2

The proof consists of two parts.

(1) Direct Part:

Set $S_0 = \sup\{S | K(R, S) \leq \varepsilon\}$. Then, we show that $S = S_0 - \gamma$ is (ε, R) -achievable for $\forall \gamma > 0$. Define the acceptance region \mathcal{A}_n as

$$\mathcal{A}_n = \left\{ \frac{1}{n} \log \frac{P_{X^n}(\mathbf{x})}{P_{\bar{X}^n}(\mathbf{x})} > R + \frac{S}{\sqrt{n}} \right\}. \quad (\text{A1})$$

Then, from Lemma 1 with $t = R + \frac{S}{\sqrt{n}}$ we have the upper bound for the type II error probability λ_n :

$$\lambda_n = \Pr \left\{ \bar{X}^n \in \mathcal{A}_n \right\} \leq e^{-nR - \sqrt{n}S}, \quad (\text{A2})$$

from which it follows that

$$\liminf_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \log \frac{1}{\lambda_n e^{nR}} \geq S. \quad (\text{A3})$$

We next evaluate the type I error probability μ_n . Noting that

$$\mu_n = \Pr \{ X^n \notin \mathcal{A}_n \} = \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S}{\sqrt{n}} \right\}, \quad (\text{A4})$$

we have

$$\limsup_{n \rightarrow \infty} \mu_n = \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S}{\sqrt{n}} \right\} \leq \varepsilon, \quad (\text{A5})$$

because $S = S_0 - \gamma$ by the definition. Hence, from Equations (A3) and (A5), $S = S_0 - \gamma$ is (ε, R) -achievable. Since $\gamma > 0$ is arbitrary, the direct part has been proved.

(2) Converse Part:

Suppose that S is (ε, R) -achievable. Then, there exists an acceptance region \mathcal{A}_n such that

$$\limsup_{n \rightarrow \infty} \mu_n \leq \varepsilon \text{ and } \liminf_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \log \frac{1}{\lambda_n e^{nR}} \geq S. \quad (\text{A6})$$

We fix this acceptance region \mathcal{A}_n . The second inequality means that for any $\gamma > 0$

$$\lambda_n \leq e^{-nR - \sqrt{n}(S - \gamma)} \quad (\text{A7})$$

holds for sufficiently large n . On the other hand, from Lemma 2 with $t = R + \frac{S - 2\gamma}{\sqrt{n}}$ it holds that

$$\mu_n + e^{nR + \sqrt{n}(S - 2\gamma)} \lambda_n \geq \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S - 2\gamma}{\sqrt{n}} \right\}. \quad (\text{A8})$$

Substituting Equation (A7) into this inequality, we have

$$\mu_n + e^{-\sqrt{n}\gamma} \geq \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S - 2\gamma}{\sqrt{n}} \right\}, \quad (\text{A9})$$

for sufficiently large n . Thus, we have

$$\limsup_{n \rightarrow \infty} \mu_n \geq \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S - 2\gamma}{\sqrt{n}} \right\}. \quad (\text{A10})$$

Here, from Equation (A6) we have

$$\varepsilon \geq \limsup_{n \rightarrow \infty} \mu_n \geq \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S - 2\gamma}{\sqrt{n}} \right\}, \quad (\text{A11})$$

which means that

$$S - 2\gamma \leq B_\varepsilon(R|\mathbf{X}||\bar{\mathbf{X}}). \quad (\text{A12})$$

Since $\gamma > 0$ is arbitrarily, the proof of the converse part has been completed. \square

Appendix B. Proof of Lemma 4

Since $P_{X_\theta^n}(\mathbf{x}) \leq e^{\frac{1}{\sqrt{n}}} P_{X^n}(\mathbf{x})$ holds for $\forall \theta \in \Theta_n^*$, we have

$$\begin{aligned} \Pr \left\{ \frac{1}{n} \log P_{X_\theta^n}(X_\theta^n) \leq z_n \right\} &\leq \Pr \left\{ \frac{1}{n} \log P_{X_\theta^n}(X_\theta^n) - \frac{1}{\sqrt{n}} \leq z_n \right\} \\ &= \Pr \left\{ \frac{1}{n} \log P_{X_\theta^n}(X_\theta^n) \leq z_n + \frac{1}{\sqrt{n}} \right\} \end{aligned} \quad (\text{A13})$$

for any z_n . By using this inequality with $z_n + \frac{1}{\sqrt{n}} \log P_{\bar{X}^n}(X_\theta^n)$ instead of z_n , we have

$$\Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq z_n \right\} \leq \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq z_n + \frac{1}{\sqrt{n}} \right\} \quad (\text{A14})$$

which completes the proof. \square

Appendix C. Proof of Lemma 5

Setting $\gamma > 0$, we define a set

$$D_n = \left\{ \mathbf{x} \in \mathcal{X}^n \mid \frac{1}{n} \log P_{X_\theta^n}(\mathbf{x}) - \frac{1}{n} \log P_{X^n}(\mathbf{x}) \leq -\frac{\gamma}{\sqrt{n}} \right\}, \quad (\text{A15})$$

for $\theta \in \Theta$. Then, it holds that

$$\begin{aligned} \Pr \{X_\theta^n \in D_n\} &= \sum_{\mathbf{x} \in D_n} P_{X_\theta^n}(\mathbf{x}) \\ &\leq \sum_{\mathbf{x} \in D_n} P_{X^n}(\mathbf{x}) e^{-\sqrt{n}\gamma} \\ &\leq e^{-\sqrt{n}\gamma}. \end{aligned} \quad (\text{A16})$$

Thus, for any real number z_n it holds that

$$\begin{aligned} &\Pr \left\{ \frac{1}{n} \log P_{X_\theta^n}(X_\theta^n) \leq z_n - \frac{\gamma}{\sqrt{n}} \right\} \\ &= \Pr \left\{ \frac{1}{n} \log P_{X_\theta^n}(X_\theta^n) \leq z_n - \frac{\gamma}{\sqrt{n}}, X_\theta^n \notin D_n \right\} + \Pr \left\{ \frac{1}{n} \log P_{X_\theta^n}(X_\theta^n) \leq z_n - \frac{\gamma}{\sqrt{n}}, X_\theta^n \in D_n \right\} \\ &\leq \Pr \left\{ \frac{1}{n} \log P_{X^n}(X_\theta^n) \leq z_n \right\} + \Pr \{X_\theta^n \in D_n\} \\ &\leq \Pr \left\{ \frac{1}{n} \log P_{X^n}(X_\theta^n) \leq z_n \right\} + e^{-\sqrt{n}\gamma}. \end{aligned} \quad (\text{A17})$$

Hence, we obtain the inequality

$$\Pr \left\{ \frac{1}{n} \log P_{X^n}(X_\theta^n) \leq z_n \right\} \geq \Pr \left\{ \frac{1}{n} \log P_{X_\theta^n}(X_\theta^n) \leq z_n - \frac{\gamma}{\sqrt{n}} \right\} - e^{-\sqrt{n}\gamma}, \quad (\text{A18})$$

from which with $z_n + \frac{1}{n} \log P_{\bar{X}^n}(X_\theta^n)$ instead of z_n it follows that

$$\Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq z_n \right\} \geq \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq z_n - \frac{\gamma}{\sqrt{n}} \right\} - e^{-\sqrt{n}\gamma} \quad (\text{A19})$$

for all $\theta \in \Theta$. This completes the proof. \square

Appendix D. Proof of Theorem 4

Setting

$$\bar{B}_\varepsilon(R, S) := \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) < R\}} dw(\theta) + \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) = R\}} \Phi_\theta(S) dw(\theta), \quad (\text{A20})$$

it suffices, in view of Theorem 2, to show two inequalities:

$$\bar{B}_\varepsilon(R, S) \geq \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S}{\sqrt{n}} \right\}, \quad (\text{A21})$$

$$\bar{B}_\varepsilon(R, S) \leq \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S}{\sqrt{n}} \right\}. \quad (\text{A22})$$

- *Proof of Equation (A21):*

By the definitions of \mathbf{X} and $\bar{\mathbf{X}}$, it holds that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S}{\sqrt{n}} \right\} \\ &= \limsup_{n \rightarrow \infty} \int_{\Theta} dw(\theta) \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} \right\} \\ &\leq \limsup_{n \rightarrow \infty} \int_{\Theta_n^*} dw(\theta) \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} \right\} + \limsup_{n \rightarrow \infty} \int_{\Theta - \Theta_n^*} dw(\theta) \\ &= \limsup_{n \rightarrow \infty} \int_{\Theta_n^*} dw(\theta) \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} \right\} \\ &\leq \limsup_{n \rightarrow \infty} \int_{\Theta_n^*} dw(\theta) \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}} \right\} \\ &\leq \limsup_{n \rightarrow \infty} \int_{\Theta} dw(\theta) \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}} \right\} \\ &\leq \int_{\Theta} dw(\theta) \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}} \right\}, \end{aligned} \quad (\text{A23})$$

where the second equality and the second inequality are due to Lemmas 3 and 4, respectively, and the last inequality is from the reverse Fatou's lemma.

Here, we define three sets:

$$\Theta_0 := \{\theta \in \Theta \mid D(P_{X_\theta} \| P_{\bar{X}}) = R\}, \quad (\text{A24})$$

$$\Theta_1 := \{\theta \in \Theta \mid D(P_{X_\theta} \| P_{\bar{X}}) < R\}, \quad (\text{A25})$$

$$\Theta_2 := \{\theta \in \Theta \mid D(P_{X_\theta} \| P_{\bar{X}}) > R\}. \quad (\text{A26})$$

Noting that, setting $X_\theta^n = (X_{\theta,1}, X_{\theta,2}, \dots, X_{\theta,n})$,

$$\frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} = \frac{1}{n} \sum_{i=1}^n \log \frac{P_{X_\theta}(X_{\theta,i})}{P_{\bar{X}}(X_{\theta,i})} \quad (\text{A27})$$

gives the arithmetic average of n i.i.d. variables with expectation

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \log \frac{P_{X_\theta}(X_{\theta,i})}{P_{\bar{X}}(X_{\theta,i})} \right] = D(P_{X_\theta} \| P_{\bar{X}}). \quad (\text{A28})$$

Then, the weak law of large numbers yields that for $\forall \theta \in \Theta_2$

$$\limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}} \right\} = 0. \quad (\text{A29})$$

Moreover, for $\forall \theta \in \Theta_0$, the central limit theorem leads to

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}} \right\} \\ &= \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{\sqrt{n}} \left(\log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} - \sqrt{n} D(P_{X_\theta} \| P_{\bar{X}}) \right) \leq S + \frac{1}{\sqrt[4]{n}} \right\} \\ &= \Phi_\theta(S). \end{aligned} \quad (\text{A30})$$

Summarizing these equalities, we obtain

$$\begin{aligned} & \int_{\Theta} dw(\theta) \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}} \right\} \\ &= \int_{\Theta_1} dw(\theta) \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}} \right\} \\ &+ \int_{\Theta_0} dw(\theta) \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} + \frac{1}{\sqrt[4]{n^3}} \right\} \\ &\leq \int_{\{\theta \mid D(P_{X_\theta} \| P_{\bar{X}}) < R\}} dw(\theta) + \int_{\{\theta \mid D(P_{X_\theta} \| P_{\bar{X}}) = R\}} \Phi_\theta(S) dw(\theta). \end{aligned} \quad (\text{A31})$$

Plugging Equation (A31) into Equation (A23) yields Equation (A21).

- *Proof of Equation (A22):*

By the definitions of \mathbf{X} and $\bar{\mathbf{X}}$, and Lemma 5 with $z_n = R + \frac{S}{\sqrt{n}}$, it holds that

$$\begin{aligned}
 & \limsup_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\bar{X}^n}(X^n)} \leq R + \frac{S}{\sqrt{n}} \right\} \\
 & \geq \liminf_{n \rightarrow \infty} \int_{\Theta} dw(\theta) \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} \right\} \\
 & \geq \liminf_{n \rightarrow \infty} \int_{\Theta} dw(\theta) \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S}{\sqrt{n}} - \frac{\gamma}{\sqrt{n}} \right\} \\
 & = \liminf_{n \rightarrow \infty} \int_{\Theta} dw(\theta) \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S - \gamma}{\sqrt{n}} \right\} \\
 & \geq \int_{\Theta} dw(\theta) \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S - \gamma}{\sqrt{n}} \right\}, \tag{A32}
 \end{aligned}$$

for any $\gamma > 0$, where the last inequality is due to Fatou's lemma. We also partition the parameter space Θ into three sets as in Equations (A24)–(A26).

Then, similarly to the derivation of Equations (A29) and (A30), we obtain

$$\liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S - \gamma}{\sqrt{n}} \right\} = \begin{cases} \Phi_\theta(S - \gamma), & \theta \in \Theta_0 \\ 1, & \theta \in \Theta_1 \end{cases} \tag{A33}$$

Thus, the right-hand side of Equation (A32) is rewritten as

$$\begin{aligned}
 & \int_{\Theta} dw(\theta) \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S - \gamma}{\sqrt{n}} \right\} \\
 & \geq \int_{\Theta_1} dw(\theta) \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S - \gamma}{\sqrt{n}} \right\} \\
 & \quad + \int_{\Theta_0} dw(\theta) \liminf_{n \rightarrow \infty} \Pr \left\{ \frac{1}{n} \log \frac{P_{X_\theta^n}(X_\theta^n)}{P_{\bar{X}^n}(X_\theta^n)} \leq R + \frac{S - \gamma}{\sqrt{n}} \right\} \\
 & = \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) < R\}} dw(\theta) + \int_{\{\theta | D(P_{X_\theta} || P_{\bar{X}}) = R\}} \Phi_\theta(S - \gamma) dw(\theta). \tag{A34}
 \end{aligned}$$

Substituting Equation (A34) into Equation (A32) and noting that $\gamma > 0$ is arbitrary, we obtain Equation (A22). \square

Appendix E. Proofs of Equations (45) and (46)

(1) Proof of Equation (45):

To prove Equation (45), we define $a(\mathbf{x})$ as

$$a(\mathbf{x}) := v\text{-ess. sup } \bar{P}_\sigma^n(\mathbf{x}), \tag{A35}$$

where $v\text{-ess. sup } f_\sigma$ denotes the essential supremum of f_σ with respect to $v(\sigma)$, i.e., $v\text{-ess. sup } f_\sigma := \inf\{\alpha | \Pr\{f_\sigma > \alpha\} = 0\}$. Thus, from the property of the essential supremum we immediately have

$$a(\mathbf{x}) \geq P_{\bar{X}^n}(\mathbf{x}), \tag{A36}$$

for $\forall n = 1, 2, \dots$.

Let $P_{\mathbf{x}}$ denote the type of $\mathbf{x} \in \mathcal{T}_{\theta, \nu}^n$. Then, noting that

$$\begin{aligned}\bar{P}_{\sigma}^n(\mathbf{x}) &= \prod_{x \in \mathcal{X}} \bar{P}_{\sigma}(x)^{N(x|\mathbf{x})} \\ &= \exp \left[\sum_{x \in \mathcal{X}} N(x|\mathbf{x}) \log \bar{P}_{\sigma}(x) \right] \\ &= \exp \left[-n \left(H(P_{\mathbf{x}}) + D(P_{\mathbf{x}}||\bar{P}_{\sigma}) \right) \right]\end{aligned}\quad (\text{A37})$$

holds, $a(\mathbf{x})$ is written as

$$\begin{aligned}a(\mathbf{x}) &= \exp \left[-n \left(H(P_{\mathbf{x}}) + v\text{-ess. inf } D(P_{\mathbf{x}}||\bar{P}_{\sigma}) \right) \right] \\ &= \exp \left[-n \left(H(P_{\mathbf{x}}) + D(P_{\mathbf{x}}||\bar{P}_{\sigma(P_{\mathbf{x}})}) \right) \right].\end{aligned}\quad (\text{A38})$$

Here, it is important to notice that $D(P||\bar{P}_{\sigma})$ is continuous in (P, \bar{P}_{σ}) and hence, owing to the assumption, $D(P||\bar{P}_{\sigma(Q)})$ is continuous in $Q \in \mathcal{P}(\mathcal{X})$. Thus, expanding $D(P||\bar{P}_{\sigma(P_{\mathbf{x}})})$ in $P_{\mathbf{x}}$ around P_{θ} leads to

$$D(P||\bar{P}_{\sigma(P_{\mathbf{x}})}) = D(P||\bar{P}_{\sigma(P_{\theta})}) + \delta_{\theta}(\nu) \quad (\mathbf{x} \in \mathcal{T}_{\theta, \nu}^n). \quad (\text{A39})$$

with some $\delta_{\theta}(\nu)$ such that $\delta_{\theta}(\nu) \rightarrow 0$ as $\nu \rightarrow 0$, because $\sum_{x \in \mathcal{X}} |P_{\theta}(x) - P_{\mathbf{x}}(x)| \leq \nu$ for $\mathbf{x} \in \mathcal{T}_{\theta, \nu}^n$.

Then, with $P_{\mathbf{x}}$ instead of P in Equation (A39) and in view of Equation (A36) for each $\mathbf{x} \in \mathcal{T}_{\theta, \nu}^n$ we have the upper bound:

$$\begin{aligned}P_{\bar{X}^n}(\mathbf{x}) &\leq a(\mathbf{x}) \\ &= \exp \left[-n \left(H(P_{\mathbf{x}}) + D(P_{\mathbf{x}}||\bar{P}_{\sigma(P_{\mathbf{x}})}) \right) \right] \\ &= \exp \left[-n \left(H(P_{\mathbf{x}}) + D(P_{\mathbf{x}}||\bar{P}_{\sigma(P_{\theta})}) - \delta_{\theta}(\nu) \right) \right] \\ &= \bar{P}_{\sigma(P_{\theta})}^n(\mathbf{x}) \exp[n\delta_{\theta}(\nu)],\end{aligned}\quad (\text{A40})$$

from which it follows that for each $\mathbf{x} \in \mathcal{T}_{\theta, \nu}^n$

$$\frac{1}{n} \log \frac{1}{P_{\bar{X}^n}(\mathbf{x})} \geq \frac{1}{n} \log \frac{1}{\bar{P}_{\sigma(P_{\theta})}^n(\mathbf{x})} - \delta_{\theta}(\nu). \quad (\text{A41})$$

Therefore, the proof of Equation (45) has been completed.

(2) *Proof of Equation (46):*

To prove Equation (46), we show the lower bound of $P_{\bar{X}^n}(\mathbf{x})$. For any $P \in \mathcal{P}(\mathcal{X})$ and any small constant $\tau > 0$, set

$$\mathcal{S}_{\tau}(P) := \left\{ \sigma \in \Sigma \mid D(P||\bar{P}_{\sigma}) < D(P||\bar{P}_{\sigma(P)}) + \tau \right\}, \quad (\text{A42})$$

then, by the definition of $v\text{-ess. inf}$,

$$c_{\tau}(P) := \int_{\mathcal{S}_{\tau}(P)} dv(\sigma) > 0 \quad (\text{A43})$$

holds. Our claim is that for any $\theta \in \Theta$ and sufficiently small $\tau > 0$ and with some positive constant $c_{\theta} > 0$

$$\inf_{\mathbf{x} \in \mathcal{T}_{\theta, \nu}^n} c_{\tau}(P_{\mathbf{x}}) \geq c_{\theta}. \quad (\text{A44})$$

To see this, consider a sequence $\{\tau_i\}_{i=1}^{\infty}$ such that $0 < \tau_1 < \tau_2 < \dots \rightarrow \tau$. Then, there exists a positive integer m such that $c_{\tau_m}(P_\theta) > 0$. Otherwise, the continuity of probability measure implies that

$$0 = \lim_{i \rightarrow \infty} c_{\tau_i}(P_\theta) = c_\tau(P_\theta) > 0, \quad (\text{A45})$$

which is a contradiction. On the other hand, in view of Equation (A42), $\sigma \in \mathcal{S}_{\tau_m}(P_\theta)$ is equivalent to

$$D(P_\theta || \bar{P}_\sigma) < D(P_\theta || \bar{P}_{\sigma(P_\theta)}) + \tau_m, \quad (\text{A46})$$

$$D(P_x || \bar{P}_\sigma) < D(P_x || \bar{P}_{\sigma(P_x)}) + \tau_m + \gamma(\nu) \quad (\forall x \in \mathcal{T}_{\theta, \nu}^n), \quad (\text{A47})$$

where Equation (A47) follows from Equation (A46) by expanding P_θ around P_x with some $\gamma(\nu) > 0$ such that $\gamma(\nu) \rightarrow 0$ as $\nu \rightarrow 0$. Therefore, all $\sigma \in \mathcal{S}_{\tau_m}(P_\theta)$ satisfy Equation (A47). Now we can take $\nu > 0$ so that $\tau_m + \gamma(\nu) < \tau$ to have

$$D(P_x || \bar{P}_\sigma) < D(P_x || \bar{P}_{\sigma(P_x)}) + \tau \quad (\forall x \in \mathcal{T}_{\theta, \nu}^n). \quad (\text{A48})$$

Therefore, $\mathcal{S}_{\tau_m}(P_\theta) \subset \mathcal{S}_\tau(P_x)$. Hence, we have

$$0 < c_{\tau_m}(P_\theta) \leq c_\tau(P_x) \quad (\forall x \in \mathcal{T}_{\theta, \nu}^n). \quad (\text{A49})$$

This is nothing but Equation (A44).

Thus, again for $\forall x \in \mathcal{T}_{\theta, \nu}^n$, we have the lower bound

$$\begin{aligned} P_{\bar{X}}^n(x) &= \int_{\Sigma} \bar{P}_\sigma^n(x) dv(\sigma) \\ &\geq \int_{\mathcal{S}_\tau(P_x)} \bar{P}_\sigma^n(x) dv(\sigma) \\ &= \int_{\mathcal{S}_\tau(P_x)} \exp[-n(H(P_x) + D(P_x || \bar{P}_\sigma))] dv(\sigma) \\ &\geq \int_{\mathcal{S}_\tau(P_x)} \exp[-n(H(P_x) + D(P_x || \bar{P}_{\sigma(P_x)}) + \tau)] dv(\sigma) \\ &= c_\tau(P_x) \exp[-n(H(P_x) + D(P_x || \bar{P}_{\sigma(P_\theta)}) + \tau - \delta_\theta(\nu))] \\ &\geq c_{\tau_m}(P_\theta) \bar{P}_{\sigma(P_\theta)}^n(x) \exp[n(\delta_\theta(\nu) - \tau)], \end{aligned} \quad (\text{A50})$$

where in the second last equality and in the last inequality we have used the continuity of $D(P_x || \bar{P}_{\sigma(P_x)})$ in P_x around P_θ and Equation (A49), respectively. From Equation (A50), we obtain

$$\frac{1}{n} \log \frac{1}{P_{\bar{X}}^n(x)} \leq \frac{1}{n} \log \frac{1}{\bar{P}_{\sigma(P_\theta)}^n(x)} + \frac{1}{n} \log \frac{1}{c_{\tau_m}(P_\theta)} + (\tau - \delta_\theta(\nu)), \quad (\text{A51})$$

for each $x \in \mathcal{T}_{\theta, \nu}^n$, which completes the proof of Equation (46). \square

Appendix F. Proof of Theorem 6

First, we prove the inequality:

$$B_0(\mathbf{X} || \bar{\mathbf{X}}) \geq \min_{1 \leq i \leq K, 1 \leq j \leq L} B_0(\mathbf{X}_i || \bar{\mathbf{X}}_j). \quad (\text{A52})$$

To do so, we arbitrarily fix R_{ij} for $1 \leq i \leq K, 1 \leq j \leq L$ so that

$$R_{ij} < B_0(\mathbf{X}_i || \bar{\mathbf{X}}_j). \quad (\text{A53})$$

Then, by the definition of $B_0(\mathbf{X}_i||\bar{\mathbf{X}}_j)$, there exists an acceptance region $\mathcal{A}_n^{(i,j)}$ satisfying

$$\lim_{n \rightarrow \infty} \mu_n^{(i,j)} = 0, \quad (\text{A54})$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n^{(i,j)}} \geq R_{ij}, \quad (\text{A55})$$

where $\mu_n^{(i,j)}$ and $\lambda_n^{(i,j)}$ are defined respectively as

$$\mu_n^{(i,j)} := \Pr \left\{ X_i^n \notin \mathcal{A}_n^{(i,j)} \right\}, \quad \lambda_n^{(i,j)} := \Pr \left\{ \bar{X}_j^n \in \mathcal{A}_n^{(i,j)} \right\}. \quad (\text{A56})$$

By using these regions, we define the acceptance region \mathcal{A}_n as

$$\mathcal{A}_n := \bigcup_{i=1}^K \left(\bigcap_{j=1}^L \mathcal{A}_n^{(i,j)} \right). \quad (\text{A57})$$

Then, we have

$$\begin{aligned} \mu_n &= \Pr \{ X^n \notin \mathcal{A}_n \} = \sum_{i=1}^K \alpha_i \Pr \left\{ X_i^n \notin \left(\bigcup_{i'=1}^K \bigcap_{j=1}^L \mathcal{A}_n^{(i',j)} \right) \right\} \\ &\leq \sum_{i=1}^K \alpha_i \Pr \left\{ X_i^n \notin \left(\bigcap_{j=1}^L \mathcal{A}_n^{(i,j)} \right) \right\} \\ &\leq \sum_{i=1}^K \sum_{j=1}^L \alpha_i \Pr \left\{ X_i^n \notin \left(\mathcal{A}_n^{(i,j)} \right) \right\} \\ &= \sum_{i=1}^K \sum_{j=1}^L \alpha_i \mu_n^{(i,j)}, \end{aligned} \quad (\text{A58})$$

from which, together with Equation (A54), we obtain

$$\lim_{n \rightarrow \infty} \mu_n = 0. \quad (\text{A59})$$

Similarly, we have

$$\lambda_n = \Pr \left\{ \bar{X}^n \in \mathcal{A}_n \right\} \leq \sum_{j=1}^L \sum_{i=1}^K \beta_j \lambda_n^{(i,j)}, \quad (\text{A60})$$

from which, together with Equation (A55), we obtain for any small $\gamma > 0$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n} \geq \min_{1 \leq i \leq K, 1 \leq j \leq L} R_{ij} - \gamma. \quad (\text{A61})$$

Since R_{ij} are arbitrary as far as Equation (A53) is satisfied, we have Equation (A52).

Next, we prove the inequality:

$$B_0(\mathbf{X}||\bar{\mathbf{X}}) \leq \min_{1 \leq i \leq K, 1 \leq j \leq L} B_0(\mathbf{X}_i||\bar{\mathbf{X}}_j). \quad (\text{A62})$$

To do so, let R be 0-achievable, then there exists an acceptance region \mathcal{A}_n satisfying

$$\lim_{n \rightarrow \infty} \mu_n = 0, \quad (\text{A63})$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n} \geq R. \quad (\text{A64})$$

We fix such an \mathcal{A}_n and consider the hypothesis testing with null hypothesis \mathbf{X}_i and alternative hypothesis $\bar{\mathbf{X}}_j$ for arbitrarily fixed i and j . Then, probabilities of type I error and type II error are given by

$$\mu_n^{(i,j)} = \Pr \{X_i^n \notin \mathcal{A}_n\}, \quad (\text{A65})$$

$$\lambda_n^{(i,j)} = \Pr \{\bar{X}_j^n \in \mathcal{A}_n\}. \quad (\text{A66})$$

Since

$$\begin{aligned} \mu_n &= \sum_{i=1}^K \alpha_i \Pr \{X_i^n \notin \mathcal{A}_n\} \\ &= \sum_{i=1}^K \alpha_i \mu_n^{(i,j)}, \end{aligned} \quad (\text{A67})$$

we have

$$\mu_n^{(i,j)} \leq \frac{\mu_n}{\alpha_i}. \quad (\text{A68})$$

From this inequality and Equation (A63) we obtain

$$\lim_{n \rightarrow \infty} \mu_n^{(i,j)} = 0. \quad (\text{A69})$$

Similar to the derivation of Equation (A68), we have

$$\lambda_n^{(i,j)} \leq \frac{\lambda_n}{\beta_j}. \quad (\text{A70})$$

Hence, from Equation (A64) we obtain

$$\begin{aligned} R &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n} \\ &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n^{(i,j)}} + \limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\beta_j} \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n^{(i,j)}}. \end{aligned} \quad (\text{A71})$$

From Equations (A69) and (A71), it follows that R is 0-achievable for the hypothesis testing with \mathbf{X}_i against $\bar{\mathbf{X}}_j$. Noting that i, j are arbitrary with $1 \leq i \leq K$ and $1 \leq j \leq L$, we obtain

$$R \leq \min_{1 \leq i \leq K, 1 \leq j \leq L} B_0(\mathbf{X}_i || \bar{\mathbf{X}}_j). \quad (\text{A72})$$

This means that Equation (A62) holds, completing the proof of Theorem 6. \square

Appendix G. Proof of Theorem 8

It suffices to show two inequalities:

$$B(\{\mathbf{X}_i\}_{i=1}^K || \{\bar{\mathbf{X}}_j\}_{j=1}^L) \leq B(\{\alpha_i, \mathbf{X}_i\}_{i=1}^K || \{\beta_j, \bar{\mathbf{X}}_j\}_{j=1}^L), \quad (\text{A73})$$

$$B(\{\mathbf{X}_i\}_{i=1}^K || \{\bar{\mathbf{X}}_j\}_{j=1}^L) \geq B(\{\alpha_i, \mathbf{X}_i\}_{i=1}^K || \{\beta_j, \bar{\mathbf{X}}_j\}_{j=1}^L). \quad (\text{A74})$$

- *Proof of Equation (A73):*

Suppose that R is 0-achievable for the compound hypothesis testing, that is, there exists an acceptance region \mathcal{A}_n such that

$$\lim_{n \rightarrow \infty} \mu_n^{(i)} = 0 \quad (i = 1, 2, \dots, K), \quad (\text{A75})$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n^{(j)}} \geq R \quad (j = 1, 2, \dots, L). \quad (\text{A76})$$

Then, the type I error probability μ_n for the hypothesis testing with mixed general sources is evaluated as follows. By the definition of μ_n and Equation (72), we have

$$\begin{aligned} \mu_n &= \Pr \{X^n \notin \mathcal{A}_n\} \\ &= \sum_{i=1}^K \alpha_i \Pr \{X_i^n \notin \mathcal{A}_n\} \\ &= \sum_{i=1}^K \alpha_i \mu_n^{(i)}, \end{aligned} \quad (\text{A77})$$

from which, together with Equation (A75), we obtain

$$\lim_{n \rightarrow \infty} \mu_n = 0. \quad (\text{A78})$$

Similarly, we have

$$\begin{aligned} \lambda_n &= \Pr \{\bar{X}^n \in \mathcal{A}_n\} \\ &= \sum_{j=1}^L \beta_j \Pr \{\bar{X}_j^n \in \mathcal{A}_n\} \\ &= \sum_{j=1}^L \beta_j \lambda_n^{(j)}. \end{aligned} \quad (\text{A79})$$

On the other hand, Equation (A76) implies

$$\lambda_n^{(j)} \leq e^{-n(R-\gamma)} \quad (n \geq n_0), \quad (\text{A80})$$

holds for any $\gamma > 0$ and all $j = 1, 2, \dots, L$. Substituting this inequality into Equation (A79) yields

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n} \geq R - \gamma. \quad (\text{A81})$$

Since $\gamma > 0$ is arbitrary, from Equations (A78) and (A81) we conclude that Equation (A73) holds.

- Proof of Equation (A74):

Suppose that R is 0-achievable for the mixed hypothesis testing, that is, there exists an acceptance region \mathcal{A}_n such that

$$\lim_{n \rightarrow \infty} \mu_n = 0, \quad (\text{A82})$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n} \geq R. \quad (\text{A83})$$

We fix such an \mathcal{A}_n and set

$$\mu_n^{(i)} = \Pr \{X_i^n \notin \mathcal{A}_n\}, \quad (\text{A84})$$

$$\lambda_n^{(j)} = \Pr \{\bar{X}_j^n \in \mathcal{A}_n\}. \quad (\text{A85})$$

Then, from Equation (72) we have

$$\begin{aligned} \mu_n &= \sum_{i=1}^K \alpha_i \Pr \{X_i^n \notin \mathcal{A}_n\} \\ &= \sum_{i=1}^K \alpha_i \mu_n^{(i)}, \end{aligned} \quad (\text{A86})$$

from which, it follows that

$$\mu_n^{(i)} \leq \frac{\mu_n}{\alpha_i} \quad (\text{A87})$$

for all $i = 1, 2, \dots, K$. From this inequality and Equation (A82), we obtain

$$\lim_{n \rightarrow \infty} \mu_n^{(i)} = 0, \quad (\text{A88})$$

for all $i = 1, 2, \dots, K$. Similarly,

$$\begin{aligned} \lambda_n &= \sum_{j=1}^L \beta_j \Pr \{\bar{X}_j^n \in \mathcal{A}_n\} \\ &= \sum_{j=1}^L \beta_j \lambda_n^{(j)}, \end{aligned} \quad (\text{A89})$$

so that we have for $j = 1, 2, \dots, L$,

$$\lambda_n^{(j)} \leq \frac{\lambda_n}{\beta_j}, \quad (\text{A90})$$

which means that

$$\begin{aligned} \frac{1}{n} \log \frac{1}{\lambda_n^{(j)}} &\geq \frac{1}{n} \log \frac{\beta_j}{\lambda_n} \\ &= \frac{1}{n} \log \frac{1}{\lambda_n} - \frac{1}{n} \log \frac{1}{\beta_j}. \end{aligned} \quad (\text{A91})$$

Noting that β_j ($j = 1, 2, \dots, L$) are constants, from Equation (A83) we obtain

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\lambda_n^{(j)}} \geq R, \quad (\text{A92})$$

for all $j = 1, 2, \dots, L$. From Equations (A88) and (A92), we conclude that Equation (A74) holds. \square

References

1. Han, T.S. *Information-Spectrum Methods in Information Theory*; Springer: New York, NY, USA, 2003.
2. Dembo, A.; Zeitouni, O. *Large Deviations Techniques and Applications*; Jones and Bartlett Publishers: Boston, MA, USA, 1993.
3. Chen, P.N. General formulas for the Neyman-Pearson type-II error exponent subject to fixed and exponential type-I error bound. *IEEE Trans. Inf. Theory* **1996**, *42*, 316–323.
4. Strassen, V. Asymptotische Abschätzungen in Shannon's Informations Theorie. In Proceedings of the Transactions of the 3rd Prague Conference on Information Theory, Prague, Czech Republic, 5–13 June 1962; pp. 687–723.
5. Hayashi, M. Information Spectrum Approach to Second-Order Coding Rate in Channel Coding. *IEEE Trans. Inf. Theory* **2009**, *55*, 4947–4966.
6. Polyanskiy, Y.; Poor, H.; Verdú, S. Channel Coding Rate in the Finite Blocklength Regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359.
7. Nomura, R.; Han, T.S. Second-Order Resolvability, Intrinsic Randomness, and Fixed-Length Source Coding for Mixed Sources: Information Spectrum Approach. *IEEE Trans. Inf. Theory* **2013**, *59*, 1–16.
8. Tan, V.; Kosut, O. On the Dispersion of Three Network Information Theory Problems. *IEEE Trans. Inf. Theory* **2014**, *60*, 881–903.
9. Watanabe, S. Second-Order Region for Gray–Wyner Network. *IEEE Trans. Inf. Theory* **2017**, *63*, 1006–1018.
10. Polyanskiy, Y.; Poor, H.V.; Verdú, S. Dispersion of the Gilbert-Elliott Channel. *IEEE Trans. Inf. Theory* **2011**, *57*, 1829–1848.
11. Nomura, R.; Han, T.S. Second-Order Slepian-Wolf Coding Theorems for Non-Mixed and Mixed Sources. *IEEE Trans. Inf. Theory* **2014**, *60*, 5553–5572.
12. Yagi, H.; Han, T.S.; Nomura, R. First- and Second-Order Coding Theorems for Mixed Memoryless Channels with General Mixture. *IEEE Trans. Inf. Theory* **2016**, *62*, 4395–4412.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).