A Spectral Representation Of The Simple Hypothesis Testing Problem

Barış Nakiboğlu, *Member, IEEE,*

Canım anneannem Fatma Canbaz'ın anısına adanmıştır.
Dedicated to the memory of my dear grandmother Fatma Canbaz.

arXiv:2601.09564v3 [cs.IT] 7 Nov 2026

***Abstract*—The minimum Type II error probability (or volume) of simple hypothesis tests with randomized detectors is expressed as a Riemann integral of the (cumulative) distribution function of the likelihood ratio for all non-negative Type I error probability (or volume) values: $\beta(\varepsilon)=\int_0^\infty |F(\imath/\tau)-\varepsilon|^+ d\tau$ for all $\varepsilon\geq 0$. The derivation relies on convex conjugation (the Legendre transform) and level set integration, and is sufficiently general to extend to tests between σ-finite measures. Measure change identities relating the $\beta(\cdot)$ functions of different hypothesis testing problems are established. Approximations for Type II and Type I volumes are derived under two different hypotheses concerning the Kolmogorov distance to normality for the log-likelihood ratio distributions. In both the Central Limit Theorem and the large deviations regimes, the resulting non-asymptotic expressions recover and extend state-of-the-art characterizations of optimal performance in the memoryless case, and improve and generalize them in the Markovian case. Finally, The distinction between the implications of the Central Limit Theorem and the Berry–Esseen Theorem for the asymptotic simple hypothesis testing problem is clarified.**

Contents

I Introduction 1
- I-A Prior Work On $\beta(\cdot|^q_w)$, $F^\vee_\ell(\cdot|^w_q)$, and $\hbar(\cdot|^q_w)$ 3
- I-B Prior Work On Approximations 5
- I-C Main Contributions 6
- I-D Outline 7

II Preliminaries 7
- II-A Convex Functions on the Real Line . . 7
- II-B The Semicontinuous Inverses 7
- II-C The Likelihood-Ratio Quantiles 8
- II-D The Primitive Entropy Spectrum 9

III Spectral Representations and Measure Change 10
- III-A Likelihood-Ratio Tests 11
- III-B A Spectral Neyman–Pearson Lemma . . 11
- III-C Measure Change Identities 13

IV Approximations for randomized-NP-function 14
- IV-A The Log-Likelihood-Ratio 15
- IV-B Gaussian Mills Ratio 15
- IV-C Kolmogorov Distance to Normality . . . 16

V Applications 18
- V-A Memoryless Case 18
- V-B Discrete Markov Case 19
- V-C Asymptotic Case 23

VI Discussion 25

Appendix 26
- A Proofs Of The Lemmas About LRQs . 26
- B Proofs Of The Lemmas About PES . . 27
- C Proofs Of 1st-Order Stein Lemmas . . . 27
- D Proof of Lemma 9 28

References 28

B. Nakiboğlu is with the Department of Electrical and Electronics Engineering Middle East Technical University 06800 Ankara, Türkiye (0000-0001-7737-5423).

This paper was presented in part at the 14th Iran Workshop on Communication and Information Theory, April 2026, and at the IEEE International Symposium on Information Theory, (ISIT) Guangzhou, China, July 2026 [DOI:10.1109/ISIT62367.2026.11653959].

I. Introduction

The following elementary decision problem illustrates our principal novelty regarding the Neyman–Pearson lemma. Let Y be a zero mean unit variance Gaussian random variable. What is the infimum of the lengths of Borel sets $\mathcal{E}$ satisfying $\mathbf{P}(\mathsf{Y}\notin\mathcal{E})\leq\varepsilon$, i.e, what is the value of $\beta(\varepsilon)$ defined in the following?

$$\beta(\varepsilon):=\inf\left\{\int_{\mathcal{E}}\mathrm{d}y : \mathcal{E}\in\mathcal{B} \text{ and } \int_{\mathbb{R}\setminus\mathcal{E}}\varphi(y)\mathrm{d}y\leq\varepsilon\right\} \quad \forall\varepsilon\in[0,1],$$

where $\mathcal{B}$ is the set of all Borel subsets of $\mathbb{R}$ and $\varphi(\cdot)$ is the Gaussian probability density function:

$$\varphi(y):=\tfrac{1}{\sqrt{2\pi}}e^{-y^2/2} \qquad \forall y\in\mathbb{R}. \tag{1}$$

The answer is intuitively obvious: we first determine the threshold $\phi(\varepsilon)$ for $\varphi(\cdot)$ such that $\mathbf{P}(\varphi(\mathsf{Y})\leq\phi(\varepsilon))=\varepsilon$, then $\beta(\varepsilon)$ is the length of the set of all y's satisfying $\varphi(y)>\phi(\varepsilon)$. The threshold $\phi(\varepsilon)$ is the ε-quantile for the random variable $\varphi(\mathsf{Y})$, i.e., $\phi(\varepsilon)=F^{-1}_{\varphi(\mathsf{Y})}(\varepsilon)$.

$$\begin{aligned} F^{-1}_{\varphi(\mathsf{Y})}(\varepsilon)&:=\inf\{\tau : F_{\varphi(\mathsf{Y})}(\tau)\geq\varepsilon\} && \forall\varepsilon\in(0,1),\\ &=\varphi\big(Q^{-1}\big(\tfrac{\varepsilon}{2}\big)\big) && \forall\varepsilon\in(0,1), \end{aligned} \tag{2}$$

where[1] $F_{\varphi(\mathsf{Y})}(\cdot)$ is the cumulative distribution function of the random variable $\varphi(\mathsf{Y})$ and $Q^{-1}(\cdot)$ is the inverse of the Gaussian Q-function defined as

$$Q(\tau):=\int_\tau^\infty\varphi(y)\mathrm{d}y \qquad \forall\tau\in\mathbb{R}. \tag{3}$$

[1]Note that $F^{-1}_{\varphi(\mathsf{Y})}:(0,1)\mapsto(0,1/\sqrt{2\pi})$ is the inverse of the function $F_{\varphi(\mathsf{Y})}$.

Thus $\{y : |y| < Q^{-1}(\varepsilon/2)\}$ is the set of all y's satisfying the condition $\varphi(y) > \phi(\varepsilon)$, and consequently

$$\beta(\varepsilon) = 2Q^{-1}(\varepsilon/2) \qquad \forall \varepsilon \in (0,1). \tag{4}$$

What we have observed until now, is implied by Neyman–Pearson lemma and there is nothing new conceptually. The expressions in (2) and (4) for $F^{-1}_{\varphi(\mathsf{Y})}(\cdot)$ and $\beta(\cdot)$ allows us to establish alternative characterizations for $\beta(\cdot)$ and to relate $F^{-1}_{\varphi(\mathsf{Y})}(\cdot)$ and $\beta(\cdot)$.

First note that $\frac{\partial}{\partial \tau} Q(\tau) = -\varphi(\tau)$, (2), and (4) imply

$$\begin{aligned} \beta'(\varepsilon) &= \tfrac{-1}{\varphi\left(\Phi^{-1}\left(\frac{\varepsilon}{2}\right)\right)} \\ &= \tfrac{-1}{F^{-1}_{\varphi(\mathsf{Y})}(\varepsilon)} \qquad \forall \varepsilon \in (0,1). \end{aligned} \tag{5a}$$

Then using $\beta(1) - \beta(\varepsilon) = \int_\varepsilon^1 \beta'(\tau) \mathrm{d}\tau$ and $\beta(1) = 0$, we get

$$\beta(\varepsilon) = \int_\varepsilon^1 \tfrac{1}{F^{-1}_{\varphi(\mathsf{Y})}(\tau)} \mathrm{d}\tau \qquad \forall \varepsilon \in (0,1). \tag{5b}$$

On the other hand, $\beta(\cdot)$ is convex because $\beta'(\cdot)$ is increasing. Furthermore, $\beta(\cdot)$ can be described as the convex conjugate of its convex conjugate on $(0,1)$. Consequently, it has the following dual characterization

$$\begin{aligned} \beta(\varepsilon) &= \sup\nolimits_{\gamma \geq 0} \hbar(\gamma) - \gamma\varepsilon \qquad \forall \varepsilon \in (0,1), && \text{(5c)} \\ &= \hbar(\gamma_\varepsilon) - \gamma_\varepsilon \varepsilon \qquad \forall \varepsilon \in (0,1), && \text{(5d)} \end{aligned}$$

where $\hbar(\gamma) = \mathbf{E}[1/\varphi(\mathsf{Y}) \wedge \gamma]$ and $\gamma_\varepsilon = \frac{1}{F^{-1}_{\varphi(\mathsf{Y})}(\varepsilon)}$. Recall that the expected value of any non-negative random variable X equals to the integral of its complementary distribution function, i.e., $\mathbf{E}[\mathsf{X}] = \int_0^\infty \mathbf{P}(\mathsf{X} > \tau) \mathrm{d}\tau$, by [1, (1.30)], [2, p.247]. Thus,

$$\begin{aligned} \hbar(\gamma) &= \int_0^\infty \mathbf{P}\Big(\Big[\tfrac{1}{\varphi(\mathsf{Y})} \wedge \gamma\Big] > \tau\Big) \mathrm{d}\tau \qquad \forall \gamma \geq 0, \\ &= \int_0^\gamma \mathbf{P}\Big(\tfrac{1}{\varphi(\mathsf{Y})} > \tau\Big) \mathrm{d}\tau \qquad \forall \gamma \geq 0. \end{aligned}$$

Hence, (5d), the fact that $\mathbf{P}((1/\varphi(\mathsf{Y})) > \tau)$ is non-increasing in τ, and $\mathbf{P}(1/\varphi(\mathsf{Y}) > \gamma_\varepsilon) = \varepsilon$, imply

$$\begin{aligned} \beta(\varepsilon) &= \int_0^{\gamma_\varepsilon} \Big[\mathbf{P}\Big(\tfrac{1}{\varphi(\mathsf{Y})} > \tau\Big) - \varepsilon\Big] \mathrm{d}\tau \qquad \forall \varepsilon \in (0,1), && \text{(5e)} \\ &= \int_0^\infty \Big|\mathbf{P}\Big(\tfrac{1}{\varphi(\mathsf{Y})} > \tau\Big) - \varepsilon\Big|^+ \mathrm{d}\tau \qquad \forall \varepsilon \in (0,1), && \text{(5f)} \end{aligned}$$

where $|x|^+ = \max\{0, x\}$.

The above problem is a simple hypothesis testing problem with deterministic detectors between the standard Gaussian measure on $\mathbb{R}$ and the Lebesgue measure on $\mathbb{R}$; and $\varphi(y)$ is the likelihood-ratio of the outcome y. Then $\beta(\varepsilon)$ is the infimum type II error volume (length) when type I error probability is less than or equal to ε.

We call the infimum of type II error volume (i.e., missed detection volume, power) as a function of type I error volume (i.e., false alarm volume, significance level) the deterministic-detector Neyman–Pearson function (dNP-function) and denote it by $\beta(\cdot\|_w^q)$ when the measures associated with the alternative and null hypotheses are q and w, respectively. If the randomized detectors are permitted then we call it the randomized-detector Neyman–Pearson function (rNP-function) and denote it by $\beta(\cdot|_w^q)$. That is for any pair of σ-finite measures (w, q) on the measurable space $(\mathcal{Y}, \mathcal{Y})$, dNP-function $\beta(\cdot\|_w^q) : \mathbb{R} \mapsto [0,\infty]$ and rNP-function $\beta(\cdot|_w^q) : \mathbb{R} \mapsto [0,\infty]$ are defined as

$$\beta(\varepsilon\|_w^q) := \inf_{\mathsf{T} \in \mathcal{T}^d(\mathcal{Y}) : \int \mathsf{T} \mathrm{d}w \leq \varepsilon} \int (1 - \mathsf{T}) \mathrm{d}q \qquad \forall \varepsilon \in \mathbb{R}, \tag{6a}$$

$$\beta(\varepsilon|_w^q) := \inf_{\mathsf{T} \in \mathcal{T}(\mathcal{Y}) : \int \mathsf{T} \mathrm{d}w \leq \varepsilon} \int (1 - \mathsf{T}) \mathrm{d}q \qquad \forall \varepsilon \in \mathbb{R}, \tag{6b}$$

where $\mathcal{T}^d(\mathcal{Y})$ is the set of all $\mathcal{Y}$-measurable deterministic tests on $\mathcal{Y}$ (i.e., $\mathcal{Y}$-measurable functions of the form $\mathsf{T} : \mathcal{Y} \mapsto \{0,1\}$), and $\mathcal{T}(\mathcal{Y})$ is the set of all $\mathcal{Y}$-measurable randomized tests on $\mathcal{Y}$ (i.e., $\mathcal{Y}$-measurable functions of the form $\mathsf{T} : \mathcal{Y} \mapsto [0,1]$). Then $\beta(\cdot) = \beta(\cdot\|_w^q)$ on $[0,\infty)$ for the case when q is the Lebesgues measure and w is the standard Gaussian measure on $\mathbb{R}$ by definition. Furthermore, $\beta(\cdot) = \beta(\cdot|_w^q)$ on $[0,\infty)$ by (4) because $\beta(\cdot|_w^q)$ is the convex envelope of $\beta(\cdot\|_w^q)$ and $\beta(\cdot)$ is convex.

The extension of definitions of $\beta(\cdot|_w^q)$ and $\beta(\cdot\|_w^q)$ from the customary setting in which both w and q are probability measures, to the case when both w and q are σ-finite measures results in a subtlety: for some σ-finite measures w and q, both $\beta(\varepsilon|_w^q)$ and $\beta(\varepsilon\|_w^q)$ are infinite for all finite ε values. We deem those cases to be void cases. We will see that non-void cases are those cases satisfying the following condition:

$$w\Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w} \geq 1\Big) + q\Big(\tfrac{\mathrm{d}w_\sim}{\mathrm{d}q} > 1\Big) < \infty, \tag{7}$$

where $q_\sim$ is the w-absolutely continuous component of q and $w_\sim$ is the q-absolutely continuous component of w, [4, §31].

If w and q are σ-finite measures, then

$$\begin{aligned} \beta(\varepsilon|_w^q) &= \int_0^\infty \Big|w\Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w} > \tau\Big) - \varepsilon\Big|^+ \mathrm{d}\tau \qquad \forall \varepsilon \geq 0, && \text{(8a)} \\ &= \int_0^\infty \Big|w\Big(\tfrac{\mathrm{d}w_\sim}{\mathrm{d}q_\sim} \leq \tfrac{1}{\tau}\Big) - \varepsilon\Big|^+ \mathrm{d}\tau \qquad \forall \varepsilon \geq 0, && \text{(8b)} \end{aligned}$$

To the best of our knowledge, the characterization of $\beta(\cdot|_w^q)$ given in (8) has not been reported before, even for probability measures w and q satisfying $w = w_\sim$ and $q = q_\sim$. Furthermore, all of the identities given in (5) for $\beta(\cdot|_w^q)$, hold with appropriate modifications in all non-void cases, i.e., when w and q satisfy (7), see (61) of Theorem 1. We provide an account of related prior results in §I-A.

Evidently, neither (2) nor (4) is true for arbitrary σ-finite measures —or for arbitrary probability measures— w and q. In fact, $\beta(\cdot|_w^q)$ is not differentiable for some ε values for certain w and q. However, $\beta(\cdot|_w^q)$ is convex and finite on $(0,\infty)$ in all non-void cases. Thus $\beta(\cdot|_w^q)$ has directional derivatives everywhere on $(0,\infty)$. Furthermore, the left and right derivatives of $\beta(\cdot|_w^q)$ are $-1/F^\wedge_\ell\big(\cdot\big|_q^w\big)$ and $-1/F^\vee_\ell\big(\cdot\big|_q^w\big)$, where $F^\wedge_\ell\big(\cdot\big|_q^w\big)$ and $F^\vee_\ell\big(\cdot\big|_q^w\big)$ are the lower semicontinuous and the upper semicontinuous quantiles for the likelihood-ratio, see (32) and (37) in the following for formal definitions. When the distribution function $F_\ell\big(\cdot\big|_q^w\big)$ is constant on a subinterval of $(0,\infty)$, the inverse function in the usual sense does not exist; that is why we have to work with the semicontinuous inverses of $F_\ell\big(\cdot\big|_q^w\big)$, i.e., likelihood-ratio quantiles $F^\wedge_\ell\big(\cdot\big|_q^w\big)$ and $F^\vee_\ell\big(\cdot\big|_q^w\big)$. Whenever $F_\ell\big(\cdot\big|_q^w\big)$ is invertible $F^{-1}_\ell\big(\cdot\big|_q^w\big) = F^\wedge_\ell\big(\cdot\big|_q^w\big) = F^\vee_\ell\big(\cdot\big|_q^w\big)$ holds, as one would expect.

The function $\hbar(\cdot)$ in (5c) is the primitive entropy spectrum of the Lebesgue measure on $\mathbb{R}$ with respect to the standard Gaussian measure on $\mathbb{R}$. For arbitrary σ-finite measures q and w on the same measurable space, the primitive entropy spectrum of q with respect to w is defined as

$$\hbar(\gamma|_w^q) := \int \Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w} \wedge \gamma\Big)\mathrm{d}w \qquad \forall \gamma \in [0,\infty). \tag{9}$$

Evidently, the primitive entropy spectrum has the following variational characterization

$$\hbar(\gamma|_w^q) = \inf_{\mathcal{E}\in\mathcal{Y}} q(\mathcal{E}) + \gamma w(\mathcal{Y}\setminus\mathcal{E}) \qquad \forall \gamma \in [0,\infty). \tag{10}$$

We obtain the spectral representation of $\hbar(\cdot|_w^q)$, using the following level set integration, which is a consequence of integration by parts, [5, Theorem 2.9.3]: for any measure[2] μ and μ-measurable function f satisfying $\operatorname{ess\,inf}_\mu f \geq 0$,

$$\int f \mathrm{d}\mu = \int_0^\infty \mu(f > \tau)\mathrm{d}\tau. \tag{11}$$

In the parlance of probability theory (i.e., when μ is a probability measure) $\mathbf{E}[f] = \int_0^\infty \mathbf{P}(f>\tau)\mathrm{d}\tau$ for any non-negative random variable f. In [1, §1.5], Gallager defines the expectation of a random variable using this identity, i.e., the tail expectation formula, yielding a deceptively simple, yet fully rigorous construction aligned with measure-theoretic foundations.

What one learns from (5) that is not explicit in (4) is how $\beta(\cdot)$ as a function is related to $\hbar(\cdot)$ and $F^{-1}_{\varphi(\mathsf{Y})}(\cdot)$ as functions. Analogously Theorem 1, in the following, describes how $\beta(\cdot|_w^q)$ as a function is related to $\hbar(\cdot|_w^q)$, $F_\ell^\wedge\big(\cdot\big|_q^w\big)$, and $F_\ell^\vee\big(\cdot\big|_q^w\big)$ as functions for a given pair of σ-finite measures q and w. In other words, Theorem 1 makes the implications of the Neyman–Pearson lemma explicit and extends the lemma and its consequences to randomized decisions between σ-finite measures.

As pointed out in [6] and later employed in [7], lossless source coding provides an operational justification for hypothesis testing problems where the measure associated with the alternative hypothesis is a σ-finite measure rather than a probability measure. Specifically, finding the minimum cardinality of a subset of a countably infinite set that contains all but ε of the probability mass of a discrete probability measure w is a hypothesis testing problem against the counting measure $\sharp$ with deterministic tests. In this context, $\beta\big(\varepsilon\big\|_w^\sharp\big)$ is the cardinality of any such minimum cardinality sets. While the counting measure $\sharp$ is not finite for countably infinite sets, it remains σ-finite for any countable set.

Furthermore, for the lossless source coding problem, the randomized-detector Neyman–Pearson function determines the value of the deterministic-detector Neyman–Pearson function. Note that for those ε values for which a minimum-cardinality-set-$\mathcal{E}$ satisfies $w(\mathcal{Y}\setminus\mathcal{E}) = \varepsilon$, we have $\beta\big(\varepsilon\big\|_w^\sharp\big) = \beta\big(\varepsilon\big|_w^\sharp\big)$. On the other hand, if $w(\mathcal{Y}\setminus\mathcal{E}) < \varepsilon$ for a minimum-cardinality-set-$\mathcal{E}$, then $\beta\big(\varepsilon\big|_w^\sharp\big) < \beta\big(\varepsilon\big\|_w^\sharp\big) < \beta\big(\varepsilon\big|_w^\sharp\big)+1$. Thus, using the fact that $\beta\big(\cdot\big\|_w^\sharp\big)$ is an integer valued function, we conclude

$$\begin{aligned}
\beta\big(\varepsilon\big\|_w^\sharp\big) &= \lceil \beta\big(\varepsilon\big|_w^\sharp\big)\rceil && \forall \varepsilon \geq 0, && \text{(12a)}\\
&= \left\lceil \int_0^\infty \Big| w\Big(\tfrac{\mathrm{d}\sharp_\sim}{\mathrm{d}w} > \tau\Big) - \varepsilon\Big|^+ \mathrm{d}\tau \right\rceil && \forall \varepsilon \geq 0, && \text{(12b)}\\
&= \left\lceil \int_0^\infty \Big| w\Big(\tfrac{\mathrm{d}w}{\mathrm{d}\sharp} \leq \tfrac{1}{\tau}\Big) - \varepsilon\Big|^+ \mathrm{d}\tau \right\rceil && \forall \varepsilon \geq 0, && \text{(12c)}
\end{aligned}$$

(12b) follows from (8a), and (12c) follows from (8b). Note that, in general $\beta(\cdot\|_w^q)$ is bounded but not necessarily determined by $\beta(\cdot|_w^q)$, see Remark 2.

In the remainder of this section, we first briefly review prior work related to $\beta(\cdot|_w^q)$, $F_\ell^\vee\big(\cdot\big|_q^w\big)$, and $\hbar(\cdot|_w^q)$ in §I-A. Next, we discuss prior work on approximations of rNP-functions in §I-B. We then highlight our main contributions in §I-C and conclude the introduction with outlining the remainder of the paper in §I-D.

A. Prior Work On $\beta(\cdot|_w^q)$, $F_\ell^\vee\big(\cdot\big|_q^w\big)$, and $\hbar(\cdot|_w^q)$

The simple hypothesis testing problem is a fundamental problem that can be motivated by various applications; our interest in the problem is from the perspective of information transmission problems [8]–[12].

The importance of hypothesis testing problem, and of the NP-functions in particular, was already clear for information transmission problems by the time of Strassen's seminal work [6]. Furthermore, as we have recalled above the lossless source coding problem provides a concrete operational justification for considering $\beta(\cdot|_w^q)$ for the case when q is a σ-finite measure, which is not finite. The hypothesis testing problem recognized to play a crucial role for understanding the behavior of the error probability in the channel coding problem, not only at rates close to capacity [6], but also at other rates [13]–[17].

The hypothesis testing problem is now a key tool for analyzing and understanding information transmission problems, [8, Ch. 1], [9, Ch. 3 and Appendix 5.A], [10, Ch. 5 and 11], [11, Ch. 1 and 4], [12, Ch. 10, 11, and Part III].

The interest in finite-length (i.e., non-asymptotic) analysis and exact asymptotic analysis in information transmission problems revived the interest in $\beta(\cdot|_w^q)$, [18]–[25]. In particular, the maximum number of messages that a code on the channel $W : \mathcal{X} \mapsto \mathcal{P}(\mathcal{Y})$ with maximum error probability ε can have, i.e., $M(W,\varepsilon)$, is bounded above as,

$$M(W,\varepsilon) \leq \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{x\in\mathcal{X}} \frac{1}{\beta\Big(\varepsilon\Big|_{W(x)}^q\Big)}, \tag{13}$$

see [20, Theorem 31]. This upper bound is studied more closely in [21]. In addition new divergences between probability measures, such as the hypothesis testing divergence in [24, p. 7044], [25, (2.6)], are introduced using $\beta(\cdot|_w^q)$ so as to analyze information transmission problems in non-asymptotic and refined asymptotic regimes, see [24]–[28].

The maximum number of messages is also bounded from below for any probability mass function p on the input set $\mathcal{X}$ and positive threshold γ in terms of the likelihood ratio

[2] All measures are assumed to be non-signed and countably-additive.

distribution as a result of the maximum-code-theorem, see [17, Theorem 2.1], as

$$M(W,\varepsilon) \geq \gamma\Big(\varepsilon - F_\ell\Big(\gamma\Big|{p\otimes W \atop p\otimes(pW)}\Big)\Big), \tag{14}$$

where pW is output marginal of the joint distribution $p\otimes W$ generated by the input distribution p and the channel W.

The likelihood-ratio quantiles $F_\ell^\wedge\big(\cdot\big|{w\atop q}\big)$ and $F_\ell^\vee\big(\cdot\big|{w\atop q}\big)$, defined as the semi-continuous inverses of likelihood ratio distribution $F_\ell\big(\cdot\big|{w\atop q}\big)$, determine the sub-differential of $\beta(\cdot|{q\atop w})$, as we will see in (61e) and (61f), and have seen in (5a) for a specific case. The upper-semicontinuous likelihood-ratio quantile $F_\ell^\vee\big(\cdot\big|{w\atop q}\big)$ and its natural logarithm (i.e., $\ln F_\ell^\vee\big(\cdot\big|{w\atop q}\big)$, the log-likelihood quantile) have been introduced and analyzed by many authors under various names. $F_\ell^\vee\big(\cdot\big|{w\atop q}\big)$ is denoted by μ in [6, (2.5)]. Inspired by [11, Ch. IV], $\ln F_\ell^\vee\big(\cdot\big|{w\atop q}\big)$ is called information spectrum divergence in [25, (2.9)], the classical entropic information spectrum in [28, (2)], the relative entropy information spectrum (and the divergence spectrum) in [24, p. 7044]. Furthermore, $\beta(\cdot|{q\atop w})$ has been bounded in terms of $F_\ell^\vee\big(\cdot\big|{w\atop q}\big)$ for any probability measures w and q provided that w is absolutely continuous in q, i.e., $w\prec q$, see [19, (2.67), (2.69)], [20, (103)], [24, Lemma 2], [25, Lemma 2.4], [28, Lemma 12]:

$$\tfrac{1}{F_\ell^\vee\left(\varepsilon\left|{w\atop q}\right.\right)} \geq \beta(\varepsilon|{q\atop w}) \geq \tfrac{\delta}{F_\ell^\vee\left(\varepsilon+\delta\left|{w\atop q}\right.\right)} \quad \forall \varepsilon\in(0,1), \delta\in(0,1-\varepsilon). \tag{15}$$

The upper and lower bounds in (15) is the state-of-the-art relation between likelihood-ratio quantiles and $\beta(\cdot|{q\atop w})$, prior to Theorem 1. It is worth noting that although the upper semicontinuous quantile $F_\ell^\vee\big(\cdot\big|{w\atop q}\big)$ appear more prominently in information theory; when used without any qualifiers, what is meant by quantile in probability theory, usually, is the lower semicontinuous quantile, see [2, p. 427].

For probability measures w and q, the primitive entropy spectrum of $\hbar\big(\cdot\big|{w\atop q}\big)$ is related to the primitive divergence spectrum $d\big(\cdot\big|{w\atop q}\big)$, employed in the spectral representation of f-divergences [29]–[32] reported in [33]–[37]:

$$d\big(\gamma\big|{w\atop q}\big) = 1\wedge\gamma - \hbar\big(\gamma\big|{w\atop q}\big) \qquad \forall\gamma\in[0,\infty). \tag{16}$$

Note that $d\big(\gamma\big|{w\atop q}\big)$ is the f-divergence $D_f(w\|q)$ corresponding to the function $f_\gamma(x)=(1\wedge\gamma)-(x\wedge\gamma)$ for each $\gamma\in[0,\infty)$. The quantity $\frac{1}{1+\gamma}\hbar\big(\gamma\big|{w\atop q}\big)$ is called Bayes loss in [34, (66)], posterior Bayes risk in [35, p.734], and a posteriori loss in [36, (2.25)]. The primitive entropy spectrum $\hbar\big(\cdot\big|{w\atop q}\big)$ is denoted by $\psi_{w,q}(\cdot)$ in [37, §III.A]. The left and right derivatives of $\hbar\big(\cdot\big|{w\atop q}\big)$ were determined in [37, Lemma 3.2] for the case when w and q are probability measures.

For any channel $W:\mathcal{X}\mapsto\mathcal{P}(\mathcal{Y})$, probability mass function p on $\mathcal{X}$, and integer $M\geq 2$, there exists a code with M messagages whose average error probability ε satisfies

$$\varepsilon \leq \hbar\Big(\tfrac{M-1}{2}\Big|{p\otimes W\atop p\otimes(pW)}\Big), \tag{17}$$

see [20, (76)] and (51a). A similar dependence test bound was suggested for the lossless source coding problem in [38, Theorem 3], establishing an upper bound to the error probability when source code can have at most M distinct outputs for any probability mass function w

$$\varepsilon \leq \hbar\big(\tfrac{1}{M}\big|{\sharp\atop w}\big). \tag{18}$$

Motivated by an expression equivalent to [20, (76)] and inspired by f-divergences [30], the divergence $E_\gamma(w\|q)$ was defined in [19, (2.141)] for all probability measures w and q such that w is absolutely continuous in q, i.e., $w\prec q$, see also [39]–[41]. One can show for all probability measures w and q satisfying the hypothsis $w\prec q$, that

$$E_\gamma(w\|q) = 1-\hbar\big(\gamma\big|{w\atop q}\big) \qquad \forall\gamma\in[0,\infty). \tag{19}$$

For all probability measures w and q satisfying $w\sim q$, i.e., satisfying both $w\prec q$ and $q\prec w$, the directional derivatives of $E_\gamma(w\|q)$ as a function of γ were determined and an integral representation of the Kulllback-Leibler divergence $D(w\|q)$ in terms of $E_\gamma(w\|q)$, similar to the more general ones for f-divergences in terms of $d\big(\gamma\big|{w\atop q}\big)$ in [33]–[37], was established in [19, Theorem 21]. Later this result is extended all f-divergences with twice differentiable f's relying on [34], see [39, Proposition 3].

The following measure change bound, recently established in [42, Proposition 1] for probability measure w and q satisfying $w\prec q$, follows from (10) and (19):

$$E_\gamma(w\|q) + \gamma q(\mathcal{E}) \geq w(\mathcal{E}) \qquad \forall\mathcal{E}\in\mathcal{Y}, \gamma>0. \tag{20}$$

None of the identities of given in (61) of Theorem 1 was reported before in this form, even for the case of arbitrary probability measures w and q. However, there are a number of closely related prior results.

The prior results closest to Theorem 1 are [43] and [44], which we became aware of after the initial arXiv and ISIT submissions of the present work [45]–[47]. For the quantum formulation of the problem under the assumption that underlying Hilbert space is finite dimensional, [43, Proposition 3.2] characterizes rNP-function $\beta(\cdot|{q\atop w})$ as its own bidual for convex conjugation.[3] Thus proposition [43, Proposition 3.2] implies (61a) and (61b) for the case when w and q are probability mass functions on a finite set $\mathcal{Y}$. Later [44] set out to express $\beta(1-\alpha|{q\atop w})$ as the Legendre transform (i.e., convex conjugate) of a function for the case when w and q are probability measures on a countable set $\mathcal{Y}$, i.e., w and q are probability mass functions. [44, Lemmas 1 and 2] establish expressions that are equivalent to (61a), (61b), and (61c). More recently, a characterization of $\beta\big(\cdot\big|{w\atop q}\big)$ that is equivalent to (61a) and (61b) of Theorem 1 was established for the relative information spectrum,[4] see [50, (4.1) and (8.35)], [51], for the case when w and q are probability measures. A relation equivalent to (61b) of Theorem 1 was reported in [19, (2.147) of Theorem 21] for probability measures w and q satisfying $w\sim q$.

Working under implicit assumption that $\beta(\cdot|{q\atop w}) = \beta(\cdot\|{q\atop w})$, which is guaranteed to hold when $w_\sim$ does not have any atoms, an expressions equivalent to (61a) of Theorem 1 and (63) were reported in [35, Theorem 21], see [35, (43)] and [35, (42)]. The conjugacy relation established in (63) is not present in their framework because of their parameterization. Nevertheless [35, (42) and (43)] are deemed to be "dual" relations in [35, Theorem 21].

[3]The dual characterization of the rNP-function in [43, Proposition 3.2] has been modified, since [45], to adopt primitive entropy spectrum form, see [48, Proposition 6.1] and [49, Lemma 1].

[4]The relative information spectrum in [50], [51] is equal to $F_\ell\big(\exp(\cdot)\big|{w\atop q}\big)$.

A closely related result was also reported in the framework of lossless source coding. [52] characterizes the optimal performance of one-to-one but not necessarily prefix free variable length binary codes in terms of a quantity denoted by $R^*(n,\epsilon)$, see [52, (5)]. The value of $R^*(n,\varepsilon)$ is determined by $\beta\big(\varepsilon\big\|_w^\sharp\big)$:

$$nR^*(n,\varepsilon) = \lceil \log_2\big(1+\beta\big(\varepsilon\big\|_w^\sharp\big)\big)\rceil - 1. \tag{21}$$

Thus the value of $R^*(n,\varepsilon)$ is determined by $\beta\big(\varepsilon\big|_w^\sharp\big)$, as a result of (12a). [52, Theorem 7] characterizes $R^*(\varepsilon)=R^*(1,\varepsilon)$ for ε values satisfying $\varepsilon = w\Big(\frac{\mathrm{d}w}{\mathrm{d}\sharp}\le\gamma\Big)$ for some $\gamma\ge 0$. Evidently, not all ε values in $(0,1)$ interval satisfy that condition. This, however, does not prevent [52, Theorem 7] from characterizing $R^*(\cdot)$ and $\beta\big(\cdot\big\|_w^\sharp\big)$ as functions on $(0,1)$ as long as all $w(\{y\})$ values are distinct for all y in $\mathcal{Y}$. When multiple y's have the same $w(\{y\})$ value, however, [52, Theorem 7] is mute about the behavior of $R^*(\cdot)$ and $\beta\big(\cdot\big\|_w^\sharp\big)$ for the values of ε corresponding to these y's. Consider for example the case when w is the uniform distribution on a finite set or when w is weighted sum of uniform distributions on disjoint sets. Evidently, (12) and (21) characterize $R^*(n,\varepsilon)$ completely even in such cases.

B. Prior Work On Approximations

In many applications one is interested in the behavior of $\beta(\cdot|_w^q)$ on a sub-interval of the positive real line under various structural assumptions on q and w. In this subsection, we consider the prior work on the problem only for the memoryless and discrete Markov cases. In the memoryless case w and q are both product measures of the form $w=\otimes_{t=1}^n w_t$ and $q=\otimes_{t=1}^n q_t$, where w_t and q_t are measures on the same measurable space $(\mathcal{Y}_t,\mathcal{Y}_t)$. The memoryless case is called stationary if $(w_t,q_t)=(w_1,q_1)$ for all t. In the time-homogeneous discrete Markov case w and q are both given in terms of an initial σ-finite measure on a countable set $\mathcal{S}$ and a discrete σ-finite transition measure from $\mathcal{S}$ to $\mathcal{S}$; that is w and q are of the form $w=w_0\otimes W^{\otimes n}$ and $q=q_0\otimes Q^{\otimes n}$. All of the results mentioned in this subsection are derived for the case when w is a probability measure. In the following, we deem an approximation non-asymptotic only when both the approximation error term and the condition on n (and on other quantities if any) are explicit in the statement of the result.

For the stationary memoryless case[5] [6, Theorem 1.1] determined $\ln\beta(\varepsilon\|_w^q)$ —and $\ln\beta(\varepsilon|_w^q)$ within the proof— up to an $o(1)$ approximation error term for every $\varepsilon\in(0,1)$. The result for $\ln\beta(\varepsilon|_w^q)$ is extended to the possibly non-stationary memoryless case under appropriate hypotheses with the additional assumption of q being a probability measure, see [22, Theorem 18]. For the time-homogeneous discrete Markov case with a finite state space $\mathcal{S}$, assuming W corresponds to an irreducible aperiodic Markov chain and loglikelihood ratios are non-lattice the same result is proved in [54] for the case when q is Markov probability measure.

[5]Although [6, Theorem 1.1] is stated for the case when $\mathcal{Y}_1$ is a finite set, proof does not rely on that assumption and the extension for arbitrary measurable space is explicitly mentioned in [6, §5].

For the possibly non-stationary memoryless case, [6, Theorem 3.1] provides an approximation for $\ln\beta(\cdot|_w^q)$ on $(\delta,1-\delta)$ whose approximation error is bounded from above uniformly by $^{140}/_{\delta^8}$ provided that $\sqrt{n}\ge{}^{140}/_{\delta^8}$, where δ is constrained by certain average moment conditions due to the approximation error term in Berry–Esseen Theorem. In most cases of interest for any positive δ, non-asymptotic approximation for $\ln\beta(\cdot|_w^q)$ on $(\delta,1-\delta)$ by [6, Theorem 3.1] has a uniform $O(1)$ approximation error. An equally tight approximation was established later in [19, Lemma 14], [44, §III.B]. For the lossless source coding problem, i.e., when $q=\sharp$, a qualitatively equivalent approximation is obtained for the memoryless case in[6] both stationary [52, (44)] and non-stationary [55, Theorems 5 and 6] cases. For the time-homogeneous discrete Markov case with a finite state-space assuming that W corresponds to an irreducible aperiodic Markov chain, the lower bound of the approximation is established in [52, Theorem 20] and the upper bound of the approximation is conjectured in [52]. More recently, for the memoryless case an asymptotic approximation for $\ln\beta(\varepsilon_n|_w^q)$ was established under the hypothesis $\big|Q^{-1}(\varepsilon_n)\big|=o(\sqrt{n})$ with an approximation error term of the form $O\Big(\frac{(Q^{-1}(\varepsilon_n))^4}{n}\Big)+O(1)$, see [23, Theorem 4].

In approximations for $\ln\beta(\varepsilon|_w^q)$, the leading term is $-D(w\|q)$ both for the memoryless and Markovian cases where $D(w\|q)$ is the Kullaback-Leibler divergence. Recall that for any probability measure w and σ-finite measure q on the same measurable space satisfying $w\prec q$, the Kullaback-Leibler divergence $D(w\|q)$ is defined as

$$D(w\|q) := \mathbf{E}_w\Big[\ln\tfrac{\mathrm{d}w}{\mathrm{d}q}\Big], \tag{22}$$

where $\mathbf{E}_w[\cdot]$ is the expected value for random variables on the probability space $(\mathcal{Y},\mathcal{Y},w)$, whenever the expectation on the right hand side exists.

For the stationary memoryless case, Csiszar and Longo[7] determined the behavior of $\beta\big(e^{-nE}\big\|_q^w\big)$ for any fixed E value in [7], with the help of [6, Theorem 1.1]. In particular, $\ln\beta\big(e^{-nE}\big\|_q^w\big)$ for $E<D(w_1\|q_1)$ case and $\ln(1-\beta\big(e^{-nE}\big\|_q^w\big))$ for $E>D(w_1\|q_1)$ case, were determined in [7] up to an $o(1)$ approximation error term for a finite measure q. The Kullback–Leibler divergence $D(w_1\|q_1)$ acts as a threshold for the change in behavior of $\beta\big(e^{-nE}\big\|_q^w\big)$ with increasing n: if $E<D(w_1\|q_1)$, then $\beta\big(e^{-nE}\big\|_q^w\big)$ converges to 0; if $E>D(w_1\|q_1)$, on the other hand, $\beta\big(e^{-nE}\big\|_q^w\big)$ converges to 1. An equally tight bound is obtained via saddle point approximation in [56, Theorem 2], for the case when $w\sim q$ and q is probability measure.

The total variation norm of a signed measure v on the measurable space $(\mathcal{Y},\mathcal{Y})$ is defined as

$$\|v\| := \sup_{\mathcal{E}\in\mathcal{Y}} v(\mathcal{E}) + |v(\mathcal{Y}\setminus\mathcal{E})|. \tag{23}$$

In [7], a measure change argument is employed via tilting and the value of E is set to $D(v\|q)$ for $v={}^{v_\alpha}\!/_{\|v_\alpha\|}$ where v_α is the tilted measure defined in (77a). For all $\alpha\in(0,1)$, a

[6]Note that for the lossless source coding problem $nR^*(n,\varepsilon)$ is essentially, $\ln\beta\Big(\cdot\big|_w^\sharp\Big)$ as a result of (12) and (21)

[7]There is a typo in [7], see [57, Remark 3.7].

non-asymptotic approximation for $\ln \beta\big(e^{-D(v\|q)+a_n}\big\|{}^{w}_{q}\big)$ with $O(\sqrt{n})$ approximation error term is established in [13, Theorem 5] assuming $|a_n| = O(\sqrt{n})$ for the possibly non-stationary memoryless case, prior to [7], by invoking Chebyshev's inequality instead of [6, Theorem 1.1]. More recently non-asymptotic approximations for $\ln \beta\big(e^{-D(v\|q)+a_n}\big\|{}^{w}_{q}\big)$ with $O(1)$ approximation error terms have been be obtained under $|a_n| = O(\sqrt{n})$ hypothesis by employing Berry–Esseen Theorem instead of Chebyshev's inequality for the memoryless case see [57]–[61]. For $\alpha \in (0,1)$ case, an asymptotic approximation of similar nature were derived earlier in [62] relying on strong large deviation results of [63], [64], rather than Berry–Esseen Theorem. For $\alpha > 1$ case, a $O(1)$ approximation for $\ln(1-\beta\big(e^{-D(v\|q)+a_n}\big\|{}^{w}_{q}\big))$ was established in [65] relying on Berry–Esseen Theorem.

C. Main Contributions

One of the main contributions of this work is the explicit characterization of implications of Neyman–Pearson lemma in Theorem 1 for the relations between $\hbar(\cdot|{}^{q}_{w})$, $F^{\wedge}_{\ell}\big(\cdot\big|{}^{w}_{q}\big)$, $F^{\vee}_{\ell}\big(\cdot\big|{}^{w}_{q}\big)$ and $\beta(\cdot|{}^{q}_{w})$ as functions for σ-finite measures w and q satisfying (7). Resulting identities in (61) and (62), have not been reported either in this form or with this generality before to the best our knowledge. In particular, the identities in (8) (i.e., (62b) and (61d)), (61g), and (61j) have not appeared before. However, as we have already pointed out statements equivalent to (61a), (61b), and (61c) have appeared in [19], [35], [44], [50], [51], as discussed in §I-A. Furthermore, for a different parametrization convex conjugate of $\beta(\cdot|{}^{q}_{w})$ was characterized in [44] for the case when w and q are probability mass functions on a countable set.

Theorem 2 establishes sufficient conditions for expressing $\beta(\cdot|{}^{q}_{w})$ either as $\beta(g(\cdot)|{}^{q}_{v})$ or as $\|q_{\sim}\| - \beta(g(\cdot)|{}^{q}_{v})$ for some an invertible function $g(\cdot)$. Such a characterization has not been reported before and works quite well in simplifying the derivations when employed with the monotonicity properties of $\beta(\cdot|{}^{q}_{w})$ described in Lemma 5 and tilting, as we demonstrate while establishing Theorem 4 and Corollary 4.

Note that with the changes of variable $\tau = e^{t}$ and $\tau = e^{-t}$, (8a) and (8b) become

$$\beta(\varepsilon|{}^{q}_{w}) = \int_{-\infty}^{\infty} e^{t} \Big| w\Big(\ln \tfrac{\mathrm{d}q_{\sim}}{\mathrm{d}w} > t\Big) - \varepsilon \Big|^{+} \mathrm{d}t \qquad \forall \varepsilon \geq 0, \quad \text{(24a)}$$

$$= \int_{-\infty}^{\infty} e^{-t} \Big| w\Big(\ln \tfrac{\mathrm{d}w_{\sim}}{\mathrm{d}q_{\sim}} \leq t\Big) - \varepsilon \Big|^{+} \mathrm{d}t \quad \forall \varepsilon \geq 0. \quad \text{(24b)}$$

If the Kolmogorov distance of the log-likelihood ratio distribution, i.e., $w\Big(\ln \frac{\mathrm{d}w_{\sim}}{\mathrm{d}q_{\sim}} \leq \cdot\Big)$, to a Gaussian distribution is small, then one can approximate $\beta(\cdot|{}^{q}_{w})$ using that Gaussian distribution with the help of (24b). Theorem 3 pursues this idea to characterize $\beta(\cdot|{}^{q}_{w})$ up to a bounded explicit multiplicative error term uniformly on all closed subsets of $(0,1)$. For the case when w and q are probability mass functions on a countable set a similar and qualitatively equivalent one-shot bound was derived in [44, §III.B].

The customary approach in bounding $\beta(\varepsilon|{}^{q}_{w})$, adopted explicitly or implicitly in much of the prior work (some of which is cited in §I-B), first invokes (15) and then bounds likelihood-ratio quantile $F^{\vee}_{\ell}\big(\cdot\big|{}^{w}_{q}\big)$. With the help of (24b), we not only avoid using likelihood-ratio quantiles completely, but also obtain tighter results with briefer analysis. Furthermore, it should be noted that Theorem 3 can also be proved by first expressing $\beta(\cdot|{}^{q}_{w})$ in terms of likelihood ratio quantiles $F^{\vee}_{\ell}\big(\cdot\big|{}^{w}_{q}\big)$ using[8] (61g) and then bounding $F^{\vee}_{\ell}\big(\cdot\big|{}^{w}_{q}\big)$ by invoking [25, (2.34)].

Kolmogorov distance to normality hypothesis for the log-likelihood ratio distribution is pursued with tilting in Theorem 4, for $v = \gamma v_{\alpha}$ and appropriately bounded τ to characterize $\beta\big(\tau e^{-D(v\|q)}\big|{}^{w}_{q}\big)$ for $\alpha \in (0,1)$ and $\|w_{\sim}\| - \beta\big(\tau e^{-D(v\|q)}\big|{}^{w}_{q}\big)$ for $\alpha > 1$, upto a bounded multiplicative error term by invoking Theorem 2 and Theorem 3. Theorem 4 or a qualitatively equivalent bound has not appeared before to the best of our knowledge.

In the memoryless case, hypothesis of Theorems 3 and 4 are implied by the Berry–Esseen Theorem and Corollaries 1 and 2 follow. Corollary 1 is qualitatively equivalent to the previously known best results of this kind both for hypothesis testing [6, Theorem 3.1], [19, Lemma 14], [44, §III.B] and for lossless source coding [52, (44), Theorems 17 and 18], [55, Theorems 5 and 6], discussed in §I-B. Likewise, Corollary 2 recovers previously known best results of its kind for both for hypothesis testing [57]–[59], [65] and for lossless source coding [60], [61], discussed in §I-B. Corollary 2 demonstrates for the $\alpha \in (0,1)$ case, in contrast to prior work, neither w nor q need to be finite measure for these approximations to hold.

For the time-homogeneous discrete Markov case if W corresponds to an irreducible aperiodic Markov chain satisfying Dobrushin–Doeblin condition, see [66, (1.1)], [67, (0.1)], then the hypothesis of Theorem 3 follows from the functional Berry–Esseen Theorem of Nagaev [67, Theorem 1] and Corollary 3 follows. For the lossless source coding problem, Corollary 3 generalizes the lower bound established in [52, Theorem 20] and establishes the improvement of [52, Theorem 19] conjectured[9] in [52], without making the finite state-space assumption of [52]. The approximation error of Corollary 3 for $\ln \beta(\cdot|{}^{q}_{w})$ is $O(1)$, which is worse than $o(1)$ approximation error derived in [54] under the additional finite state-space and non-lattice loglikelihood ratio assumptions. Without non-lattice loglikelihood ratio assumption [71, Theorem 9.6] with $o(\sqrt{n})$ approximation error, which is qualitatively equivalent to [53], is the best approximation we are aware of, even for the finite state-space case.

Corollary 4 establishes for the time-homogeneous discrete Markov case with a countable state-space the approximation corresponding to the one reported in Corollary 2 for the memoryless case, relying on the concept of R-positivity [68]–[70] and Doob's h-transform. Corollary 4 improves previously known best approximations in the same regime both for hypothesis testing [71]–[74] and lossless source coding [74]–[76], which are, except for [72], derived under the additional finite state-space assumption.

[8] Whenever $\|q_{\sim}\|$ is finite, (61j) can be used instead of (61g).

[9] [52] points out that [52, Theorems 19] can be improved but only by "extending the involved proof of" [52, Theorems 17].

Corollaries 1 and 3 are non-asymptotic results following from the Berry–Esseen Theorems; they can be interpreted as third order non-asymptotic Stein's lemmas for the memoryless and discrete Markovian cases. Theorem 5 establishes the second order Stein's lemma for the case when log-likelihood-ratio of a standard simple hypothesis testing problem satisfies the Central Limit Theorem, using only Theorem 3 and the fact that convergence in distribution to the normal distribution is metrized by the Kolmogorov distance. For the Markov case the fact that Central Limit Theorem implies the second order Stein's lemma was previously pointed out in [71], as well, see [71, Corollary 7.2 and Theorem 9.6]. Theorem 5 establishes the same result far more generally.

D. Outline

In §II, we present preliminary results that are later employed in our analysis. We start by recalling certain well-known properties of convex functions in §II-A. We introduce the semicontinuous inverses of monotonic functions and explain their construction intuitively in §II-B. Then we introduce the likelihood-ratio distribution $F_\ell\big(\cdot\big|{}^{w}_{q}\big)$ and its semicontinous inverses, i.e., the likelihood-ratio quantiles $F^{\wedge}_\ell\big(\cdot\big|{}^{w}_{q}\big)$ and $F^{\vee}_\ell\big(\cdot\big|{}^{w}_{q}\big)$, and review certain elementary properties of the likelihood-ratio quantiles in §II-C. We conclude the preliminaries section by deriving elementary properties of the primitive entropy spectrum $\hbar\big(\cdot\big|{}^{w}_{q}\big)$ in §II-D.

We start §III, by discussing certain elementary observations about $\beta(\cdot\|{}^{q}_{w})$ and $\beta(\cdot|{}^{q}_{w})$. Then in §III-A, we obtain an upper bound on $\beta(\cdot|{}^{q}_{w})$ by analyzing performance of a randomized likelihood-ratio test. In §III-B, we prove Theorem 1, which is essentially the Neyman–Pearson lemma for σ-finite measures, with the additional identities we discussed above including (8). In §III-C, we prove Theorem 2, which establishes sufficient conditions for expressing $\beta(\cdot|{}^{q}_{w})$ in terms of $\beta(\cdot|{}^{q}_{v})$ for an auxiliary measure v.

In §IV, we first review certain well-known properties of Gaussian Mills ratio and then derive Theorems 3 and 4 under small Kolmogorov distance to normality hypothesis both without and with tilting, respectively.

§V is devoted to applications. In §V-A, we consider the memoryless case and establish Corollaries 1 and 2. In §V-B, we consider time-homogeneous discrete Markov case with possibly countably infinite number of states and establish Corollaries 3 and 4. In §V-C, we introduce the standard simple asymptotic hypothesis testing problem, in its complete measure theoretic generality and show that the Central Limit Theorem for the log-likelihood-ratio implies the second order Stein's lemma in Theorem 5.

In §VI, we discuss the broader implications of the spectral approach proposed in this paper for information transmission problems, and highlight several interesting technical questions that have became more manageable as a result of this work.

II. Preliminaries

A. Convex Functions on the Real Line

Assumption 1. The function $f:\mathbb{R}\mapsto\mathbb{R}\cup\{\infty\}$ takes at least one finite value and has an affine minorant, i.e., $\exists x_0\in\mathbb{R}$ such that $f(x_0)<\infty$ and $\exists a,b\in\mathbb{R}$ such that $f(x)\geq ax+b$ holds for all $x\in\mathbb{R}$. (f in not necessarily convex.)

For any function f satisfying Assumption 1, subdiffential of f at x is, [77, (E.1.4.1)],

$$\partial f(x):=\{s\in\mathbb{R}: f(y)\geq f(x)+s(y-x) \text{ for all } y\in\mathbb{R}\}. \quad (25)$$

If f is convex and x is in the interior of the interval on which f is finite, then $\partial f(x)$ is the closed interval between the left and right derivatives of f at x by [78, Theorem I.4.1.1]:

$$\partial f(x):=[\mathrm{D}_- f(x),\mathrm{D}_+ f(x)] \qquad \forall x\in\mathtt{int}(\mathtt{dom}(f)). \quad (26)$$

If f is closed (i.e., lower semicontinuous) and a is the left end point of the domain then $\partial f(a)=(-\infty,\mathrm{D}_+f(a)]$, see [78, p.22]. If f is closed and b is the right end point of the domain then $\partial f(b)=[\mathrm{D}_-f(b),\infty)$.

If $f:\mathbb{R}\mapsto\mathbb{R}\cup\{\infty\}$ is a closed (i.e., lower semicontinuous) convex function and $[a,b]\subset\mathtt{dom}(f)$, then

$$f(x)=f(a)+\int_a^x f'(\tau)\mathrm{d}\tau \qquad \forall x\in[a,b], \quad (27)$$

where f' can be either D_-f or D_+f, see [78, Remark I.4.2.5]. Furthermore, [78, Theorem I.4.2.4] asserts that for any $a,b\in\mathtt{dom}(f)$ such that $a<b$, there exists a $x\in(a,b)$ such that

$$f(b)=f(a)+(b-a)f' \qquad f'\in\partial f(x). \quad (28)$$

For any function f satisfying Assumption 1, the convex conjugate f^* of f is defined as,

$$f^*(\lambda):=\sup\{\lambda x-f(x): x\in\mathtt{dom}(f)\} \qquad \forall\lambda\in\mathbb{R}, \quad (29)$$

see [78, §I.6].

B. The Semicontinuous Inverses of Monotonic Functions

Inverse of a function, in the usual sense, is defined only for bijective (i.e., one-to-one and onto) functions. For non-bijective monotone functions the generalized (i.e., semi-continuous) inverses constitute a partial remedy, see [79]–[85].

Definition 1. For any non-decreasing function $f:\mathbb{R}\mapsto\overline{\mathbb{R}}$ the u.s.c. (i.e., upper-semi-continuous) inverse $f^{\wedge}:\mathbb{R}\mapsto\overline{\mathbb{R}}$ and the l.s.c. (i.e., lower-semi-continuous) inverse $f^{\vee}:\mathbb{R}\mapsto\overline{\mathbb{R}}$ are defined as

$$f^{\vee}(y):=\inf\{x\in\mathbb{R}: f(x)>y\} \qquad \forall y\in\mathbb{R}, \quad (30a)$$
$$f^{\wedge}(y):=\inf\{x\in\mathbb{R}: f(x)\geq y\} \qquad \forall y\in\mathbb{R}. \quad (30b)$$

For any non-increasing function $g:\mathbb{R}\mapsto\overline{\mathbb{R}}$ the u.s.c. inverse $g^{\vee}:\mathbb{R}\mapsto\overline{\mathbb{R}}$ and the l.s.c. inverse $g^{\wedge}:\mathbb{R}\mapsto\overline{\mathbb{R}}$ are defined as

$$g^{\vee}(y):=\inf\{x\in\mathbb{R}: g(x)<y\} \qquad \forall y\in\mathbb{R}, \quad (31a)$$
$$g^{\wedge}(y):=\inf\{x\in\mathbb{R}: g(x)\leq y\} \qquad \forall y\in\mathbb{R}. \quad (31b)$$

Remark 1. Recall that $\inf\emptyset=\infty$, $\sup\emptyset=-\infty$, and the inequality $\inf\mathcal{A}\leq\sup\mathcal{A}$ holds only for $\mathcal{A}\neq\emptyset$ case; in that case $\inf\mathcal{A}\leq x\leq\sup\mathcal{A}$ for all $x\in\mathcal{A}$, as well.

Customarily, only the non-decreasing functions are studied because corresponding claims for the non-increasing ones are evident. In addition, the u.s.c. inverse $f^{\vee}$ is usually called the right continuous inverse and the l.s.c. inverse $f^{\wedge}$ is usually

called the left continuous inverse. Excellent overviews and derivations of some of the fundamental results and some applications can be found in [83]–[85].

The definitions above are describing an operation that is quite easy to grasp intuitively. Let $h:\mathbb{R}\mapsto\overline{\mathbb{R}}$ be a monotonic function. We extend the graph of $h(\cdot)$ by drawing vertical lines between $(x,\lim_{\gamma\uparrow x}h(\gamma))$ and $(x,\lim_{\gamma\downarrow x}h(\gamma))$ at the points where[10] the left limit $\lim_{\gamma\uparrow x}h(\gamma)$ is not equal to the right limit $\lim_{\gamma\downarrow x}h(\gamma)$; those lines are not present in the graph of $h(\cdot)$ to ensure that $h(x)$ has a single value for each $x\in\mathbb{R}$. Note that extended graph is connected. Then the inversion operation simply takes the mirror image of the extended graph with respect to $y=x$ line, i.e., with respect to the graph of the identity function, and replace any vertical lines in the mirror image (i.e., horizontal lines in the original graph) with upper most point for $h^{\vee}(\cdot)$ and the lower most point for $h^{\wedge}(\cdot)$. Then following three observations are evident from the description of the semicontinuous inversion given above:

- The semicontinuous inverses $h^{\vee}(\cdot)$ and $h^{\wedge}(\cdot)$ would be the same if replace $h(\cdot)$ with its lower semicontinuous closure $\underline{\mathtt{cl}h}(\cdot)$, or with its upper semicontinuous closure $\overline{\mathtt{cl}h}(\cdot)$, because $h(\cdot)$, $\underline{\mathtt{cl}h}(\cdot)$, and $\overline{\mathtt{cl}h}(\cdot)$ share the same the extended graph, where the lower semicontinuous closure $\underline{\mathtt{cl}h}$ of h and the upper semicontinuous closure $\overline{\mathtt{cl}h}$ of h are defined as
$$\underline{\mathtt{cl}h}(x):=(\lim_{\gamma\uparrow x}h(\gamma))\wedge(\lim_{\gamma\downarrow x}h(\gamma))\quad\forall x\in\mathbb{R},$$
$$\overline{\mathtt{cl}h}(x):=(\lim_{\gamma\uparrow x}h(\gamma))\vee(\lim_{\gamma\downarrow x}h(\gamma))\quad\forall x\in\mathbb{R},$$
see [77, Definition B.1.2.4] for a more general definition.
- $h^{\wedge}(\cdot)$ is lower semicontinuous on $\mathbb{R}$ and $h^{\vee}(\cdot)$ is upper semicontinuous on $\mathbb{R}$.
- The semicontinuous inversion is involutive in the sense that the u.s.c. inverses of both $h^{\wedge}(\cdot)$ and $h^{\vee}(\cdot)$ equal to $\overline{\mathtt{cl}h}(\cdot)$ and the l.s.c. inverses of both $h^{\wedge}(\cdot)$ and $h^{\vee}(\cdot)$ equal to $\underline{\mathtt{cl}h}(\cdot)$.

C. The Likelihood-Ratio Quantiles

Recall that the Lebesgue decomposition theorem holds not only for finite measures but also for σ-finite measures, see [87, Theorem 5.5.3] and [4, §32, Theorem C]. As a consequence, Radon-Nikodym Theorem for finite measures generalizes to the σ-finite measures, see [5, §3.2], [4, §31, Theorem B]. Thus for any two σ-finite measures w and ν satisfying $w\prec\nu$, there is a non-negative ν-measurable function $\frac{\mathrm{d}w}{\mathrm{d}\nu}$, called the Radon-Nikodym derivative of w with respect to ν.

Definition 2. For any pair of σ-finite measures (w,q) on the same measurable space, say $(\mathcal{Y},\mathcal{Y})$, the likelihood-ratio $\ell_{w,q}(\cdot)$ and the likelihood-ratio distribution $F_\ell\big(\cdot\big|_q^w\big)$ are defined as
$$\ell_{w,q}:=\begin{cases}\infty & \text{if } \frac{\mathrm{d}q}{\mathrm{d}(w+q)}=0,\\ \frac{\mathrm{d}w_\sim}{\mathrm{d}q} & \text{if } \frac{\mathrm{d}q}{\mathrm{d}(w+q)}>0,\end{cases}\tag{32a}$$
$$F_\ell\big(\tau\big|_q^w\big):=\int\mathbb{1}_{\{\ell_{w,q}\leq\tau\}}\mathrm{d}w\qquad\forall\tau\in\mathbb{R},\tag{32b}$$

[10]For any extended real valued monotonic function both the right limit and the left limit exist at every point in $\mathbb{R}$, by [86, Theorem 4.29].

where $w_\sim$ is the q-absolutely continuous component of w.

The likelihood-ratio distribution $F_\ell\big(\cdot\big|_q^w\big)$ is "insidiously self-referential" as [10, p. 57] deems entropy to be while discussing asymptotic equipartition property. Nevertheless, $F_\ell\big(\cdot\big|_q^w\big)$ is a well–defined monotonic function for any pair of σ-finite measures (w,q) on a common measurable space. Furthermore,
$$F_\ell\big(\tau\big|_q^{\lambda w}\big)=\lambda F_\ell\big(\tfrac{\tau}{\lambda}\big|_q^w\big)\qquad\forall\tau\in\mathbb{R},\tag{33a}$$
$$F_\ell\big(\tau\big|_{\lambda q}^w\big)=F_\ell\big(\lambda\tau\big|_q^w\big)\qquad\forall\tau\in\mathbb{R},\tag{33b}$$
where λ is positive constant, i.e., $\lambda\in(0,\infty)$.

If replace $w+q$ with any other σ-finite reference measure ν satisfying $(w+q)\prec\nu$, then resulting likelihood-ratio is equal to the one given in (32a) $(w+q)$-a.e.; hence the resulting $F_\ell\big(\cdot\big|_q^w\big)$ will be equal to the one given in (32b). Furthermore, with the understanding that $\frac{1}{0}$ and $\frac{1}{\infty}$ stand for ∞ and 0, respectively, we have
$$\ell_{q,w}=\tfrac{1}{\ell_{w,q}}\qquad\qquad(w+q)\text{-a.e..}\tag{34}$$

If w is a probability measure, then $F_\ell\big(\cdot\big|_q^w\big)$ is the cumulative distribution of the extended real valued random variable $\ell_{w,q}$ in the probability space $(\mathcal{Y},\mathcal{Y},w)$, as its name suggests. However, we do not assume either w or q to be a probability measure; we only assume both of them to be σ-finite measures. Note that the likelihood-ratio distribution $F_\ell\big(\cdot\big|_q^w\big)$ depends only on the absolutely continuous components of w and q in each other, as an immediate consequence of Definition 2:
$$F_\ell\big(\cdot\big|_q^w\big)=F_\ell\big(\cdot\big|_{q_\sim}^{w_\sim}\big),\tag{35}$$
where $q_\sim$ is the w-absolutely continuous component of q and $w_\sim$ is the q-absolutely component of w. Furthermore, for any σ-finite measure $\mu\prec q$ satisfying $\frac{\mathrm{d}\mu}{\mathrm{d}q}\leq 1$ $q_\sim$-a.e., the inequality $\ell_{w,\mu}\geq\ell_{w,q}$ holds q-almost-everywhere. Thus
$$F_\ell\big(\cdot\big|_\mu^w\big)\leq F_\ell\big(\cdot\big|_q^w\big)\qquad\forall\mu:\tfrac{\mathrm{d}\mu}{\mathrm{d}q}\leq 1\quad q_\sim\text{-a.e.}\tag{36}$$
We call the cases when the likelihood-ratio distribution is a function of the form $F_\ell\big(\cdot\big|_q^w\big):\mathbb{R}\mapsto\{0,\lambda\}$ to be the trivial case. The case when $F_\ell\big(\tau\big|_q^w\big)=0$ for all $\tau\in\mathbb{R}$ is called the singular case because it happens iff w and q are mutually singular, i.e., $w\perp q$. Recall that $w\perp q\Leftrightarrow\|w_\sim\|=0\Leftrightarrow\|q_\sim\|=0$. We call the case when the likelihood-ratio distribution is a surjective (onto) function of the form $F_\ell\big(\cdot\big|_q^w\big):\mathbb{R}\mapsto\{0,\lambda\}$ to be the trivial-non-singular case. Note that the trivial-non-singular case happens iff $w\not\perp q$ and $\frac{\mathrm{d}w_\sim}{\mathrm{d}q}=\zeta$ holds $w_\sim$-a.e. some $\zeta\in(0,\infty)$. Note that in the trivial-non-singular case $\|w_\sim\|=\lambda$ and $\|q_\sim\|=\lambda/\zeta$

Although the likelihood-ratio distribution $F_\ell\big(\cdot\big|_q^w\big)$ is well-defined, it might not have an inverse in the usual sense. Nevertheless, the monotonicity of $F_\ell\big(\cdot\big|_q^w\big)$ ensures that it has semicontinuous inverses, called likelihood-ratio quantiles.

Definition 3. For any pair of σ-finite measures (w,q) on the same measurable space, say $(\mathcal{Y},\mathcal{Y})$, the likelihood-ratio quantiles (LRQs) $F_\ell^{\wedge}\big(\cdot\big|_q^w\big):\mathbb{R}\mapsto\overline{\mathbb{R}}$ and $F_\ell^{\vee}\big(\cdot\big|_q^w\big):\mathbb{R}\mapsto\overline{\mathbb{R}}$ are defined as
$$F_\ell^{\wedge}\big(\varepsilon\big|_q^w\big):=\inf\big\{\gamma\in\mathbb{R}:F_\ell\big(\gamma\big|_q^w\big)\geq\varepsilon\big\}\qquad\forall\varepsilon\in\mathbb{R},\tag{37a}$$
$$F_\ell^{\vee}\big(\varepsilon\big|_q^w\big):=\inf\big\{\gamma\in\mathbb{R}:F_\ell\big(\gamma\big|_q^w\big)>\varepsilon\big\}\qquad\forall\varepsilon\in\mathbb{R}.\tag{37b}$$

Note that the likelihood-ratio quantiles are nothing but the semicontinuous inverses of likelihoods ratio distribution $F_\ell(\cdot|{}^{w}_{q})$; this is why we imitate the notation in [79], [84] and denote them with $F^{\wedge}_\ell(\cdot|{}^{w}_{q})$ and $F^{\vee}_\ell(\cdot|{}^{w}_{q})$.

As a result of (35), we have

$$F^{\vee}_\ell(\cdot|{}^{w}_{q}) = F^{\vee}_\ell(\cdot|{}^{w_\sim}_{q_\sim}) \quad \text{and} \quad F^{\wedge}_\ell(\cdot|{}^{w}_{q}) = F^{\wedge}_\ell(\cdot|{}^{w_\sim}_{q_\sim}). \tag{38}$$

Furthermore, evident inclusion relations for the sets we are talking the infimum over in (37) imply,

$$F^{\wedge}_\ell(\varepsilon|{}^{w}_{q}) \le F^{\vee}_\ell(\varepsilon|{}^{w}_{q}) \le F^{\wedge}_\ell(\varepsilon+\delta|{}^{w}_{q}) \qquad \forall \varepsilon\in\mathbb{R}, \delta>0. \tag{39}$$

Let the LRQ-halflines[11] $\mathcal{A}^{\wedge}_\ell(\varepsilon|{}^{w}_{q})$ and $\mathcal{A}^{\vee}_\ell(\varepsilon|{}^{w}_{q})$ be

$$\mathcal{A}^{\wedge}_\ell(\varepsilon|{}^{w}_{q}) := \{\gamma\in\mathbb{R} : F_\ell(\gamma|{}^{w}_{q}) \ge \varepsilon\} \qquad \forall \varepsilon\in\mathbb{R}, \tag{40a}$$
$$\mathcal{A}^{\vee}_\ell(\varepsilon|{}^{w}_{q}) := \{\gamma\in\mathbb{R} : F_\ell(\gamma|{}^{w}_{q}) > \varepsilon\} \qquad \forall \varepsilon\in\mathbb{R}. \tag{40b}$$

Then

$$\mathcal{A}^{\wedge}_\ell(\varepsilon+\delta|{}^{w}_{q}) \subset \mathcal{A}^{\vee}_\ell(\varepsilon|{}^{w}_{q}) \subset \mathcal{A}^{\wedge}_\ell(\varepsilon|{}^{w}_{q}) \qquad \forall \varepsilon\in\mathbb{R}, \delta>0. \tag{41}$$

Since $F_\ell(\cdot|{}^{w}_{q})$ is non-decreasing, both $\mathcal{A}^{\wedge}_\ell(\varepsilon|{}^{w}_{q})$ and $\mathcal{A}^{\vee}_\ell(\varepsilon|{}^{w}_{q})$ are halflines directed towards ∞. Thus

$$\inf \mathcal{A}^{\wedge}_\ell(\varepsilon|{}^{w}_{q}) = \sup(\mathbb{R}\setminus \mathcal{A}^{\wedge}_\ell(\varepsilon|{}^{w}_{q})), \tag{42a}$$
$$\inf \mathcal{A}^{\vee}_\ell(\varepsilon|{}^{w}_{q}) = \sup(\mathbb{R}\setminus \mathcal{A}^{\vee}_\ell(\varepsilon|{}^{w}_{q})). \tag{42b}$$

As a result of (37), (40), and (42), we have the following alternative characterization of LRQs

$$F^{\wedge}_\ell(\varepsilon|{}^{w}_{q}) = \sup\{\gamma\in\mathbb{R} : F_\ell(\gamma|{}^{w}_{q}) < \varepsilon\} \qquad \forall \varepsilon\in\mathbb{R}, \tag{43a}$$
$$F^{\vee}_\ell(\varepsilon|{}^{w}_{q}) = \sup\{\gamma\in\mathbb{R} : F_\ell(\gamma|{}^{w}_{q}) \le \varepsilon\} \qquad \forall \varepsilon\in\mathbb{R}. \tag{43b}$$

Lemma 1. *For any pair of σ-finite measures (w,q) on the same measurable space $(\mathcal{Y},\mathcal{Y})$, LRQs $F^{\wedge}_\ell(\cdot|{}^{w}_{q})$ and $F^{\vee}_\ell(\cdot|{}^{w}_{q})$, defined in* (37)*, are non-decreasing functions on $\mathbb{R}$ satisfying*

$$\lim_{\tau\uparrow\varepsilon} F^{-1}_\ell(\tau|{}^{w}_{q}) = F^{\wedge}_\ell(\varepsilon|{}^{w}_{q}) \qquad \forall \varepsilon\in\mathbb{R}, \tag{44a}$$
$$\lim_{\tau\downarrow\varepsilon} F^{-1}_\ell(\tau|{}^{w}_{q}) = F^{\vee}_\ell(\varepsilon|{}^{w}_{q}) \qquad \forall \varepsilon\in\mathbb{R}, \tag{44b}$$
$$w_\sim(\ell_{w,q} \le F^{\wedge}_\ell(\varepsilon|{}^{w}_{q})) \ge \varepsilon \qquad \forall \varepsilon \le \|w_\sim\|, \tag{44c}$$
$$w(\ell_{w,q} < F^{\vee}_\ell(\varepsilon|{}^{w}_{q})) \le \varepsilon \qquad \forall \varepsilon \ge 0, \tag{44d}$$
$$F^{\vee}_\ell(\varepsilon|{}^{w}_{q}) = F^{\wedge}_\ell(\varepsilon|{}^{w}_{q}) \qquad \forall \varepsilon\in\mathbb{R}\setminus\mathcal{D}, \tag{44e}$$
$$[F^{\wedge}_\ell(\varepsilon|{}^{w}_{q}), F^{\vee}_\ell(\varepsilon|{}^{w}_{q})] \subset (0,\infty) \qquad \forall \varepsilon\in(0,\|w_\sim\|), \tag{44f}$$

for some countable set $\mathcal{D}\subset[0,\|w_\sim\|]$, where $F^{-1}_\ell(\cdot|{}^{w}_{q})$ in (44a) *and* (44b) *can be either $F^{\wedge}_\ell(\cdot|{}^{w}_{q})$ or $F^{\vee}_\ell(\cdot|{}^{w}_{q})$.*

Proof of Lemma 1 is presented in Appendix A. Let us proceed with elaborating on some of the implications of Lemma 1, briefly. If $F^{\wedge}_\ell(\varepsilon|{}^{w}_{q}) \neq F^{\vee}_\ell(\varepsilon|{}^{w}_{q})$ for an $\varepsilon\in\mathbb{R}$, then $[F^{\wedge}_\ell(\varepsilon|{}^{w}_{q}), F^{\vee}_\ell(\varepsilon|{}^{w}_{q})) \neq \emptyset$ by (39). Furthermore, additivity of w implies

$$w(\ell_{w,q} \le F^{\wedge}_\ell(\varepsilon|{}^{w}_{q})) \le F_\ell(\gamma|{}^{w}_{q}) \le w(\ell_{w,q} < F^{\vee}_\ell(\varepsilon|{}^{w}_{q}))$$

for all γ in $[F^{\wedge}_\ell(\varepsilon|{}^{w}_{q}), F^{\vee}_\ell(\varepsilon|{}^{w}_{q}))$. Then (44c) and (44d) imply

$$F_\ell(\gamma|{}^{w}_{q}) = \varepsilon \quad \forall \varepsilon\in[0,\|w_\sim\|],\ \gamma\in[F^{\wedge}_\ell(\varepsilon|{}^{w}_{q}), F^{\vee}_\ell(\varepsilon|{}^{w}_{q})). \tag{45}$$

The behavior of $F^{\wedge}_\ell(\cdot|{}^{w}_{q})$ and $F^{\vee}_\ell(\cdot|{}^{w}_{q})$ are interesting only on $[0,\|w_\sim\|]$. What happens outside that range is determined by (39), (44a), (44b), and the values of $F^{\wedge}_\ell(\cdot|{}^{w}_{q})$ and $F^{\vee}_\ell(\cdot|{}^{w}_{q})$ on the boundary

$$F^{\wedge}_\ell(0|{}^{w}_{q}) = -\infty, \tag{46a}$$
$$F^{\vee}_\ell(0|{}^{w}_{q}) = \operatorname{ess\,inf}_{w_\sim} \ell_{w,q}, \tag{46b}$$
$$F^{\wedge}_\ell(\|w_\sim\||{}^{w}_{q}) = \operatorname{ess\,sup}_{w_\sim} \ell_{w,q}, \tag{46c}$$
$$F^{\vee}_\ell(\|w_\sim\||{}^{w}_{q}) = \infty. \tag{46d}$$

The effect of certain elementary modifications of w and q on the likelihood ratio distribution was described (33) and (36). Lemma 2, in the following, describes the effect of the very same modifications on the LRQs. For the case when w and q are probability measures, (47c) follows from similar identities reported before see [24, Lemma 3] and [25, Lemma 2.2].

Lemma 2. *For any pair of σ-finite measures (w,q) on the same measurable space $(\mathcal{Y},\mathcal{Y})$ the LRQs defined in* (37) *satisfy*

$$F^{-1}_\ell(\cdot|{}^{\lambda w}_{q}) = \lambda F^{-1}_\ell(\tfrac{\cdot}{\lambda}|{}^{w}_{q}) \qquad \forall \lambda\in(0,\infty), \tag{47a}$$
$$F^{-1}_\ell(\cdot|{}^{w}_{\lambda q}) = \tfrac{1}{\lambda} F^{-1}_\ell(\cdot|{}^{w}_{q}) \qquad \forall \lambda\in(0,\infty), \tag{47b}$$
$$F^{-1}_\ell(\cdot|{}^{w}_{\mu}) \ge F^{-1}_\ell(\cdot|{}^{w}_{q}) \qquad \forall \mu : \tfrac{\mathrm{d}\mu}{\mathrm{d}q} \le 1 \quad q_\sim\text{-}a.e, \tag{47c}$$

where $q_\sim$ is the w-absolutely component of q.

Each relation in Lemma 2 follows from the corresponding relation in (33) or (36), as an immediate consequence of the definition of LRQs in (37).

D. The Primitive Entropy Spectrum

Recall that for any pair of σ-finite measures (w,q) on the same measurable space, say $(\mathcal{Y},\mathcal{Y})$, the primitive entropy spectrum of w with respect to q is defined in (9) as

$$\hbar(\gamma|{}^{w}_{q}) = \int \left(\tfrac{\mathrm{d}w_\sim}{\mathrm{d}q} \wedge \gamma\right)\mathrm{d}q \qquad \forall \gamma\in[0,\infty). \tag{48a}$$

Thus for all σ-finite measures w and q, we have

$$\hbar(\gamma|{}^{w}_{q}) = \inf_{\mathcal{E}\in\mathcal{Y}} w(\mathcal{E}) + \gamma q(\mathcal{Y}\setminus\mathcal{E}) \qquad \forall \gamma\in[0,\infty), \tag{48b}$$
$$\hbar(0|{}^{w}_{q}) = 0, \tag{48c}$$
$$\hbar(\gamma|{}^{w}_{q}) = \hbar(\gamma|{}^{w_\sim}_{q_\sim}) \qquad \forall \gamma\in[0,\infty), \tag{48d}$$
$$\hbar(\gamma|{}^{w}_{q}) = w(\ell_{w,q}\le\gamma) + \gamma q(\ell_{w,q}>\gamma) \qquad \forall \gamma\in(0,\infty), \tag{48e}$$
$$\hbar(\gamma|{}^{w}_{q}) = \gamma\hbar(1/\gamma|{}^{q}_{w}) \qquad \forall \gamma\in(0,\infty), \tag{48f}$$

Note that $\hbar(\gamma|{}^{w}_{q})$ is non-decreasing in γ. Thus (48f) implies

$$(1\vee\gamma)\hbar(1|{}^{w}_{q}) \ge \hbar(\gamma|{}^{w}_{q}) \ge (1\wedge\gamma)\hbar(1|{}^{w}_{q}) \quad \forall \gamma\in(0,\infty). \tag{49}$$

Thus for any $\gamma\in(0,\infty)$, we have

$$\hbar(1|{}^{w}_{q}) = 0 \iff \hbar(\gamma|{}^{w}_{q}) = 0, \tag{50a}$$
$$\hbar(1|{}^{w}_{q}) \in (0,\infty) \iff \hbar(\gamma|{}^{w}_{q}) \in (0,\infty), \tag{50b}$$
$$\hbar(1|{}^{w}_{q}) = \infty \iff \hbar(\gamma|{}^{w}_{q}) = \infty. \tag{50c}$$

We deem the case when $\hbar(\gamma|{}^{w}_{q}) = \infty$ for all $\gamma\in(0,\infty)$ to be the void case for primitive entropy spectrum $\hbar(\cdot|{}^{w}_{q})$. Thus a necessary and sufficient condition for the non-void cases for $\hbar(\cdot|{}^{w}_{q})$ is (7), i.e., $\hbar(1|{}^{w}_{q}) < \infty$. On the other hand the singular case $w \perp q$ corresponds to $\hbar(1|{}^{w}_{q}) = 0$ case. More generally in the trivial case the entropy spectrum has the following simple expression $\hbar(\gamma|{}^{w}_{q}) = \|w_\sim\| \wedge (\gamma\|q_\sim\|)$ for all $\gamma\in(0,\infty)$.

[11] We call $\mathcal{A}^{\wedge}_\ell(\varepsilon|{}^{w}_{q})$ and $\mathcal{A}^{\vee}_\ell(\varepsilon|{}^{w}_{q})$ halflines with the understanding that $\emptyset$ and $\mathbb{R}$ are halflines, as well.

Using the level set integration formula given in (11), we can express $\hbar\big(\gamma\big|{}^{w}_{q}\big)$ as Riemann integrals in terms of likelihood ratio distributions. To see why let us first express $w(\ell_{w,q}\leq\gamma)$ as such an integral.

$$\begin{aligned} w(\ell_{w,q}\leq\gamma) &\overset{(a)}{=} \int \mathbb{1}_{\{\ell_{w,q}\leq\gamma\}}\ell_{w,q}\mathrm{d}q, \\ &\overset{(b)}{=} \int_0^\infty q\big(\mathbb{1}_{\{\ell_{w,q}\leq\gamma\}}\ell_{w,q}>\tau\big)\mathrm{d}\tau, \\ &\overset{(c)}{=} \int_0^\gamma q\big(\mathbb{1}_{\{\ell_{w,q}\leq\gamma\}}\ell_{w,q}>\tau\big)\mathrm{d}\tau, \\ &= \int_0^\gamma q(\gamma\geq\ell_{w,q}>\tau)\mathrm{d}\tau, \end{aligned}$$

where (a) follows from (32); (b) follows from (11); (c) holds because $q\big(\mathbb{1}_{\{\ell_{w,q}\leq\gamma\}}\ell_{w,q}>\tau\big)=0$ for $\tau\geq\gamma$. Then (48e) implies

$$\hbar\big(\gamma\big|{}^{w}_{q}\big) = \int_0^\gamma q(\ell_{w,q}>\tau)\mathrm{d}\tau \qquad \forall\gamma\in(0,\infty). \tag{51a}$$

Then (48f) implies

$$\hbar\big(\gamma\big|{}^{w}_{q}\big) = \gamma\int_0^{1/\gamma} w(\ell_{q,w}>s)\mathrm{d}s \qquad \forall\gamma\in(0,\infty).$$

Then applying the change of variable $s=\tau/\gamma$, we get

$$\hbar\big(\gamma\big|{}^{w}_{q}\big) = \int_0^1 w\Big(\ell_{q,w}>\tfrac{\tau}{\gamma}\Big)\mathrm{d}\tau \qquad \forall\gamma\in(0,\infty). \tag{51b}$$

Note that integrands $q(\ell_{w,q}>\tau)$ and $w(\ell_{q,w}>\tau/\gamma)$ in (51a) and (51b) can be replaced by $q(\ell_{w,q}\geq\tau)$ and $w(\ell_{q,w}\geq\tau/\gamma)$ because monotonic functions have at most countably many discontinuities by [86, Theorem 4.30] and any countable set has zero Lebesgue measure. Thus using (32) and (34), we get

$$\hbar\big(\gamma\big|{}^{w}_{q}\big) = \int_0^\gamma F_\ell\big(\tfrac{1}{\tau}\big|{}^{q}_{w}\big)\mathrm{d}\tau \qquad \forall\gamma\in(0,\infty), \tag{51c}$$

$$\hbar\big(\gamma\big|{}^{w}_{q}\big) = \int_0^1 F_\ell\big(\tfrac{\gamma}{\tau}\big|{}^{w}_{q}\big)\mathrm{d}\tau \qquad \forall\gamma\in(0,\infty). \tag{51d}$$

Lemma 3. *If a pair of σ-finite measures (w,q) satisfies* (7)*, i.e., $\hbar\big(1\big|{}^{w}_{q}\big)<\infty$, then the primitive entropy spectrum $\hbar\big(\cdot\big|{}^{w}_{q}\big)$ satisfies*

$$\lim_{\gamma\downarrow0}\hbar\big(\gamma\big|{}^{w}_{q}\big) = \hbar\big(0\big|{}^{w}_{q}\big), \tag{52a}$$

$$\lim_{\gamma\uparrow\infty}\hbar\big(\gamma\big|{}^{w}_{q}\big) = \|w_\sim\|, \tag{52b}$$

$$\mathrm{D}_-\hbar\big(\gamma\big|{}^{w}_{q}\big) = q(\ell_{w,q}\geq\gamma) \qquad \forall\gamma\in(0,\infty), \tag{52c}$$

$$\mathrm{D}_+\hbar\big(\gamma\big|{}^{w}_{q}\big) = q(\ell_{w,q}>\gamma) \qquad \forall\gamma\in[0,\infty), \tag{52d}$$

where $\mathrm{D}_-\hbar\big(\gamma\big|{}^{w}_{q}\big)$ and $\mathrm{D}_+\hbar\big(\gamma\big|{}^{w}_{q}\big)$ are left and right derivatives of $\hbar\big(\cdot\big|{}^{w}_{q}\big)$ at γ. Hence, either $\hbar\big(\cdot\big|{}^{w}_{q}\big)=0$, or $\hbar\big(\cdot\big|{}^{w}_{q}\big)$ is a non-decreasing, finite valued, continuous, and concave function, which is monotonically increasing on $\gamma\in[0,\operatorname{ess\,sup}_{w_\sim}\ell_{w,q}]$.

Lemma 4. *Let w_0, w_1, q_0, q_1 be σ-finite measures with a common σ-finite reference measure ν and γ_0, γ_1 be non-negative real numbers. If both $\frac{\mathrm{d}w_0}{\mathrm{d}\nu}\leq\frac{\mathrm{d}w_1}{\mathrm{d}\nu}$ and $\gamma_0\frac{\mathrm{d}q_0}{\mathrm{d}\nu}\leq\gamma_1\frac{\mathrm{d}q_1}{\mathrm{d}\nu}$ holds ν-a.e. then*

$$\hbar\big(\gamma_0\big|{}^{w_0}_{q_0}\big)\leq\hbar\big(\gamma_1\big|{}^{w_1}_{q_1}\big). \tag{53a}$$

If $\gamma_\alpha=\alpha\gamma_1+(1-\alpha)\gamma_0$, $w_\alpha=\alpha w_1+(1-\alpha)w_0$, and $q_\alpha=\frac{\alpha\gamma_1}{\gamma_\alpha}q_1+\frac{(1-\alpha)\gamma_0}{\gamma_\alpha}q_0$, for some $\alpha\in[0,1]$, then

$$\alpha\hbar\big(\gamma_1\big|{}^{w_1}_{q_1}\big)+(1-\alpha)\hbar\big(\gamma_0\big|{}^{w_0}_{q_0}\big)\leq\hbar\big(\gamma_\alpha\big|{}^{w_\alpha}_{q_\alpha}\big). \tag{53b}$$

Hence $\hbar\big(\gamma\big|{}^{w}_{q}\big)$ is non-decreasing and concave in the pair $(w,\gamma q)$. Furthermore, for any positive real λ, i.e., $\lambda\in(0,\infty)$,

$$\hbar\big(\gamma\big|{}^{\lambda w}_{q}\big)=\lambda\hbar\big(\tfrac{\gamma}{\lambda}\big|{}^{w}_{q}\big) \qquad \forall\gamma\in[0,\infty), \tag{53c}$$

$$\hbar\big(\gamma\big|{}^{w}_{\lambda q}\big)=\hbar\big(\lambda\gamma\big|{}^{w}_{q}\big) \qquad \forall\gamma\in[0,\infty). \tag{53d}$$

Proof of Lemmas 3 and 4 is presented in Appendix B.

III. The Spectral Representations and A Measure Change Lemma For Simple Hypothesis Testing

Our primary goal in this section is to establishing a spectral Neyman–Pearson lemma relating $\beta(\cdot|{}^{q}_{w})$ to $\hbar(\cdot|{}^{q}_{w})$, $F^{\wedge}_{\ell}\big(\cdot\big|{}^{w}_{q}\big)$, and $F^{\vee}_{\ell}\big(\cdot\big|{}^{w}_{q}\big)$, see Theorem 1, and a measure change lemma for the simple hypothesis testing problem expressing $\beta(\cdot|{}^{q}_{w})$ in terms of a $\beta(\cdot|{}^{q}_{v})$ for a σ-finite measure v satisfying certain conditions, see Theorem 2. To that end let us first point out certain elementary facts about rNP-function $\beta(\cdot|{}^{q}_{w})$ and dNP-function $\beta(\cdot\|{}^{q}_{w})$ defined in (6). Note that

$$\beta(\cdot|{}^{q}_{w})\leq\beta(\cdot\|{}^{q}_{w}) \qquad \forall w,q, \tag{54}$$

because any deterministic detector is also randomized detector, i.e., $\mathcal{T}^d(\mathcal{Y})\subset\mathcal{T}(\mathcal{Y})$. Definitions in (6) follow the framework of [6] and [7], allowing q's that are not probability measures and parameterizing $\beta(\cdot|{}^{q}_{w})$ with ε rather than $1-\varepsilon$, unlike the majority of the recent work on the subject [19]–[25], i.e., $\beta(\varepsilon|{}^{q}_{w})=\beta_{1-\varepsilon}(w\|q)$. Extension to the case when w is a finite measure rather than a probability measures were already mentioned explicitly in [6] and q being a σ-finite measure is implicitly allowed in both [6] and [7].

Allowing both w and q to be σ-finite measures, however, introduces a nuance one needs to handle; that is for certain w and q choices $\beta(\varepsilon|{}^{q}_{w})$ is infinite for all finite ε values, as we have mention in §I. We deem this case to be the void case for $\beta(\cdot|{}^{q}_{w})$. Recall that we have called the case when $\hbar\big(1\big|{}^{w}_{q}\big)=\infty$ to be the void case for $\hbar\big(\cdot\big|{}^{w}_{q}\big)$ because in that case $\hbar\big(\gamma\big|{}^{w}_{q}\big)=\infty$ for all positive γ. On the other hand, definitions of $\beta(\cdot|{}^{q}_{w})$ and $\hbar\big(\cdot\big|{}^{w}_{q}\big)$ in (6) and (9) imply

$$\beta(\varepsilon|{}^{q}_{w})+\varepsilon\geq\hbar\big(1\big|{}^{w}_{q}\big) \qquad \forall\varepsilon\geq0. \tag{55}$$

Thus whenever the condition in (7) is violated, i.e., whenever $\hbar\big(1\big|{}^{w}_{q}\big)=\infty$, we are in the void case for $\beta(\cdot|{}^{q}_{w})$, as well. When (7) holds, i.e., $\hbar\big(1\big|{}^{w}_{q}\big)<\infty$, the function $\beta(\cdot|{}^{q}_{w})$ is finite on $(0,\infty)$, as we will see in Theorem 1.

Note that $\beta\big(\cdot\big|{}^{q_\sim}_{w_\sim}\big)\leq\beta(\cdot|{}^{q}_{w})$ because $\int\mathsf{T}\mathrm{d}w_\sim\leq\int\mathsf{T}\mathrm{d}w$ and $\int(1-\mathsf{T})\mathrm{d}q_\sim\leq\int(1-\mathsf{T})\mathrm{d}q$ for all $\mathsf{T}\in\mathcal{T}(\mathcal{Y})$. On the other hand, $\beta(\cdot|{}^{q}_{w})\leq\beta\big(\cdot\big|{}^{q_\sim}_{w_\sim}\big)$ because we can set $\mathsf{T}=0$ whenever $\ell_{w,q}=\infty$ to ensure $\int\mathsf{T}\mathrm{d}w_\sim=\int\mathsf{T}\mathrm{d}w$ and $\mathsf{T}=1$ whenever $\ell_{w,q}=0$, to ensure $\int(1\text{–}\mathsf{T})\mathrm{d}q_\sim=\int(1\text{–}\mathsf{T})\mathrm{d}q$, without changing the values of $\int\mathsf{T}\mathrm{d}w_\sim$ and $\int(1\text{–}\mathsf{T})\mathrm{d}q_\sim$. Same argument applies to $\beta(\cdot\|{}^{q}_{w})$, as well. Thus

$$\beta(\cdot|{}^{q}_{w})=\beta\big(\cdot\big|{}^{q_\sim}_{w_\sim}\big), \tag{56a}$$

$$\beta(\cdot\|{}^{q}_{w})=\beta\big(\cdot\big\|{}^{q_\sim}_{w_\sim}\big). \tag{56b}$$

Consequently, behavior of $\beta(\cdot\|^q_w)$ and $\beta(\cdot|^q_w)$ are interesting only on $[0,\|w_\sim\|]$. What happens outside that range is determined by the monotonicity of $\beta(\cdot\|^q_w)$ and $\beta(\cdot|^q_w)$ and the values on the boundary:

$$\beta(\varepsilon\|^q_w) = \beta(\varepsilon|^q_w) = \infty \qquad \forall \varepsilon < 0, \tag{56c}$$
$$\beta(0\|^q_w) = \beta(0|^q_w) = \|q_\sim\|, \tag{56d}$$
$$\beta(\varepsilon\|^q_w) = \beta(\varepsilon|^q_w) = 0 \qquad \forall \varepsilon \geq \|w_\sim\|, \tag{56e}$$

where $\|q_\sim\|$ can be infinite for the case when q is a σ-finite measure. When q is a finite measure both $\beta(\cdot\|^q_w)$ and $\beta(\cdot|^q_w)$ have a discontinuity at zero, because $\|q_\sim\| \leq \|q\|$.

Furthermore, the effect of certain elementary modifications of w and q on $\beta(\cdot|^q_w)$ and $\beta(\cdot\|^q_w)$ given below follow from definitions in (6) and evident pointwise relation that are valid for all $\mathsf{T} \in \mathcal{T}(\mathcal{Y})$.

Lemma 5. *For any pair of σ-finite measures (w,q) on the same measurable space $(\mathcal{Y},\mathcal{Y})$, $\beta(\cdot|^q_w)$ defined in* (6b) *satisfies*

$$\beta\big(\cdot\big|^{\lambda q}_w\big) = \lambda\beta(\cdot|^q_w) \qquad \forall \lambda \in (0,\infty), \tag{57a}$$
$$\beta(\cdot|^q_{\lambda w}) = \beta\big(\tfrac{\cdot}{\lambda}\big|^q_w\big) \qquad \forall \lambda \in (0,\infty), \tag{57b}$$
$$\beta(\cdot|^\mu_w) \leq \beta(\cdot|^q_w) \qquad \forall \mu : \tfrac{\mathrm{d}\mu}{\mathrm{d}q} \leq 1 \quad q_\sim\text{-a.e}, \tag{57c}$$
$$\beta\big(\cdot\big|^q_\mu\big) \geq \beta(\cdot|^q_w) \qquad \forall \mu : \tfrac{\mathrm{d}w_\sim}{\mathrm{d}\mu} \leq 1 \quad \mu\text{-a.e}, \tag{57d}$$

where $w_\sim$ is the q-absolutely component of w and $q_\sim$ is the w-absolutely component of q. Furthermore, (57) *hold as is if $\beta(\cdot|\cdot)$'s are replaced by $\beta(\cdot\|\cdot)$'s.*

In the singular case, i.e., when $w \perp q$, NP-functions satisfy $\beta(\varepsilon\|^q_w) = \beta(\varepsilon|^q_w) = 0$ for all $\varepsilon \geq 0$. Note that a trivial-non-singular case can simultaneously be a void case. In the trivial-non-singular-non-void case, rNP-function has the following closed form expression

$$\beta(\varepsilon|^q_w) = \|q_\sim\|\left|1 - \tfrac{\varepsilon}{\|w_\sim\|}\right|^+ \qquad \forall \varepsilon \geq 0. \tag{58}$$

In the following, we will primarily focus on rNP-function $\beta(\cdot|^q_w)$, nevertheless it is worth mentioning certain intuitive but subtle observations about dNP-function $\beta(\cdot\|^q_w)$ and rNP-function $\beta(\cdot|^q_w)$. If $w_\sim$ (and hence $q_\sim$) does not have any atoms[12] then $\beta(\cdot\|^q_w) = \beta(\cdot|^q_w)$. If the likelihood-ratio distribution $F_\ell\big(\cdot\big|^w_q\big)$ has a discontinuity at a value of $\ell_{w,q}$ and if there are only finitely many outputs y taking that value, i.e., there exists a γ satisfying both

$$\lim_{\tau\uparrow\gamma} F_\ell\big(\tau\big|^w_q\big) \neq F_\ell\big(\gamma\big|^w_q\big) \qquad \text{and} \qquad \big|\mathcal{B}\big(\gamma|^w_q\big)\big| < \infty,$$

where $\mathcal{B}\big(\gamma|^w_q\big) = \{y \in \mathcal{Y} : \ell_{w,q} = \gamma\}$. Then the inequality in (54) is strict for some ε. If $\mathcal{B}\big(\gamma|^w_q\big)$ is not a finite set, then the issue is more nuanced, e.g., (54) can be satisfied as an equality even for the discrete case. Let both w and q be the geometric probability mass function with parameter $1/2$ supported on positive integers. Then $\beta(\varepsilon|^q_w) = \beta(\varepsilon\|^q_w) = |1-\varepsilon|^+$ for all $\varepsilon \geq 0$. See Remark 2, for a worst-case bound on $\beta(\varepsilon\|^q_w) - \beta(\varepsilon|^q_w)$.

[12] An event $\mathcal{E} \in \mathcal{Y}$ is an atom of $w_\sim$ iff $\mathcal{E}$ satisfies both $w_\sim(\mathcal{E}) > 0$ and the condition $w_\sim(\mathcal{E} \cap \mathcal{S}) \in \{0, w_\sim(\mathcal{E})\}$ for all $\mathcal{S} \in \mathcal{Y}$.

A. Likelihood-Ratio Tests

Lemma 6. *For any pair of σ-finite measures (w,q) satisfying* (7)*, $\varepsilon \in (0,\|w_\sim\|)$, and $\gamma \in \Big[1/F^\vee_\ell\big(\varepsilon\big|^w_q\big), 1/F^\wedge_\ell\big(\varepsilon\big|^w_q\big)\Big]$, let $\mathsf{T}^{q,w}_{\varepsilon,\gamma}$ be*

$$\mathsf{T}^{q,w}_{\varepsilon,\gamma} := \begin{cases} 1 & \text{if } \ell_{q,w} > \gamma, \\ \frac{\varepsilon - w(\ell_{q,w} > \gamma)}{w(\ell_{q,w} = \gamma)} & \text{if } \ell_{q,w} = \gamma \text{ and } w(\ell_{q,w} = \gamma) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{59}$$

Then $\mathsf{T}^{q,w}_{\varepsilon,\gamma}$ is a randomized likelihood-ratio test, $\mathsf{T}^{q,w}_{\varepsilon,\gamma} \in \mathcal{T}(\mathcal{Y})$. Furthermore, $\mathsf{T}^{q,w}_{\varepsilon,\gamma}$ satisfies

$$\int \mathsf{T}^{q,w}_{\varepsilon,\gamma}\,\mathrm{d}w = \varepsilon, \tag{60a}$$
$$\int (1 - \mathsf{T}^{q,w}_{\varepsilon,\gamma})\,\mathrm{d}q = \hbar(\gamma|^q_w) - \gamma\varepsilon. \tag{60b}$$

Proof. If $w \perp q$, then $\|w_\sim\| = 0$ and the lemma is void. If $w \not\perp q$ then and $\varepsilon \in (0,\|w_\sim\|)$, then $\big[F^\wedge_\ell\big(\varepsilon\big|^w_q\big), F^\vee_\ell\big(\varepsilon\big|^w_q\big)\big]$ is a non-empty subset of $(0,\infty)$ by Lemma 1. Thus (44c), (44d), and (44f) imply

$$w(\ell_{w,q} < 1/\gamma) \leq \varepsilon \leq w(\ell_{w,q} \leq 1/\gamma).$$

Thus using (34) and $\big[F^\wedge_\ell\big(\varepsilon\big|^w_q\big), F^\vee_\ell\big(\varepsilon\big|^w_q\big)\big] \subset (0,\infty)$ we get

$$0 \leq \varepsilon - w(\ell_{q,w} > \gamma) \leq w(\ell_{q,w} = \gamma).$$

Then $\mathsf{T}^{q,w}_{\varepsilon,\gamma} \in [0,1]$ and $\mathsf{T}^{q,w}_{\varepsilon,\gamma} \in \mathcal{T}(\mathcal{Y})$. Furthermore,

$$\begin{aligned} \int \mathsf{T}^{q,w}_{\varepsilon,\gamma}\,\mathrm{d}w &\overset{(a)}{=} (\varepsilon - w(\ell_{q,w} > \gamma)) + \int \mathbb{1}_{\{\ell_{q,w} > \gamma\}}\,\mathrm{d}w, \\ &= \varepsilon. \end{aligned}$$
$$\begin{aligned} \int (1 - \mathsf{T}^{q,w}_{\varepsilon,\gamma})\,\mathrm{d}q &\overset{(b)}{=} (w(\ell_{q,w} \geq \gamma) - \varepsilon)\gamma + \int \mathbb{1}_{\{\ell_{q,w} < \gamma\}}\,\mathrm{d}q, \\ &\overset{(c)}{=} \hbar(\gamma|^q_w) - \gamma\varepsilon, \end{aligned}$$

where (a) and (b) follow the definitions from (32) and (59); (c) follows (48e), with the roles of w and q interchanged. □

B. A Spectral Neyman–Pearson Lemma

Theorem 1. *If a pair of σ-finite measures (w,q) satisfies* (7)*, i.e., $\hbar\big(1\big|^w_q\big) < \infty$, then the restriction of $\beta(\cdot|^q_w)$ to $[0,\infty)$ is a non-increasing, convex, and continuous function satisfying*

$$\beta(\varepsilon|^q_w) = \sup_{\gamma \geq 0} \hbar(\gamma|^q_w) - \gamma\varepsilon \qquad \forall \varepsilon \in \mathbb{R}, \tag{61a}$$
$$= \hbar(\gamma_\varepsilon|^q_w) - \gamma_\varepsilon\varepsilon \qquad \forall \varepsilon > 0, \tag{61b}$$
$$= \int_0^{\gamma_\varepsilon} \Big[w\Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w} > \tau\Big) - \varepsilon\Big]\,\mathrm{d}\tau \qquad \forall \varepsilon \geq 0, \tag{61c}$$
$$= \int_0^{\infty} \Big|w\Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w} > \tau\Big) - \varepsilon\Big|^+ \mathrm{d}\tau \qquad \forall \varepsilon \in \mathbb{R}, \tag{61d}$$
$$\mathrm{D}_-\beta(\varepsilon|^q_w) = \tfrac{-1}{F^\wedge_\ell\left(\varepsilon\right|^w_q)} \qquad \forall \varepsilon > 0, \tag{61e}$$
$$\mathrm{D}_+\beta(\varepsilon|^q_w) = \tfrac{-1}{F^\vee_\ell\left(\varepsilon\right|^w_q)} \qquad \forall \varepsilon \geq 0, \tag{61f}$$

for any $\gamma_\varepsilon \in \left[\frac{1}{F^\vee_\ell\left(\varepsilon\right|^w_q)}, \frac{1}{(F^\wedge_\ell\left(\varepsilon\right|^w_q) \vee 0)}\right]$. Furthermore, the restriction of $\beta(\cdot|^q_w)$ to $(0,\|w_\sim\|)$ is an invertible function with the

image $(0,\|q_\sim\|)$ *and there exists a* $\mathsf{T}_\varepsilon\in\mathcal{T}(\mathcal{Y})$ *satisfying both* $\int \mathsf{T}_\varepsilon \mathrm{d}w=\varepsilon$ *and* $\int(1-\mathsf{T}_\varepsilon)\mathrm{d}q=\beta(\varepsilon|^q_w)$ *for all* $\varepsilon\in[0,\|w_\sim\|]$.

$$\beta(\varepsilon|^q_w)=\int_\varepsilon^{\|w_\sim\|} g(\tau)\mathrm{d}\tau \qquad \forall\varepsilon\in[0,\|w_\sim\|], \tag{61g}$$

$$\partial\beta(\varepsilon|^q_w)\subset(-\infty,0) \qquad \forall\varepsilon\in(0,\|w_\sim\|), \tag{61h}$$

for any $g:(0,\|w_\sim\|)\mapsto(0,\infty)$ *s.t.* $\frac{1}{F^\vee_\ell(\cdot|^w_q)}\le g(\cdot)\le\frac{1}{F^\wedge_\ell(\cdot|^w_q)}$. *If* $q_\sim$ *is a finite measure, i.e.,*$\|q_\sim\|<\infty$, *then*

$$\beta(\varepsilon|^q_w)=\|q_\sim\|-\int_0^\varepsilon g(\tau)\mathrm{d}\tau \qquad \forall\varepsilon\ge 0. \tag{61j}$$

Furthermore, (61a) *and* (61d) *hold even when* (7) *is violated, i.e., when* $\hbar\big(1\big|^w_q\big)=\infty$.

In (61c) and (61d), the '$>$' signs can be replaced by '$\ge$' signs because monotonic functions have at most countably many discontinuities by [86, Theorem 4.30] and any countable set has zero Lebesgue measure. Hence (32) and (34) imply

$$\beta(\varepsilon|^q_w)=\int_0^{\gamma_\varepsilon}\big[F_\ell\big(\tfrac{1}{\tau}\big|^w_q\big)-\varepsilon\big]\mathrm{d}\tau \qquad \forall\varepsilon\ge 0, \tag{62a}$$

$$=\int_0^{\infty}\big|F_\ell\big(\tfrac{1}{\tau}\big|^w_q\big)-\varepsilon\big|^+\mathrm{d}\tau \qquad \forall\varepsilon\in\mathbb{R}. \tag{62b}$$

As an immediate consequence of Theorem 1, we can express the convex conjugate $\beta^*(\cdot|^q_w)$ of $\beta(\cdot|^q_w)$, as defined in (29), in terms of the primitive entropy spectrum, as follows

$$\beta^*(\lambda|^q_w)=\begin{cases}-\hbar(-\lambda|^q_w) & \text{if } \lambda\le 0,\\ \infty & \text{if } \lambda>0.\end{cases} \tag{63}$$

On the other hand, the lower semicontinuous inverse $\beta^\wedge(\cdot|^q_w)$ is defined via (31b), because $\beta(\cdot|^q_w)$ is non-increasing. $\beta^\wedge(\cdot|^q_w)$ can be determined using injectivity of $\beta(\cdot|^q_w)$ on $[0,\|w_\sim\|]$ and (56), to be[13]

$$\beta^\wedge(\cdot|^q_w)=\beta\big(\cdot\big|^w_q\big), \tag{64}$$

where $\beta\big(\cdot\big|^w_q\big)$ is defined via (6b), as well. Thus[14]

$$\beta\big(\beta(\varepsilon|^q_w)\big|^w_q\big)=\varepsilon \qquad \forall\varepsilon\in(0,\|w_\sim\|), \tag{65a}$$

$$\beta\big(\beta\big(\tau\big|^w_q\big)\big|^q_w\big)=\tau \qquad \forall\tau\in(0,\|q_\sim\|). \tag{65b}$$

That is the rNP-function $\beta(\cdot|^q_w):(0,\|w_\sim\|)\mapsto(0,\|q_\sim\|)$ and the rNP-function$\beta\big(\cdot\big|^w_q\big):(0,\|q_\sim\|)\mapsto(0,\|w_\sim\|)$ are inverses of one another.

If the likelihood-ratio distribution $F_\ell\big(\cdot\big|^w_q\big)$ is constant on a subinterval of $\big(F^\vee_\ell\big(0\big|^w_q\big),F^\wedge_\ell\big(\|w_\sim\|\big|^w_q\big)\big)$, then the restriction of $F_\ell\big(\cdot\big|^w_q\big)$ to $\big(F^\vee_\ell\big(0\big|^w_q\big),F^\wedge_\ell\big(\|w_\sim\|\big|^w_q\big)\big)$ is not invertible and the left and right derivatives of $\beta(\cdot|^q_w)$ differ at some point in $(0,\|w_\sim\|)$, as a result of (61e) and (61f). On the other hand, any discontinuities of $F_\ell\big(\cdot\big|^w_q\big)$ in $\big(F^\vee_\ell\big(0\big|^w_q\big),F^\wedge_\ell\big(\|w_\sim\|\big|^w_q\big)\big)$, corresponds to a subinterval of $(0,\|w_\sim\|)$ on which $F^\wedge_\ell\big(\cdot\big|^w_q\big)$ and $F^\vee_\ell\big(\cdot\big|^w_q\big)$ are equal to the same constant and $\beta(\cdot|^q_w)$ is an affine function on the said interval by (61e) and (61f).

[13]One can define upper semicontinuous inverse of $\beta\big(\cdot\big|^q_w\big)$ using (31a), as well. The value of semicontinuous inverses may differ only at zero: the upper-semicontinuous inverse will have the value ∞, and the lower-semicontinuous will have the value $\|w_\sim\|$ at zero.

[14]If $\|w_\sim\|<\infty$, then $\|w_\sim\|$ in (65a) and 0 in (65b) can be included in the intervals. If $\|q_\sim\|<\infty$, then 0 in (65a) and $\|q_\sim\|$ in (65b) can be included in the intervals.

Proof Theorem 1. Let us first prove a lower bound on $\beta(\cdot|^q_w)$

$$\begin{aligned}\beta(\varepsilon|^q_w)&\overset{(a)}{=}\inf_{\mathsf{T}\in\mathcal{T}(\mathcal{Y})}\sup_{\gamma\ge0}\int(1-\mathsf{T})\mathrm{d}q+\gamma\bigg(\int\mathsf{T}\mathrm{d}w-\varepsilon\bigg),\\ &\overset{(b)}{\ge}\sup_{\gamma\ge0}\inf_{\mathsf{T}\in\mathcal{T}(\mathcal{Y})}\int(1-\mathsf{T})\mathrm{d}q+\gamma\bigg(\int\mathsf{T}\mathrm{d}w-\varepsilon\bigg),\\ &\overset{(c)}{\ge}\sup_{\gamma\ge0}-\gamma\varepsilon+\inf_{\mathsf{T}\in\mathcal{T}(\mathcal{Y})}\int\Big[(1-\mathsf{T})\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w}+\gamma\mathsf{T}\Big]\mathrm{d}w,\\ &\overset{(d)}{=}\sup_{\gamma\ge0}-\gamma\varepsilon+\int\Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w}\wedge\gamma\Big)\mathrm{d}w,\\ &=\sup_{\gamma\ge0}\hbar(\gamma|^q_w)-\gamma\varepsilon.\end{aligned} \tag{66}$$

where (a) follows from (6b); (b) follows from max-min inequality; (c) holds because both $\mathsf{T}\le 1$ and $\frac{\mathrm{d}q_\sim}{\mathrm{d}q}\le 1$ hold q-a.e.; (d) follows from the fact that $\mathsf{T}:\mathcal{Y}\mapsto[0,1]$ for tests in $\mathcal{T}(\mathcal{Y})$.

If $\varepsilon\in(0,\|w_\sim\|)$, then (61a) and (61b) follow from (60) of Lemma 6 and (66). Then (61c) follows from (51a) with the roles of w and q interchanged and (61b). Furthermore, (61d) follow from (61a) and (61c) because the integrand in (61c) is non-negative for all $\tau<1/F^\wedge_\ell\big(\varepsilon\big|^w_q\big)$ by (34) and (44c); and non-positive for all $\tau>1/F^\vee_\ell\big(\varepsilon\big|^w_q\big)$ by (34) and (44d).

If $\varepsilon\ge\|w_\sim\|$, then (61a) and (61b), follow from (56e) and (66) because $\hbar(0|^q_w)=0$ by (52a). Furthermore, (61d) follows from $w\Big(\frac{\mathrm{d}q_\sim}{\mathrm{d}w}\ge 0\Big)\le\|w_\sim\|$. For $\varepsilon=\|w_\sim\|$ case (61c) follows from (32), (34), (46c) and (46d). For $\varepsilon>\|w_\sim\|$ case, (61c) holds because $F^\wedge_\ell\big(\varepsilon\big|^w_q\big)=F^\vee_\ell\big(\varepsilon\big|^w_q\big)=\infty$.

For $\varepsilon=0$ case (61a) and (61d) follow from applying (51a) and (52b) of Lemma 3, to $\hbar(\gamma|^q_w)$ rather than $\hbar\big(\gamma\big|^w_q\big)$. For $\varepsilon=0$ case (61c) follows from (32), (34), (46a), and (46b).

$\beta(\cdot|^q_w)$ is convex, non-increasing, and lower-semicontinous because it is the supremum of such functions by (61a). Then $\lim_{\varepsilon\downarrow0}\beta(\varepsilon|^q_w)=\beta(0|^q_w)$ by the lower-semicontinuity and monotonicity of $\beta(\cdot|^q_w)$. Furthermore, $\beta(\cdot|^q_w)$ is continuous on $(0,\infty)$ because any convex function is continuous on the interior of the region that it is finite by [87, Theorem 6.3.4]. and $\beta(\varepsilon|^q_w)<\infty$ for all $\varepsilon>0$ by Lemma 6 and (49).

One can prove the remaining assertions of the theorem, via measure theoretic arguments using (61c). We, however, we will rely on the conjugacy relation implied by (61a) to prove those assertions. Let $f(\cdot):\mathbb{R}\mapsto(-\infty,\infty]$ be

$$f(x):=\begin{cases}-\hbar(-x|^q_w) & \text{if } x\le 0,\\ \infty & \text{if } x>0.\end{cases} \tag{67}$$

Then $f^*(\cdot)=\beta(\cdot|^q_w)$, i.e., $\beta(\cdot|^q_w)$ is the convex conjugate of $f(\cdot)$, as a result of (61a) and the definition convex conjugation given in (29). Furthermore, $f(\cdot)$ is a closed convex function (i.e., lower semicontinuous and convex function not identically equal to ∞) by Lemma 3. Then $\beta(\cdot|^q_w)$ is a closed convex function by [77, Theorem E.1.1.2] and convex conjugate of $\beta(\cdot|^q_w)$ is $f(\cdot)$, i.e., $\beta^*(\cdot|^q_w)=f(\cdot)$ by [77, Theorem E.1.3.5]. Furthermore, as a result of [77, Corollary E.1.4.4], [78, Proposition I.6.1.2].

$$x\in\partial\beta(\varepsilon|^q_w)\quad\iff\quad\varepsilon\in\partial f(x).$$

Thus as result of (67) for all $\varepsilon>0$ we have

$$-\lambda\in\partial\beta(\varepsilon|^q_w)\quad\iff\quad\varepsilon\in[\mathrm{D}_+\hbar(\lambda|^q_w),\mathrm{D}_-\hbar(\lambda|^q_w)].$$

Thus as result of Lemma 3 for all $\varepsilon > 0$ we have

$$-\lambda \in \partial\beta(\varepsilon|^q_w) \quad \iff \quad \varepsilon \in \left[w\left(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w} > \lambda\right), w\left(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w} \geq \lambda\right)\right].$$

Then for $\varepsilon > 0$ (61e) and (61f) follow from (34) and Lemma 1 and for $\varepsilon = 0$ (61f) follows from (44b) and because $\mathrm{D}_+\beta(\cdot|^q_w)$ is continuous from the right by [78, Theorem I.4.2.1].

(61h) follows from (44f), (61e), and (61f). Furthermore, (61g) and (61j) follow from (27), (56), (61e), and (61f). Note that (61e), (61f), (61g), and (61h) imply that $\beta(\cdot|^q_w)$ is decreasing and injective on $[0, \|w_\sim\|]$.

(61a) continues to hold when (7) is violated because $\hbar(\varepsilon|^w_q) = \infty$ for all $\varepsilon \geq 0$ and $\beta(\gamma|^q_w) = \infty$ for all $\gamma > 0$ in that case. Then (61d) holds as result of (51a) and (50c). □

The state-of-the-art bound relating $F^\vee_\ell(\cdot|^w_q)$ and $\beta(\cdot|^q_w)$ for arbitrary probability measures w and q, i.e., (15), can be improved slightly and extended to the case when w and q are σ-finite measures using Theorem 1. Since $F^\vee_\ell(\cdot|^w_q)$ is non-decreasing by Lemma 1, (61g) implies

$$\frac{\left|\|w_\sim\| - \varepsilon\right|^+}{F^\vee_\ell(\varepsilon|^w_q)} \geq \beta(\varepsilon|^q_w) \qquad \forall \varepsilon > 0. \tag{68}$$

On the other hand, since $\beta(\cdot|^w_q)$ is convex its supporting line any point and hence at $\varepsilon+\delta$ lies below $\beta(\cdot|^w_q)$ itself. Thus for any γ in $\left[1/F^\vee_\ell(\varepsilon+\delta|^w_q), 1/F^\wedge_\ell(\varepsilon+\delta|^w_q)\right]$, (61e), and (61f) imply

$$\beta(\varepsilon|^q_w) \geq \delta\gamma + \beta(\varepsilon+\delta|^w_q) \qquad \forall \delta \in (0, \|w_\sim\| - \varepsilon). \tag{69}$$

Remark 2. Theorem 1 characterizes $\beta(\cdot|^q_w)$ rather than $\beta(\cdot\|^q_w)$. Nevertheless, we can bound $\beta(\cdot\|^q_w)$ in terms of $\beta(\cdot|^q_w)$ using Lemma 6, deterministic likelihood ratio tests, (44d), and (54):

$$\beta(\varepsilon|^q_w) \leq \beta(\varepsilon\|^q_w) \leq \beta(\varepsilon|^q_w) + \zeta(\varepsilon) \qquad \forall \varepsilon \in (0, \|w_\sim\|), \tag{70a}$$

$$\zeta(\varepsilon) := \inf_{\mathcal{A}\in\mathfrak{B}} \max_{\mathcal{E}\in\mathcal{A}} q(\mathcal{E}) \qquad \forall \varepsilon \in (0, \|w_\sim\|), \tag{70b}$$

where $\mathfrak{B}$ is the set of all $\mathcal{Y}$-measurable finite partitions of the set $\{y \in \mathcal{Y} : \frac{\mathrm{d}w_\sim}{\mathrm{d}q}(y) = F^\vee_\ell(\varepsilon|^w_q)\}$. If $w_\sim$ does not have any atoms then $\zeta(\varepsilon) = 0$ for all $\varepsilon \in (0, \|w_\sim\|)$. If $\mathcal{Y}$ is a countably separated σ-algebra, then all singletons are in $\mathcal{Y}$ by [5, Theorem 6.5.7]. If in addition all atoms of $w_\sim$ are singletons, then $\zeta(\varepsilon)$ has the following less daunting expression

$$\zeta(\varepsilon) = \max\left\{ q(\{y\}) : \tfrac{\mathrm{d}w_\sim}{\mathrm{d}q}(y) = F^\vee_\ell(\varepsilon|^w_q)\right\} \quad \forall \varepsilon \in (0, \|w_\sim\|).$$

The use of max is justified because we are actually taking a supremum over countable set for which supremum is achieved even for the case when there are uncountably many y's in $\mathcal{Y}$ satisfying $\frac{\mathrm{d}w_\sim}{\mathrm{d}q}(y) = F^\vee_\ell(\varepsilon|^w_q)$ by [86, Theorem 4.30]. Note that (70) was already pointed by Strassen, see [6, (2.7)] and the following unnumbered equation in [6]. It is wort mentioning that the upper bound in (70) is not tight even for deterministic likelihood ratio tests: if w and q are geometric probability mass functions with parameter $1/2$ supported on positive and non-negative integers, respectively, then $\frac{\mathrm{d}w}{\mathrm{d}q}(y) = 2$ hold w-a.e. and $\beta(\varepsilon|^q_w) = \beta(\varepsilon\|^q_w) = \left|\frac{1-\varepsilon}{2}\right|^+$ for all $\varepsilon \geq 0$; yet $\zeta(\varepsilon) = 1/2$ for all $\varepsilon \in (0,1)$. Many of these observations, as well as some more advanced ideas going beyond likelihood ratio tests are discussed in [19, Appendix B].

C. Measure Change Identities

Assumption 2. w, q, and v are σ-finite measures satisfying both $\hbar(1|^w_q) \in (0, \infty)$ and $v \sim w_\sim$, and there exist a function f of the form $f : (0,\infty) \mapsto (0,\infty)$ satisfying

$$\tfrac{\mathrm{d}v}{\mathrm{d}q} = f\left(\tfrac{\mathrm{d}w_\sim}{\mathrm{d}q}\right) \qquad w_\sim\text{-a.e.}, \tag{71a}$$

where $w_\sim$ is the q-absolutely continuous component of w. Furthermore, $\varkappa(\varepsilon, v)$ and $\overline{\varkappa}(\varepsilon, v)$ are defined in terms of v and $\mathsf{T}^{q,w}_{\varepsilon,\gamma}$ defined in (59) for any $\gamma \in \left[1/F^\vee_\ell(\varepsilon|^w_q), 1/F^\wedge_\ell(\varepsilon|^w_q)\right]$ as

$$\varkappa(\varepsilon, v) := \int \mathsf{T}^{q,w}_{\varepsilon,\gamma}\,\mathrm{d}v \qquad \forall \varepsilon \in (0, \|w_\sim\|), \tag{71b}$$

$$\overline{\varkappa}(\varepsilon, v) := \int (1 - \mathsf{T}^{q,w}_{\varepsilon,\gamma})\,\mathrm{d}v \qquad \forall \varepsilon \in (0, \|w_\sim\|). \tag{71c}$$

Note that (71b) and (71c) imply

$$\varkappa(\varepsilon, v) + \overline{\varkappa}(\varepsilon, v) = \|v\| \qquad \forall \varepsilon \in (0, \|w_\sim\|). \tag{72}$$

The Csiszar-dual $\widetilde{f} : (0,\infty) \mapsto \mathbb{R}$ of a function f of the form $f : (0,\infty) \mapsto \mathbb{R}$ is defined as, see [35, (2)], [88, (29)].

$$\widetilde{f}(x) := x f\left(\tfrac{1}{x}\right) \qquad \forall x \in (0,\infty). \tag{73}$$

Theorem 2. *Let measures w, q, v and function $f(\cdot)$ satisfy Assumption 2. Then the Csiszar-dual $\widetilde{f}$ defined in* (73) *satisfies*

$$\tfrac{\mathrm{d}v}{\mathrm{d}w} = \widetilde{f}\left(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w}\right) \qquad q_\sim\text{-a.e.} \tag{74a}$$

If $f(\cdot)$ is increasing and bijective and $\hbar(1|^v_q) < \infty$, then

$$\beta(\varepsilon|^q_w) = \beta(\varkappa(\varepsilon, v)|^q_v) \qquad \forall \varepsilon \in (0, \|w_\sim\|). \tag{74b}$$

If $f(\cdot)$ is decreasing and bijective and $\|q_\sim\| < \infty$, then

$$\beta(\varepsilon|^q_w) = \|q_\sim\| - \beta(\overline{\varkappa}(\varepsilon, v)|^q_v) \qquad \forall \varepsilon \in (0, \|w_\sim\|). \tag{74c}$$

If $\widetilde{f}(\cdot)$ is increasing and bijective and $\hbar(1|^v_w) < \infty$, then

$$\varepsilon = \beta(\overline{\varkappa}(\varepsilon, v)|^w_v) \qquad \forall \varepsilon \in (0, \|w_\sim\|). \tag{74d}$$

If $\widetilde{f}(\cdot)$ is decreasing and bijective, and $\|w_\sim\| < \infty$, then

$$\varepsilon = \|w_\sim\| - \beta(\varkappa(\varepsilon, v)|^w_v) \qquad \forall \varepsilon \in (0, \|w_\sim\|). \tag{74e}$$

Proof of Theorem 2. First note that,

$$\tfrac{\mathrm{d}v}{\mathrm{d}w} = \tfrac{\mathrm{d}q_\sim}{\mathrm{d}w}\tfrac{\mathrm{d}v}{\mathrm{d}q_\sim} \qquad v\text{-a.e.},$$

$$\tfrac{1}{\frac{\mathrm{d}q_\sim}{\mathrm{d}w}} = \tfrac{\mathrm{d}w_\sim}{\mathrm{d}q} \qquad v\text{-a.e.}.$$

Then (74a) follows from (71a) and (73). On the other hand, Lemma 6 and Theorem 1 imply

$$\int (1 - \mathsf{T}^{q,w}_{\varepsilon,\gamma})\mathrm{d}q = \beta(\varepsilon|^q_w) \qquad \forall \varepsilon \in (0, \|w_\sim\|), \tag{75}$$

$$\int \mathsf{T}^{q,w}_{\varepsilon,\gamma}\,\mathrm{d}w = \varepsilon \qquad \forall \varepsilon \in (0, \|w_\sim\|). \tag{76}$$

Furthermore, $\gamma \in (0,\infty)$ for all $\varepsilon \in (0, \|w_\sim\|)$ by (44f) and Lemma 6.

When $f(\cdot)$ is increasing and bijective (59), (71a), and (71b) imply $1/f(1/\gamma) \in \left[1/F^\vee_\ell(\varkappa(\varepsilon,v)|^v_q), 1/F^\wedge_\ell(\varkappa(\varepsilon,v)|^v_q)\right]$ and

$$\mathsf{T}^{q,w}_{\varepsilon,\gamma} = \mathsf{T}^{q,v}_{\varkappa(\varepsilon,v),1/f(1/\gamma)} \qquad q\text{-a.e.}.$$

Then (74b) follows from (75), Lemma 6, and Theorem 1 whenever $\hbar\big(1\big|{}^{v}_{q}\big)$ is finite.

When $f(\cdot)$ is decreasing and bijective (59), (71a), and (71c) imply ${}^{1}\!/_{f({}^{1}\!/_{\gamma})} \in \Big[{}^{1}\!/_{F^{\vee}_{\ell}\left(\overline{\varkappa}(\varepsilon,v)\middle|{}^{v}_{q}\right)}, {}^{1}\!/_{F^{\wedge}_{\ell}\left(\overline{\varkappa}(\varepsilon,v)\middle|{}^{v}_{q}\right)}\Big]$ and

$$1-\mathsf{T}^{q,w}_{\varepsilon,\gamma} = \mathsf{T}^{q,v}_{\overline{\varkappa}(\varepsilon,v),{}^{1}\!/_{f({}^{1}\!/_{\gamma})}} \qquad q\text{-a.e..}$$

Then (74c) follows from (75), Lemma 6, and Theorem 1 whenever $\|q_{\sim}\|$ is finite.

When $\widetilde{f}(\cdot)$ is increasing and bijective (59), (71c), and (74a) imply ${}^{1}\!/_{\widetilde{f}(\gamma)} \in \big[{}^{1}\!/_{F^{\vee}_{\ell}(\overline{\varkappa}(\varepsilon,v)|^{v}_{w})}, {}^{1}\!/_{F^{\wedge}_{\ell}(\overline{\varkappa}(\varepsilon,v)|^{v}_{w})}\big]$ and

$$1-\mathsf{T}^{q,w}_{\varepsilon,\gamma} = \mathsf{T}^{w,v}_{\overline{\varkappa}(\varepsilon,v),{}^{1}\!/_{\widetilde{f}(\gamma)}} \qquad w\text{-a.e..}$$

Then (74d) follows from (76), Lemma 6, and Theorem 1 whenever $\hbar(1|^{v}_{w})$ is finite.

When $\widetilde{f}(\cdot)$ is decreasing and bijective (59), (71b), and (74a) imply ${}^{1}\!/_{\widetilde{f}(\gamma)} \in \big[{}^{1}\!/_{F^{\vee}_{\ell}(\varkappa(\varepsilon,v)|^{v}_{w})}, {}^{1}\!/_{F^{\wedge}_{\ell}(\varkappa(\varepsilon,v)|^{v}_{w})}\big]$ and

$$\mathsf{T}^{q,w}_{\varepsilon,\gamma} = \mathsf{T}^{w,v}_{\varkappa(\varepsilon,v),{}^{1}\!/_{\widetilde{f}(\gamma)}} \qquad w\text{-a.e..}$$

Then (74e) follows from (76), Lemma 6, and Theorem 1 whenever $\|w_{\sim}\|$ is finite.

Remark 3. There is a somewhat longer and indirect way to prove (74b), (74c), (74d), and (74e), which is easier to confirm. Here is the proof of (74b) in this alternative way.

$$\begin{aligned}\hbar\Big(\tfrac{1}{f({}^{1}\!/_{\gamma})}\Big|{}^{q}_{v}\Big) &\overset{(a)}{=} q\Big(\tfrac{\mathrm{d}v}{\mathrm{d}q} > f({}^{1}\!/_{\gamma})\Big) + \tfrac{1}{f({}^{1}\!/_{\gamma})}\, v\Big(\tfrac{\mathrm{d}v}{\mathrm{d}q} \le f({}^{1}\!/_{\gamma})\Big)\\ &\overset{(b)}{=} q\Big(\tfrac{\mathrm{d}w_{\sim}}{\mathrm{d}q} > \tfrac{1}{\gamma}\Big) + \tfrac{1}{f({}^{1}\!/_{\gamma})}\, v\Big(\tfrac{\mathrm{d}w_{\sim}}{\mathrm{d}q} \le \tfrac{1}{\gamma}\Big)\\ &\overset{(c)}{=} \beta(\varepsilon|^{q}_{w}) + \tfrac{1}{f({}^{1}\!/_{\gamma})}\varkappa(\varepsilon,v),\end{aligned}$$

where (a) follows from the definition of primitive entropy spectrum in (9), (b) follows from (71a) because $f(\cdot)$ is increasing and invertible, (c) follows from (59), (71b), and (75). Thus $\beta(\varkappa(\varepsilon,v)|^{q}_{v}) \ge \beta(\varepsilon|^{q}_{w})$ by (61a). On the other hand $\beta(\varkappa(\varepsilon,v)|^{q}_{v}) \le \beta(\varepsilon|^{q}_{w})$ by (6b), (71b), and (75) because $\mathsf{T}^{q,w}_{\varepsilon,\gamma} \in \mathcal{T}(\mathcal{Y})$, i.e., $\mathsf{T}^{q,w}_{\varepsilon,\gamma}$ is a randomized test, by Lemma 6. Thus (74b) holds for all $\varepsilon \in (0, \|w_{\sim}\|)$. □

The tilted measures $v^{w,q}_{\alpha}$, defined in the following, satisfy the monotonicity assumptions of Theorem 2 on $f(\cdot)$ and $\widetilde{f}(\cdot)$ at the same time for all $\alpha \in \mathbb{R}\setminus\{0,1\}$. Usually, we will also assume $\|v_{\alpha}\|$ to be finite to ensure finite primitive entropy hypotheses of Theorem 2.

Definition 4. For any pair of σ-finite measures (w,q) satisfying $w \not\perp q$ and real number α, the tilted measure $v^{w,q}_{\alpha}$ is defined as the following $w_{\sim}$-equivalent measure

$$\tfrac{\mathrm{d}v^{w,q}_{\alpha}}{\mathrm{d}q} := \big(\tfrac{\mathrm{d}w_{\sim}}{\mathrm{d}q}\big)^{\alpha} \qquad q\text{-a.e..} \tag{77a}$$

Domain of the tilted measure's finiteness $\mathfrak{I}^{w}_{q}$ is defined as

$$\mathfrak{I}^{w}_{q} := \{\alpha \in \mathbb{R} : \|v^{w,q}_{\alpha}\| < \infty\}. \tag{77b}$$

Note that the tilted measure $v^{w,q}_{\alpha}$ also satisfies

$$\tfrac{\mathrm{d}v^{w,q}_{\alpha}}{\mathrm{d}w} = \big(\tfrac{\mathrm{d}q_{\sim}}{\mathrm{d}w}\big)^{1-\alpha} \qquad w\text{-a.e..} \tag{78}$$

Furthermore, $\ln\|v^{w,q}_{\alpha}\|$ is a convex function as a result of Hölder's inequality [5, Theorem 2.11.1]. Thus $\mathfrak{I}^{w}_{q}$ is an interval and $\mathfrak{I}^{w}_{q}$ is non-empty iff $\ln\|v^{w,q}_{\alpha}\|$ as a function of α satisfies Assumption 1. In the following, we will denote $v^{w,q}_{\alpha}$ simply by v_{α}, whenever w and q are clear from the context.

Remark 4. In the trivial-non-singular case $\frac{\mathrm{d}w_{\sim}}{\mathrm{d}q} = \zeta$ holds $w_{\sim}$-a.e. for some $\zeta \in (0,\infty)$. Thus $v_{\alpha} = \zeta^{\alpha-1}w_{\sim} = \zeta^{\alpha}q_{\sim}$ holds $\forall \alpha \in \mathbb{R}$. Furthermore, trivial-non-singular case is non-void iff $\|q_{\sim}\| \in (0,\infty)$, which is equivalent to $\|w_{\sim}\| \in (0,\infty)$ for the trivial-non-singular case. In the trivial-non-singular-non-void case $\zeta = {}^{\|w_{\sim}\|}\!/_{\|q_{\sim}\|}$, and $\varkappa(\varepsilon, v_{\alpha}) = \varepsilon\zeta^{\alpha-1}$ for all $\alpha \in \mathbb{R}$ and $\varepsilon \in (0, \|w_{\sim}\|)$.

Note that $v_{\alpha} \sim q_{\sim}$ holds by definition and v_{α} satisfies Assumption 2 at least for all $\alpha \in \mathfrak{I}^{w}_{q}$. Furthermore,

(a) If $\alpha < 0$, then $f(\cdot)$ is bijective and decreasing.
(b) If $\alpha > 0$, then $f(\cdot)$ is bijective and increasing.
(c) If $\alpha < 1$, then $\widetilde{f}(\cdot)$ is bijective and increasing.
(d) If $\alpha > 1$, then $\widetilde{f}(\cdot)$ is bijective and decreasing.

Thus if σ-finite measure w and q satisfy $\hbar\big(1\big|{}^{w}_{q}\big) \in (0,\infty)$ and $\alpha \in \mathfrak{I}^{w}_{q}$, then Theorem 2 implies for all $\varepsilon \in (0, \|w_{\sim}\|)$, that

$$\beta(\varepsilon|^{q}_{w}) = \begin{cases} \|q_{\sim}\| - \beta\big(\overline{\varkappa}(\varepsilon,v_{\alpha})\big|^{q}_{v_{\alpha}}\big) & \text{if } \alpha < 0,\\ \overline{\varkappa}(\varepsilon,v_{\alpha}) & \text{if } \alpha = 0,\\ \beta\big(\varkappa(\varepsilon,v_{\alpha})\big|^{q}_{v_{\alpha}}\big) & \text{if } \alpha > 0,\end{cases} \tag{79a}$$

$$\varepsilon = \begin{cases} \beta\big(\overline{\varkappa}(\varepsilon,v_{\alpha})\big|^{w}_{v_{\alpha}}\big) & \text{if } \alpha < 1,\\ \varkappa(\varepsilon,v_{\alpha}) & \text{if } \alpha = 1,\\ \|w_{\sim}\| - \beta\big(\varkappa(\varepsilon,v_{\alpha})\big|^{w}_{v_{\alpha}}\big) & \text{if } \alpha > 1.\end{cases} \tag{79b}$$

Furthermore, invoking (65), we can express the function $\varkappa(\cdot, v_{\alpha}) : (0, \|w_{\sim}\|) \mapsto (0, \|v_{\alpha}\|)$ in two distinct ways:

$$\varkappa(\varepsilon,v_{\alpha}) = \begin{cases} \|v_{\alpha}\| - \beta\big(\|q_{\sim}\| - \beta(\varepsilon|^{q}_{w})\big|^{v_{\alpha}}_{q}\big) & \text{if } \alpha < 0,\\ \|v_{\alpha}\| - \beta(\varepsilon|^{q}_{w}) & \text{if } \alpha = 0,\\ \beta\big(\beta(\varepsilon|^{q}_{w})\big|^{v_{\alpha}}_{q}\big) & \text{if } \alpha > 0,\end{cases} \tag{79c}$$

$$\varkappa(\varepsilon,v_{\alpha}) = \begin{cases} \|v_{\alpha}\| - \beta(\varepsilon|^{v_{\alpha}}_{w}) & \text{if } \alpha < 1,\\ \varepsilon & \text{if } \alpha = 1,\\ \beta(\|w_{\sim}\| - \varepsilon|^{v_{\alpha}}_{w}) & \text{if } \alpha > 1.\end{cases} \tag{79d}$$

Theorem 2 and Definition 4 are inspired by the information theorists approach to large deviations analysis that can be traced back to [13]–[17]. Customarily the tilted measure is defined as a probability measure of the form ${}^{v_{\alpha}}\!/_{\|v_{\alpha}\|}$, e.g., [13, Theorem 5 and (3.27)] considering orders in $(0,1)$. The primary advantage of avoiding the scaling is that not only memorylessness but also Markovian structure (and probably other structural assumption on w and q) are inherited by the tilted measure when there is no scaling, see §V-A and §V-B.

IV. Approximations for randomized-NP-function

In this section, we approximate the rNP-function $\beta(\cdot|^{q}_{w})$ under hypotheses on the likelihood-ratio distribution that are implied by the Berry–Esseen type bounds non-asymptotically and Central Limit Theorem asymptotically. Since likelihood-ratios typically follow a multiplicative rather than additive structure (e.g., memoryless and Markovian models) these assumptions are more naturally formulated and interpreted using the distribution of the log-likelihood ratio rather than the likelihood ratio.

In §IV-A, we first define the log-likelihood ratio and its distribution, and then express $\beta(\cdot|_w^q)$ in terms of the log-likelihood ratio distribution $F_{\mathsf{L}}(\cdot|_q^w)$ using Theorem 1. In §IV-B, we define Gaussian Mills ratio and review certain well-known bounds on it that will be helpful both in our analysis and in interpreting our results. Finally, in §IV-C we first approximate $\beta(\cdot|_w^q)$ assuming $F_{\mathsf{L}}(\cdot|_q^w)$ is close to a Gaussian distribution in Kolmogorov distance and then bound $\beta(\cdot|_q^w)$ assuming $F_{\mathsf{L}}(\cdot|_w^{\gamma v_\alpha})$ is close to a Gaussian distribution in Kolmogorov distance.

A. The Log-Likelihood-Ratio

Definition 5. For any pair of σ-finite measures (w,q) on the same measurable space, say $(\mathcal{Y},\mathcal{Y})$, the log-likelihood-ratio $\mathsf{L}_{w,q}(\cdot)$ and the log-likelihood-ratio distribution $F_{\mathsf{L}}(\cdot|_q^w)$ are defined as

$$\mathsf{L}_{w,q}:=\begin{cases}\infty & \text{if } \frac{\mathrm{d}q}{\mathrm{d}(w+q)}=0,\\ \ln\frac{\mathrm{d}w_\sim}{\mathrm{d}q} & \text{if } \frac{\mathrm{d}q}{\mathrm{d}(w+q)}>0,\end{cases} \tag{80a}$$

$$F_{\mathsf{L}}(\tau|_q^w):=\int \mathbb{1}_{\{\mathsf{L}_{w,q}\le\tau\}}\mathrm{d}w \qquad \forall\tau\in\mathbb{R}, \tag{80b}$$

where $w_\sim$ isthe q-absolutely continuous component of w.

Note that definitions in (32) and (80) imply

$$\mathsf{L}_{w,q}=\ln\ell_{w,q} \qquad (w+q)\text{-a.e.}, \tag{81}$$

$$F_{\mathsf{L}}(\tau|_q^w)=F_\ell(e^\tau|_q^w) \qquad \forall\tau\in\mathbb{R}. \tag{82}$$

Thus identities (34) and (35) imply

$$\mathsf{L}_{w,q}=-\mathsf{L}_{q,w} \qquad (w+q)\text{-a.e.} \tag{83}$$

$$F_{\mathsf{L}}(\tau|_q^w)=F_{\mathsf{L}}(\tau|_{q_\sim}^{w_\sim}) \qquad \forall\tau\in\mathbb{R}. \tag{84}$$

Remark 5. $F_{\mathsf{L}}^\wedge(\cdot|_q^w):\mathbb{R}\mapsto\overline{\mathbb{R}}$ and $F_{\mathsf{L}}^\vee(\cdot|_q^w):\mathbb{R}\mapsto\overline{\mathbb{R}}$, i.e., the log-likelihood ratio quantiles, can be defined using the definition of the likelihood ratio quantiles by replacing $F_\ell(\cdot|_q^w)$ with $F_{\mathsf{L}}(\cdot|_q^w)$ in (37). Then $F_{\mathsf{L}}^\wedge(\cdot|_q^w)=\ln F_\ell^\wedge(\cdot|_q^w)$ and $F_{\mathsf{L}}^\vee(\cdot|_q^w)=\ln F_\ell^\vee(\cdot|_q^w)$, with the understanding that $\ln(-\infty)$, $\ln 0$, and $\ln\infty$ stand for $-\infty$, $-\infty$, and ∞, respectively. Then (61g) implies

$$\beta(\varepsilon|_w^q)=\int_\varepsilon^{\|w_\sim\|} e^{-f(\tau)}\mathrm{d}\tau \qquad \forall\varepsilon\in[0,\|w_\sim\|], \tag{85}$$

for any $f:(0,\|w_\sim\|)\mapsto\mathbb{R}$ s.t. $F_{\mathsf{L}}^\wedge(\cdot|_q^w)\le f(\cdot)\le F_{\mathsf{L}}^\vee(\cdot|_q^w)$.

For any pair of σ-finite measures w and q —not necessarily satisfying (7)— first invoking (62b), i.e., (8), and then applying the change of variable $\tau=e^{-t}$, we get

$$\beta(\varepsilon|_w^q)=\int_{-\infty}^{\infty} e^{-t}\big|F_{\mathsf{L}}(t|_q^w)-\varepsilon\big|^+\mathrm{d}t \qquad \forall\varepsilon\ge 0. \tag{86}$$

For example, if w is a probability measure with a finite $\mathbf{E}_w[\mathsf{L}_{w,q}^2]$, then the Chebychev-Cantelli inequality and (86) lead to the following non-asymptotic Stein's lemma, which improves similar results reported before in [19, Lemma 15], [20, Lemma 59], [89, (12)]. Proof of Lemma 7, which is elementary, is presented in Appendix C for completeness.

Lemma 7. *Let w be a probability measure and q be a σ-finite measure satisfying $\mathbf{E}_w[(\mathsf{L}_{w,q})^2]<\infty$, then $\mathbf{E}_w[\mathsf{L}_{w,q}]\in\mathbb{R}$ and*

$$\ln\tfrac{\Delta}{1+\Delta}-\sigma\sqrt{\tfrac{\varepsilon+\Delta}{1-\varepsilon}}\le\ln\tfrac{\beta(\varepsilon|_w^q)}{1-\varepsilon}+\mathbf{E}_w[\mathsf{L}_{w,q}]\le\sigma\sqrt{\tfrac{1-\varepsilon}{\varepsilon}} \tag{87a}$$

for all $\varepsilon\in(0,1)$, $\Delta>0$, where $\sigma:=\sqrt{\mathbf{E}_w\Big[\big(\mathsf{L}_{w,q}-\mathbf{E}_w[\mathsf{L}_{w,q}]\big)^2\Big]}$.

Remark 6. $\ln\frac{1-\varepsilon}{\beta(\varepsilon|_w^q)}$ is called ε-hypothesis testing divergence and denoted by $D_h^\varepsilon(w\|q)$, see [24, p. 7044], [25, (2.6)]. Note that for any probability measure w, monotonicity of divided differences for convex functions and convexity of $\beta(\cdot|_w^q)$ imply that $D_h^\varepsilon(w\|q)$ is decreasing in ε and $D_h^\varepsilon(w\|q)\ge-\ln\|q_\sim\|$. Note that inequality holds as an equality for all $\varepsilon\in(0,1)$ in the trivial case provided that $w\prec q$.

B. Gaussian Mills Ratio

The Gaussian Mills ratio is defined as the ratio of Gassian Q-function and Gaussian probability density:

$$\Psi(\tau):=\tfrac{Q(\tau)}{\varphi(\tau)} \qquad \forall\tau\in\mathbb{R}. \tag{88}$$

Then $\Psi(\cdot):\mathbb{R}\mapsto(0,\infty)$ is a decreasing strictly convex and invertible function because $\ln\Psi(\cdot):\mathbb{R}\mapsto\mathbb{R}$ is such a function by [90]. Strikingly its multiplicative inverse $1/\Psi(\cdot)$ is a convex functions as well, see [91].

Note that $\ln\Psi(\tau+\delta)>\ln\Psi(\tau)+\frac{\Psi'(\tau)}{\Psi(\tau)}\delta$ for all $\tau\in\mathbb{R}$ and $\delta\ne0$ where $\Psi'(\tau)=\frac{\mathrm{d}}{\mathrm{d}\tau}\Psi(\tau)$ because $\ln\Psi(\cdot)$ is strictly convex. Since $\ln\Psi(\cdot)$ is strictly decreasing as well, we have

$$\Psi(\tau)>\Psi(\tau+\delta)>\Psi(\tau)e^{\frac{\Psi'(\tau)}{\Psi(\tau)}\delta} \qquad \forall\tau\in\mathbb{R},\ \delta>0, \tag{89}$$

On the other hand $\Psi'(\tau)/\Psi(\tau)$ is non-decreasing because $\ln\Psi(\cdot)$ is convex. Then using the identity $\Psi'(\tau)=\tau\Psi(\tau)-1$ we get

$$\Psi(\tau)>\Psi(\tau+\delta)>\Psi(\tau)e^{(x-\frac{1}{\Psi(x)})\delta} \quad \forall\tau\ge x,\delta>0,x\in\mathbb{R}. \tag{90}$$

On that other hand $1/\Psi(-1.3)\le0.19$. Hence

$$1.5\delta\ge\ln\tfrac{\Psi(\tau)}{\Psi(\tau+\delta)}\ge0 \qquad \forall\tau\ge-1.3,\ \delta\ge0. \tag{91}$$

Furthermore,

$$\tfrac{2}{\tau+\sqrt{\tau^2+4}}\le\Psi(\tau)\le\tfrac{2}{\tau+\sqrt{\tau^2+8/\pi}} \qquad \forall\tau\ge0, \tag{92}$$

by [92, (10), (11), (12)] and [93, 7.1.13]. Let the function $g(\cdot,\cdot):\mathbb{R}\times(0,1)\mapsto(0,\infty)$ be

$$g(x,\varepsilon):=e^{\frac{x^2}{2}}Q\big(x-Q^{-1}(\varepsilon)\big), \tag{93}$$

$$=\tfrac{Q\left(x-Q^{-1}(\varepsilon)\right)}{\varphi(x)\sqrt{2\pi}}, \tag{94}$$

$$=\Psi\big(x-Q^{-1}(\varepsilon)\big)\varphi\big(Q^{-1}(\varepsilon)\big)e^{xQ^{-1}(\varepsilon)} \tag{95}$$

Then

$$\tfrac{\partial}{\partial\varepsilon}g(x,\varepsilon)=-e^{xQ^{-1}(\varepsilon)}$$

$$=\tfrac{-g(x,\varepsilon)}{\Psi(x-Q^{-1}(\varepsilon))}\tfrac{1}{\varphi(Q^{-1}(\varepsilon))}. \tag{96}$$

We denote the Gaussian cumulative distribution function by $\Phi(\cdot)$, i.e.,

$$\Phi(x):=\int_{-\infty}^{x}\varphi(\tau)\mathrm{d}\tau \qquad \forall x\in\mathbb{R}. \tag{97}$$

Note that (3) and the fact that $\varphi(\tau)=\varphi(-\tau)$ for $\tau\in\mathbb{R}$ imply

$$\Phi(x)=Q(-x) \qquad \forall x\in\mathbb{R}. \tag{98}$$

C. Kolmogorov Distance to Normality

Theorem 3. *If w and q are σ-finite measures satisfying*

$$\sup_{x\in\mathbb{R}} \left|F_{\mathsf{L}}\big(\xi+\sigma x\big|_q^w\big) - \Phi(x)\right| \leq \delta, \tag{99a}$$

for a $\xi\in\mathbb{R}$, a $\sigma>0$, and a $\delta\in(0,1)$, then

$$\beta(\varepsilon|_w^q) \leq e^{-\xi+\frac{\sigma^2}{2}}\, Q\big(\sigma - Q^{-1}(\varepsilon-\delta)\big) \quad \forall \varepsilon\in(\delta, 1+\delta), \tag{99b}$$

$$\beta(\varepsilon|_w^q) \geq e^{-\xi+\frac{\sigma^2}{2}}\, Q\big(\sigma - Q^{-1}(\varepsilon+\delta)\big) \quad \forall \varepsilon\in(0, 1-\delta), \tag{99c}$$

Furthermore, if $\varepsilon\in(\delta, 1-\delta-Q(\sigma))$, then

$$\beta(\varepsilon|_w^q) = e^{-\xi+\sigma\tau}\Psi(\sigma-\tau)\varphi(\tau)\Delta, \tag{99d}$$

where $\tau = Q^{-1}(\varepsilon)$, and Δ satisfies $|\ln\Delta| \leq \frac{\delta(\sigma+0.5)\sqrt{2\pi}}{(1-\varepsilon-\delta)\wedge(\varepsilon-\delta)}$.

Proof of Theorem 3. Note that $\|w_\sim\| \in [1-\delta, 1+\delta]$ by (99a) because $\|w_\sim\| = \lim_{x\uparrow\infty} F_{\mathsf{L}}\big(x\big|_q^w\big)$. Thus w and q satisfy (7).

$$\begin{aligned}
\beta(\varepsilon|_w^q) &\overset{(a)}{=} \sigma\int_{-\infty}^{\infty} e^{-(\xi+\sigma x)}\left|F_{\mathsf{L}}\big(\xi+\sigma x\big|_q^w\big)-\varepsilon\right|^+ \mathrm{d}x,\\
&\overset{(b)}{\leq} e^{-\xi}\sigma\int_{-\infty}^{\infty} e^{-\sigma x}\left|\Phi(x)+\delta-\varepsilon\right|^+ \mathrm{d}x,\\
&\overset{(c)}{=} e^{-\xi}\sigma\int_{\Phi^{-1}(\varepsilon-\delta)}^{\infty} e^{-\sigma x}\left(\Phi(x)+\delta-\varepsilon\right)\mathrm{d}x,
\end{aligned}$$

where (a) follows from (86) for $t=\xi+\sigma x$, (b) follows from the hypothesis in (99a), (c) holds because $\Phi(\cdot):\mathbb{R}\mapsto(0,1)$ is an invertible and increasing function. Then (99b) follows from (100) established below for $\lambda = \Phi^{-1}(\varepsilon-\delta)$.

$$\begin{aligned}
\int_\lambda^\infty e^{-\sigma x}(\Phi(x)-\Phi(\lambda))\mathrm{d}x &= \int_\lambda^\infty e^{-\sigma x}\int_\lambda^x \varphi(\tau)\mathrm{d}\tau\mathrm{d}x,\\
&= \int_\lambda^\infty \varphi(\tau)\int_\tau^\infty e^{-\sigma x}\mathrm{d}x\mathrm{d}\tau,\\
&= \tfrac{1}{\sigma}\int_\lambda^\infty \varphi(\tau)e^{-\sigma\tau}\mathrm{d}\tau\\
&= \tfrac{1}{\sigma}\int_\lambda^\infty e^{\frac{\sigma^2}{2}}\varphi(\tau+\sigma)\mathrm{d}\tau,\\
&= \tfrac{1}{\sigma}e^{\frac{\sigma^2}{2}}\,Q(\sigma+\lambda).
\end{aligned} \tag{100}$$

(99c) is established using the lower bound on $F_{\mathsf{L}}\big(\xi-\sigma x\big|_q^w\big)$ in (99a) by following a similar analysis. It is worth mentioning that that both (99b) and (99c) can be derived using (85) instead of (86).

To establish (99d), first note that $\phi\in(\varepsilon-\delta,\varepsilon+\delta)$ such that $\beta(\varepsilon|_w^q) = e^{-\xi}g(\sigma,\phi)$ as a result of intermediate value theorem and the definition of $g(\cdot,\cdot)$ in (93), by (99b) and (99c). Since $\ln g(\cdot,\cdot)$ is continuously differentiable in its second argument, as a result of Taylor's theorem there exists a ϕ_0 between ϕ and ε such that

$$\ln g(\sigma,\phi) = \ln g(\sigma,\varepsilon) + (\phi-\varepsilon)\tfrac{g_\varepsilon'(\sigma,\phi_0)}{g(\sigma,\phi_0)}. \tag{101}$$

On the other hand using (96) we get

$$\begin{aligned}
\left|\tfrac{g_\varepsilon'(\sigma,\phi_0)}{g(\sigma,\phi_0)}\right| &= \tfrac{1}{\Psi(\sigma-Q^{-1}(\phi_0))}\tfrac{\Psi(\left|Q^{-1}(\phi_0)\right|)}{(1-\phi_0)\wedge\phi_0}\\
&\overset{(a)}{\leq} \tfrac{1}{\Psi(\sigma-Q^{-1}(\phi_0))}\tfrac{\sqrt{\pi/2}}{(1-\phi_0)\wedge\phi_0},\\
&\overset{(b)}{\leq} \tfrac{1}{\Psi(\sigma-Q^{-1}(\varepsilon+\delta))}\tfrac{\sqrt{\pi/2}}{(1-\phi_0)\wedge\phi_0},\\
&\overset{(c)}{\leq} \tfrac{1}{\Psi(2\sigma)}\tfrac{\sqrt{\pi/2}}{(1-\phi_0)\wedge\phi_0},\\
&\overset{(d)}{\leq} (\sigma+\sqrt{\sigma^2+1})\tfrac{\sqrt{\pi/2}}{(1-\phi_0)\wedge\phi_0},\\
&\overset{(e)}{\leq} (\sigma+0.5)\tfrac{\sqrt{2\pi}}{(1-\phi_0)\wedge\phi_0},
\end{aligned} \tag{102}$$

where (a) holds because $\Psi(\tau)\leq\sqrt{\pi/2}$ for all $\tau>0$ by (92), (b) holds because $\Psi(\cdot)$ and $Q^{-1}(\cdot)$ are both decreasing functions and $\phi_0\in(\varepsilon-\delta,\varepsilon+\delta)$, (c) holds because $\Psi(\cdot)$ and $Q^{-1}(\cdot)$ are both decreasing functions, $\varepsilon+\delta\leq 1-Q(\sigma)$ by hypothesis, and $Q^{-1}(1-x) = -Q^{-1}(x)$ for all $x\in(0,1)$, (d) follows from (92), (e) follows from $\sqrt{\sigma^2+1}\leq\sigma+1$. Thus (99d) follows from (101), (102), and $\phi_0\in(\varepsilon-\delta,\varepsilon+\delta)$. □

Theorem 3 assumes the cumulative distribution function of the log-likelihood ratio $\mathsf{L}_{w,q}$ in the measure space $(\mathcal{Y},\mathcal{Y},w)$ is close to that of a normal (Gaussian) distribution to bound $\beta(\cdot|_w^q)$. One can bound $\beta(\cdot|_w^q)$ or $\beta\big(\cdot\big|_q^w\big)$ using Theorem 2 by first bounding $\beta(\cdot|_v^q)$ and $\beta(\cdot|_v^w)$ using Theorem 3, provided that the cumulative distribution function of both $\ln f\big(e^{\mathsf{L}_{w,q}}\big)$ and $\ln\widetilde{f}\big(e^{-\mathsf{L}_{w,q}}\big)$ in the measure space $(\mathcal{Y},\mathcal{Y},v)$ are close to normal distributions. If $v=\gamma v_\alpha$ for a positive constant γ and tilted measure v_α for an $\alpha\in\mathbb{R}\setminus\{0,1\}$ then these conditions are one and the same; that is the main idea behind Theorem 4 given in the following.

Theorem 4. *Let w and q be σ-finite measures satisfying $\hbar\big(1\big|_q^w\big)\in(0,\infty)$ and assume that an $\alpha\in\mathcal{I}_q^w\cap(0,\infty)$ satisfies*

$$\sup_{x\in\mathbb{R}} \left|F_{\mathsf{L}}(\xi+\sigma x|_w^{\gamma v_\alpha}) - \Phi(x)\right| \leq \delta, \tag{103a}$$

for a $\gamma>0$, a $\xi\in\mathbb{R}$, a $\sigma>0$, and a $\delta\in(0,1/8)$. Let $\eta(\cdot)$ be

$$\eta(\epsilon) := e^{\frac{\alpha\xi-\ln\gamma}{1-\alpha}+\frac{\alpha^2\sigma^2}{2(1-\alpha)^2}}\,Q\Big(\tfrac{\alpha\sigma}{|1-\alpha|}-Q^{-1}(\epsilon)\Big) \tag{103b}$$

for all $\epsilon\in(0,1)$ for ξ and σ of (103a)*.*

If $\alpha\in(0,1)$, then for all $\epsilon\in(3\delta,1-3\delta)$ there exists a $\phi\in(\epsilon-3\delta,\epsilon+3\delta)$ such that

$$\beta\big(\eta(\epsilon)\big|_q^w\big) = e^{-\xi+\frac{\sigma^2}{2}}\,Q\big(\sigma-Q^{-1}(1-\phi)\big). \tag{103c}$$

If $\alpha\in(1,\infty)$, then for all $\epsilon\in(2\delta,1-2\delta)$ there exists a $\phi\in(\epsilon-2\delta,\epsilon+2\delta)$ such that

$$\beta\big(\eta(\epsilon)\big|_q^w\big) = \|w_\sim\| - e^{-\xi+\frac{\sigma^2}{2}}\,Q\big(\sigma-Q^{-1}(\phi)\big). \tag{103d}$$

Furthermore, if $\delta\leq 1/10$ and $\widetilde{\eta}\in[\eta(3/5),\eta(2/5)]$, then $\exists\Delta$ satisfying both $|\ln\Delta|\leq\frac{0.43}{1\wedge\alpha}+(\sigma+2.3)15\sqrt{2\pi}\delta$ and

$$\beta\big(\widetilde{\eta}\big|_q^w\big) = \begin{cases}\theta(\alpha,\sigma)\widetilde{\eta}^{\frac{\alpha-1}{\alpha}}\gamma^{-\frac{1}{\alpha}}\Delta & \text{if } \alpha\in(0,1),\\ \|w_\sim\| - \theta(\alpha,\sigma)\widetilde{\eta}^{\frac{\alpha-1}{\alpha}}\gamma^{-\frac{1}{\alpha}}\Delta & \text{if } \alpha>1,\end{cases} \tag{103e}$$

where $\theta(\alpha,\sigma) = (1/2\pi)^{\frac{1}{2\alpha}}\left(\Psi(\alpha\sigma/|\alpha-1|)\right)^{\frac{1-\alpha}{\alpha}}\Psi(\sigma)$.

Remark 7. Note that for all $\alpha\in(0,\infty)$ as σ increases $\theta(\alpha,\sigma)$ is bounded above and below tightly by bounds on the Gaussian Mills ratio given in (92): $\theta\sim\frac{\alpha}{|1-\alpha|}\big(\sigma\sqrt{2\pi}\big)^{-1/\alpha}$ as $\sigma\to\infty$.

Proof of Theorem 4. Let us first establish (103c) and (103d). To that end we will first invoke Theorem 3 to bound $\beta\big(\cdot\big|_{\gamma v_\alpha}^{w}\big)$ and $\beta\big(\cdot\big|_{\gamma v_\alpha}^{q}\big)$, and then employ Theorem 2.

Note that (103a) ensures that the hypothesis of Theorem 3 is satisfied for $\beta\big(\cdot\big|_{\gamma v_\alpha}^{w}\big)$. Thus Theorem 3 and the intermediate value theorem imply that for all $\epsilon\in(\delta,1-\delta)$ there exists a $\phi\in(\epsilon-\delta,\epsilon+\delta)$ satisfying

$$\beta\big(\epsilon\big|_{\gamma v_\alpha}^{w}\big)=e^{-\xi+\frac{\sigma^2}{2}}Q\big(\sigma-Q^{-1}(\phi)\big)\quad=e^{-\xi}g(\sigma,\phi).\quad(104)$$

In order to apply Theorem 3 to bound $\beta\big(\cdot\big|_{\gamma v_\alpha}^{q}\big)$, we need to first ensure that the hypothesis of Theorem 3 is satisfied for $\beta\big(\cdot\big|_{\gamma v_\alpha}^{q}\big)$. To that end first note that (77a) and (78) imply

$$\begin{aligned}\mathsf{L}_{\gamma v_\alpha,w}&=\ln\gamma+(\alpha-1)\mathsf{L}_{w,q}&&v_\alpha\text{-a.e.},\\ \mathsf{L}_{\gamma v_\alpha,q}&=\ln\gamma+\alpha\mathsf{L}_{w,q}&&v_\alpha\text{-a.e.}.\end{aligned}$$

Thus

$$\mathsf{L}_{\gamma v_\alpha,q}=\tfrac{\alpha}{\alpha-1}\mathsf{L}_{\gamma v_\alpha,w}+\tfrac{1}{1-\alpha}\ln\gamma\qquad v_\alpha\text{-a.e.}.$$

Then (103a) implies[15]

$$\sup_{x\in\mathbb{R}}\Big|F_{\mathsf{L}}\Big(\tfrac{\ln\gamma-\alpha\xi}{1-\alpha}+\tfrac{\alpha\sigma}{|\alpha-1|}x\Big|_{w}^{\gamma v_\alpha}\Big)-\Phi(x)\Big|\le\delta.$$

Thus Theorem 3 and the intermediate value theorem imply that for all $\varkappa\in(\delta,1-\delta)$ there exists a $\epsilon\in(\varkappa-\delta,\varkappa+\delta)$ satisfying

$$\beta\big(\varkappa\big|_{\gamma v_\alpha}^{q}\big)=\eta(\epsilon).\qquad(105)$$

Note that $\eta(\cdot)$ defined in (103b) is invertible, decreasing, and continuous on $(0,1)$. On the other hand $\gamma\|v_\alpha\|\in(1-\delta,1+\delta)$ by (103a) and $\beta\big(\cdot\big|_{\gamma v_\alpha}^{q}\big)$ is invertible, decreasing, and continuous on $(0,\gamma\|v_\alpha\|)$. Thus (105) defines an invertible increasing and continuous function, i.e., $\zeta(\cdot):=\eta^{-1}\big(\beta\big(\cdot\big|_{\gamma v_\alpha}^{q}\big)\big)$, on $(\delta,1-\delta)$ satisfying $|\zeta(\varkappa)-\varkappa|\le\delta$ for all $\varkappa$ in $(\delta,1-\delta)$. Then $\zeta^{-1}(\cdot)$ is increasing on its domain and $(2\delta,1-2\delta)$ is a subset of the domain of $\zeta^{-1}(\cdot)$. Hence

$$\begin{aligned}\beta\big(\zeta^{-1}(\epsilon)\big|_{\gamma v_\alpha}^{q}\big)&=\eta(\epsilon)&&\forall\epsilon\in(2\delta,1-2\delta),&(106)\\ \big|\epsilon-\zeta^{-1}(\epsilon)\big|&\le\delta&&\forall\epsilon\in(2\delta,1-2\delta).&(107)\end{aligned}$$

Furthermore, (65) and (106) imply

$$\zeta^{-1}(\epsilon)=\beta\big(\eta(\epsilon)\big|_{q}^{\gamma v_\alpha}\big)\qquad\forall\epsilon\in(2\delta,1-2\delta).\qquad(108)$$

On the other hand, $\eta(\epsilon)\in(0,\|q_\sim\|)$ for all $\epsilon\in(2\delta,1-2\delta)$, thus $\beta\big(\eta(\epsilon)\big|_{q}^{w}\big)\in(0,\|w_\sim\|)$ for all ϵ in $(2\delta,1-2\delta)$. Applying (74b) of Theorem 2 for $f(x)=\gamma x^\alpha$ for $\varepsilon=\beta\big(\eta(\epsilon)\big|_{q}^{w}\big)$ and invoking (65), we get

$$\eta(\epsilon)=\beta\big(\varkappa\big(\beta\big(\eta(\epsilon)\big|_{q}^{w}\big),\gamma v_\alpha\big)\big|_{\gamma v_\alpha}^{q}\big)\qquad\forall\epsilon\in(2\delta,1-2\delta).$$

[15]Note that Kolmogorov distance between a continuous function f and a monotonic function does not change if the monotonic function is replaces by its upper-semicontinuous or lower-semicontinuous variants. Thus for any continuous function $f(\cdot)$ we have

$$\sup_{x\in\mathbb{R}}\big|F_{\mathsf{L}}\big(x\big|_{q}^{v}\big)-f(x)\big|=\sup_{x\in\mathbb{R}}\big|\lim\nolimits_{\tau\uparrow x}F_{\mathsf{L}}\big(\tau\big|_{q}^{v}\big)-f(x)\big|$$

Invoking (65) again, this time for $\beta\big(\cdot\big|_{\gamma v_\alpha}^{q}\big)$, we get

$$\beta\big(\eta(\epsilon)\big|_{q}^{\gamma v_\alpha}\big)=\varkappa\big(\beta\big(\eta(\epsilon)\big|_{q}^{w}\big),\gamma v_\alpha\big)\qquad\forall\epsilon\in(2\delta,1-2\delta).$$

Thus (107) and (108) imply

$$\big|\epsilon-\varkappa\big(\beta\big(\eta(\epsilon)\big|_{q}^{w}\big),\gamma v_\alpha\big)\big|\le\delta\qquad\forall\epsilon\in(2\delta,1-2\delta).\qquad(109)$$

Since $\gamma\|v_\alpha\|\in(1-\delta,1+\delta)$ by (103a), using triangle inequality together with (72), we get

$$\big|1-\epsilon-\overline{\varkappa}\big(\beta\big(\eta(\epsilon)\big|_{q}^{w}\big),\gamma v_\alpha\big)\big|\le2\delta\quad\forall\epsilon\in(2\delta,1-2\delta).\quad(110)$$

(65), (74d) of Theorem 2 for $f(x)=\gamma x^\alpha$, (104), and (110) imply (103c); and (65), (74e) of Theorem 2 for $f(x)=\gamma x^\alpha$, (104), and (109) imply (103d).

We will establish (103e) by invoking (103c) and (103d) together with Taylor's theorem. Note that (103b), (103c), and (103d) can be expressed in terms of $g(\cdot,\cdot)$ defined in (93) as

$$\begin{aligned}\eta(\epsilon)&=e^{\frac{\alpha\xi-\ln\gamma}{1-\alpha}}g\Big(\tfrac{\alpha\sigma}{|1-\alpha|},\epsilon\Big),&(111a)\\ \beta\big(\eta(\epsilon)\big|_{q}^{w}\big)&=e^{-\xi}g(\sigma,1-\phi),&(111b)\\ \beta\big(\eta(\epsilon)\big|_{q}^{w}\big)&=\|w_\sim\|-e^{-\xi}g(\sigma,\phi).&(111c)\end{aligned}$$

Let ϵ be the unique solution of $\widetilde{\eta}=\eta(\epsilon)$ in $[2/5,3/5]$ and ϕ be the corresponding number in $(\epsilon-3\delta,\epsilon+3\delta)$ satisfying (103c) if $\alpha\in(0,1)$ and (103d) if $\alpha\in(1,\infty)$. Since $\ln g(\cdot,\cdot)$ is continuously differentiable in its second argument, there exists ϵ_0 between ϵ and ϕ satisfying

$$\ln g(\sigma,1-\phi)=\ln g(\sigma,1-\epsilon)+\tfrac{g'_\varepsilon(\sigma,1-\epsilon_0)}{g(\sigma,1-\epsilon_0)}(\epsilon-\phi)\qquad(112a)$$

if $\alpha\in(0,1)$ and satisfying

$$\ln g(\sigma,\phi)=\ln g(\sigma,\epsilon)+\tfrac{g'_\varepsilon(\sigma,\epsilon_0)}{g(\sigma,\epsilon_0)}(\phi-\epsilon),\qquad(112b)$$

for $\alpha\in(1,\infty)$. Note that $\epsilon_0\in(1/10,9/10)$ in either case. Let us proceed with bounding $g'_\varepsilon(\sigma,\epsilon_0)$ for an $\epsilon_0\in(1/10,9/10)$:

$$\begin{aligned}\Big|\tfrac{g'_\varepsilon(\sigma,\epsilon_0)}{g(\sigma,\epsilon_0)}\Big|&=\tfrac{1}{\Psi(\sigma-Q^{-1}(\epsilon_0))}\tfrac{\Psi\left(\left|Q^{-1}(\epsilon_0)\right|\right)}{(1-\epsilon_0)\wedge\epsilon_0}\\ &\overset{(a)}{\le}\tfrac{1}{\Psi(\sigma+Q^{-1}(\epsilon_0))}\tfrac{\sqrt{\pi/2}}{(1-\epsilon_0)\wedge\epsilon_0},\\ &\overset{(b)}{\le}\tfrac{5\sqrt{2\pi}}{\Psi(\sigma+Q^{-1}(1/10))},\\ &\overset{(c)}{\le}\tfrac{5\sqrt{2\pi}}{\Psi(\sigma+1.3)},\\ &\overset{(d)}{\le}(\sigma+1.3+\sqrt{(\sigma+1.3)^2+4})2.5\sqrt{2\pi},\\ &\overset{(e)}{\le}(\sigma+2.3)5\sqrt{2\pi},\end{aligned}\qquad(113)$$

where (a) holds because $\Psi(\tau)\le\sqrt{\pi/2}$ for all $\tau>0$ by (92),(b) holds because $\Psi(\cdot)$ and $Q^{-1}(\cdot)$ are both decreasing functions and $\epsilon_0\in(0.1,0.9)$, (c) follows from $Q^{-1}(1/10)<1.3$ and the monotonicity of $\Psi(\cdot)$, (d) follows from (92), (e) follows from $\sqrt{\tau^2+4}\le\tau+2$ for $\tau\ge0$.

On the other hand, using (91), (95), and invoking worst case assumptions via $Q^{-1}(2/5)\le0.26$, we get

$$\begin{aligned}\Big|\tfrac{1-\alpha}{\alpha}\ln g\Big(\tfrac{\alpha\sigma}{|1-\alpha|},\epsilon\Big)+\ln\tfrac{g(\sigma,1-\epsilon)}{\theta(\alpha,\sigma)}\Big|&\le\tfrac{0.43}{\alpha}\quad\text{if }\alpha<1,&(114a)\\ \Big|\tfrac{1-\alpha}{\alpha}\ln g\Big(\tfrac{\alpha\sigma}{|1-\alpha|},\epsilon\Big)+\ln\tfrac{g(\sigma,\epsilon)}{\theta(\alpha,\sigma)}\Big|&\le0.43\quad\text{if }\alpha>1.&(114b)\end{aligned}$$

(103e) follows from (111), (112), (113), and (114). □

V. Applications

Under appropriate structural assumptions on w and q, the hypotheses of Theorems 3 and 4 follow from Berry–Essen theorems non-asymptotically and from Central Limit Theorem asymptotically. We will explore these implications for possibly non-stationary (i.e., time-inhomogeneous) memorlyless case in §V-A and for time-homogeneous discrete (i.e., countable state-space) Markov case in §V-B. Then, we will describe the asymptotic simple hypothesis testing problem and show that the Central Limit Theorem and Theorem 3 imply a second order Stein's lemma in §V-C.

A. Memoryless Case

In the memoryless case, w and q are assumed to be product measures. Since both the normalized tilted measure and the likelihood ratio inherit the product structure, the finite length implications of Theorems 3 and 4 are immediate even in the non-stationary case, provided that appropriate moment conditions hold. In models with memory, relating tilting operation to the structural assumptions of the model is the main technical challenge, as we will see for the Markov case in §V-B.

The product measure is determined by the measure of the product events, see [5, Theorem 3.3.1]. The product of two σ-finite measures is a σ-finite measure because Cartesian product of two countable sets is also countable. A more detail discussion products of measures can be found in [5, §3.3].

Assumption 3. w and q are σ-finite product measures of the form $w=\otimes_{t=1}^{n} w_t$ and $q=\otimes_{t=1}^{n} q_t$ for some positive integer n, where w_t and q_t are σ-finite measures on the measurable space $(\mathcal{Y}_t,\mathcal{Y}_t)$, for each $t\in\{1,\dots,n\}$.

The memorylessness assumption implies the additivity log-likelihood ratios: if w and q satisfy Assumption 3, then

$$\mathsf{L}_{w,q}=\sum_{t=1}^{n}\mathsf{L}_{w_t,q_t} \qquad w_\sim\text{-a.e..} \tag{115}$$

If w_t's are probability measures, then (115) and Fubini's theorem [5, Theorem 3.4.4] imply the additivity of $D(w\|q)$ and $\ln\|v_\alpha^{w,q}\|$ privided that corresponding summands, i.e., $D(w_t\|q_t)$'s or $\ln\|v_\alpha^{w_t,q_t}\|$'s, are finite.[16]

Corollary 1. *Let (w,q) be a pair of measures satisfying Assumption 3, for which w_t is a probability measure satisfying $|D(w_t\|q_t)|<\infty$ for all t, and let ξ, σ, T, and δ be*

$$\xi=D(w\|q) \tag{116a}$$
$$\sigma^2=\int(\mathsf{L}_{w,q}-D(w\|q))^2\mathrm{d}w \tag{116b}$$
$$T=\sum_{t=1}^{n}\int|\mathsf{L}_{w_t,q_t}-D(w_t\|q_t)|^3\mathrm{d}w, \tag{116c}$$
$$\delta=0.56^{T}/\sigma^3. \tag{116d}$$

If $\sigma>0$ and $\delta\in(0,1/2)$, then (99b)*,* (99c)*, and* (99d) *of Theorem 3 hold for ξ, σ, and δ defined above.*

Remark 8. If (w,q) satisfies Assumption 3, w_t is a probability measure satisfying $|D(w_t\|q_t)|<\infty$ for all t, and $\sigma=0$, then we are in the trivial-non-singular case, $\mathsf{L}_{w,q}=D(w\|q)$ holds w-a.s., $D(w\|q)=-\ln\|q_\sim\|$, and (58) imply

$$\beta(\varepsilon|_w^q)=e^{-D(w\|q)}|1-\varepsilon|^+ \qquad \forall\varepsilon\ge 0.$$

Hence (99b), (99c), and (99d) of Theorem 3 hold for $\delta=0$.

Proof of Corollary 1. Note that $w_t\prec q_t$ because $|D(w_t\|q_t)|$ is finite. Thus $\mathsf{L}_{w_t,q_t}=\ln\frac{\mathrm{d}w_t}{\mathrm{d}q_t}$ holds w-a.s. Furhtmore, we can calculate the first moment and the variance of $\mathsf{L}_{w,q}$ under w using (115) and Fubini's theorem [5, Theorem 3.4.4]:

$$D(w\|q)=\sum_{t=1}^{n}D(w_t\|q_t)$$
$$\sigma^2=\sum_{t=1}^{n}\int\Big(\ln\tfrac{\mathrm{d}w_t}{\mathrm{d}q_t}-D(w_t\|q_t)\Big)^2\mathrm{d}w_t.$$

Furthermore (99a) holds for ξ, σ, and δ defined in (116) as a result of Berry–Esseen Theorem [94]–[96]. □

Implications of (99d) for $\beta(\cdot|_w^q)$ on closed subintervals of $(0,1)$ becomes more apparent when considered for the stationary memoryless case, i.e., when $w_t=w_1$ and $q_t=q_1$ for all t for some w_1 and q_1. Note that $D(w\|q)=nD(w_1\|q_1)$, $\sigma=\sigma_1\sqrt{n}$, and $\delta=\delta_1/\sqrt{n}$ for some σ_1 and δ_1 in the stationary case. Thus[17] for all $n\ge(\frac{\sigma_1\delta_1+1/\sqrt{2\pi}}{\epsilon_1\sigma_1})^2$ the following equality holds for all $\varepsilon\in[\epsilon_1,1-\epsilon_1]$

$$\beta(\varepsilon|_w^q)=\tfrac{\Delta(\varepsilon)}{\sqrt{2\pi}}e^{-nD(w_1\|q_1)+\sqrt{n}\sigma_1\tau+\frac{\tau^2}{2}}\Psi\big(\sqrt{n}\sigma_1-\tau\big), \tag{117}$$

where $\tau=Q^{-1}(\varepsilon)$, $|\ln\Delta(\varepsilon)|\le\frac{(\sigma_1\sqrt{n}+0.5)\delta_1\sqrt{2\pi}}{\epsilon_1\sqrt{n}-\delta_1}$ and the Gaussian Mills ratio $\Psi(\sqrt{n}\sigma_1-\tau)$ is bounded tightly by (92) because $n\ge(\frac{\sigma_1\delta_1+1/\sqrt{2\pi}}{\epsilon_1\sigma_1})^2$ ensures $\sqrt{n}\sigma_1\ge Q^{-1}(\epsilon_1)\ge\tau$.

Corollary 2. *Let (w,q) be a pair of measures satisfying Assumption 3 for (w_t,q_t) pairs satisfying $\|v_\alpha^{w_t,q_t}\|<(0,\infty)$ for all t for some $\alpha\in(0,\infty)\setminus\{1\}$, and let ξ, σ, T, and δ be*

$$\gamma=1/\|v_\alpha^{w,q}\| \tag{118a}$$
$$\xi=D(v\|w) \tag{118b}$$
$$\sigma^2=(\alpha-1)^2\int(\mathsf{L}_{w,q}-\mathbf{E}_v[\mathsf{L}_{w,q}])^2\mathrm{d}v, \tag{118c}$$
$$T=|\alpha-1|^3\sum_{t=1}^{n}\int|\mathsf{L}_{w_t,q_t}-\mathbf{E}_v[\mathsf{L}_{w_t,q_t}]|^3\mathrm{d}v, \tag{118d}$$
$$\delta=0.56^{T}/\sigma^3, \tag{118e}$$

where $v=\gamma v_\alpha^{w,q}$. If $|\xi|<\infty$, $\sigma>0$, and $\delta\in(0,1/8)$ then (103c)*,* (103d)*, and* (103e) *of Theorem 4 hold for γ, ξ, σ, δ defined above.*

If the hypotheses of Corollary 2 hold, then

$$D(v\|w)=(\alpha-1)\mathbf{E}_v[\mathsf{L}_{w,q}]+\ln\gamma, \tag{119a}$$
$$D(v\|q)=\alpha\mathbf{E}_v[\mathsf{L}_{w,q}]+\ln\gamma, \tag{119b}$$
$$=\tfrac{\alpha\xi-\ln\gamma}{\alpha-1} \tag{119c}$$

for $v=\gamma v_\alpha^{w,q}$. The identities in (119) are present in [13, Theorem 5] for the case when $\mathcal{Y}$ is a finite set and w and q are probability measures. Note that the expression on the right hand side of (119c) shows up in (103b) defining $\eta(\cdot)$.

[16] The additivity of $\ln\|v_\alpha^{w,q}\|$ holds even when w_t's are not necessarily probability measures.

[17] To convert the hypothesis of (99d) to a lower bound on $\sqrt{n}$, we employ the fact that $Q(x)\le\frac{1}{x\sqrt{2\pi}}$ for all $x\ge 0$.

Remark 9. If (w,q) satisfies Assumption 3 for pairs (w_t,q_t) satisfying $\|v_\alpha^{w_t,q_t}\|\in(0,\infty)$ for all t for some $\alpha\in(0,\infty)$, and $|\xi|<\infty$, $\sigma=0$, then we are in the trivial-non-singular case, $\mathsf{L}_{w,q}=\ln\|w_\sim\|-\ln\|q_\sim\|$ holds w-a.s., $D(v\|w)=-\ln\|w_\sim\|$, $D(v\|q)=-\ln\|q_\sim\|$, and (58) imply

$$\beta\big(\tau\big|{}^{w}_{q}\big)=e^{-D(v\|w)}\Big|1-\tau e^{D(v\|q)}\Big|^{+}\qquad\forall\tau\geq 0. \tag{120}$$

Hence, (103c) (103d), and (103e) of Theorem 4 holds for $\delta=0$.

Proof of Corollary 2. Since $\|v_\alpha^{w_t,q_t}\|<(0,\infty)$ for all t, (115) and Fubini's theorem [5, Theorem 3.4.4] imply

$$\begin{aligned}\|v_\alpha^{w,q}\|&=\prod\nolimits_{t=1}^{n}\|v_\alpha^{w_t,q_t}\|,\\ \mathsf{L}_{v,w}&=\sum\nolimits_{t=1}^{n}[(\alpha-1)\mathsf{L}_{w_t,q_t}-\ln\|v_\alpha^{w_t,q_t}\|]\qquad v\text{-a.s..}\end{aligned}$$

Consequently, v inherits the product structure of w and q, i.e.,

$$v=\bigotimes\nolimits_{t=1}^{n}\tfrac{v_\alpha^{w_t,q_t}}{\|v_\alpha^{w_t,q_t}\|}.$$

Thus $\mathsf{L}_{v,w}$ is a sum of random variables that are independent under v. Hence (103a) holds for γ, ξ, σ, δ defined in (118) as a result of Berry–Esseen Theorem [94]–[96]. □

Implications of (103e) for $\beta(\cdot|{}^{q}_{w})$ in the memoryless case becomes more apparent when considered for the stationary case: If $w_t=w_1$ and $q_t=q_1$ for all t for some w_1 and q_1 satisfying $\|v_\alpha^{w_1,q_1}\|<(0,\infty)$ for some $\alpha\in(0,\infty)\setminus\{1\}$, then $\ln\|v_\alpha^{w,q}\|=n\ln\|v_\alpha^{w_1,q_1}\|$, $D(v\|w)=nD(v_1\|w_1)$ for $v_1=v_\alpha^{w_1,q_1}/\|v_\alpha^{w_1,q_1}\|$, $\sigma=\sigma_1\sqrt{n}$ and $\delta=\delta_1/\sqrt{n}$. Let $\tau(\cdot)$ be

$$\tau(\epsilon):=\tfrac{\alpha^2\sigma^2}{2(1-\alpha)^2}+\ln Q\Big(\tfrac{\alpha\sigma}{|1-\alpha|}-Q^{-1}(\epsilon)\Big).$$

For all $n\geq(10\delta_1)^2$ and $\tau_1\in[\tau(3/5),\tau(2/5)]$, then $\exists\Delta$ satisfying both $|\ln\Delta|\leq\frac{0.43}{1\wedge\alpha}+(\sigma_1+(2.3/n))15\sqrt{2\pi}\delta_1$

$$\begin{aligned}&\beta\Big(e^{\tau_1-nD(v_1\|q_1)}\Big|{}^{w}_{q}\Big)\\&\quad=\begin{cases}\Delta\theta_n e^{\frac{\alpha-1}{\alpha}\tau_1-nD(v_1\|w_1)}, & \text{if } \alpha\in(0,1),\\ \|w_{1,\sim}\|^n-\Delta\theta_n e^{\frac{\alpha-1}{\alpha}\tau_1-nD(v_1\|w_1)} & \text{if } \alpha>1,\end{cases}\end{aligned}$$

where $\theta_n=(1/2\pi)^{\frac{1}{2\alpha}}\big(\Psi(\alpha\sigma_1\sqrt{n}/|\alpha-1|)\big)^{\frac{1-\alpha}{\alpha}}\Psi(\sigma_1\sqrt{n})$. As a result of (92): $\theta\sim\frac{\alpha}{|1-\alpha|}\big(\sigma_1\sqrt{2\pi n}\big)^{-1/\alpha}$ as $n\to\infty$.

B. Discrete Markov Case

Tests between time-homogeneous σ-finite Markov measures provide a relatively simple model with memory for the simple hypothesis testing problem. To investigate the problem for Markov measures at a level of generality commensurate with our framework thus far, we must —at the very least— consider σ-bounded transition measures. We defer this generalization to §V-C and consider only the discrete models, i.e., models with a countable state-space, which are rich enough to illustrate some of the most important subtleties of the problem and of ensuring the hypotheses of Theorems 3 and 4.

Let us recall standard notational conventions about discrete Markov measures, first. For countable sets $\mathcal{X}$ and $\mathcal{Y}$, a discrete transition measure W from $\mathcal{X}$ into $\mathcal{Y}$, is a function of the from $W:\mathcal{X}\times 2^{\mathcal{Y}}\mapsto[0,\infty]$, satisfying $W(x;\mathcal{E})=\sum_{y\in\mathcal{E}}W(x;\{y\})$ for all $x\in\mathcal{X}$ and $\mathcal{E}\subset\mathcal{Y}$. With customary abuse of notation, we use $W(x;y)$, rather than $W(x;\{y\})$ for singletons and employ W to denote both the transition measure described above and the corresponding countably infinite dimensional matrix with non-negative entries whose rows and columns correspond to elements of $\mathcal{X}$ and $\mathcal{Y}$, respectively. With this interpretation, we define the support $\mathfrak{S}_W$ of the discrete transition measure W as the set of (x,y) pairs for which $W(x;y)$ is positive:

$$\mathfrak{S}_W:=\{(x,y)\in\mathcal{X}\times\mathcal{Y}:\,W(x;y)>0\}. \tag{121}$$

A discrete transition measure W is called

- σ-finite iff $W(x;y)<\infty\ \forall x\in\mathcal{X},y\in\mathcal{Y}$,
- σ-bounded iff $\sup_{x\in\mathcal{X},y\in\mathcal{Y}}W(x;y)<\infty$,
- bounded iff $\sup_{x\in\mathcal{X}}\sum_y W(x;y)<\infty$,
- a transition probability iff $\sum_y W(x;y)=1\ \forall x\in\mathcal{X}$.

For any discrete σ-finite transition measure W from $\mathcal{Y}$ into $\mathcal{Y}$ and $n\in\mathbb{Z}_+$, discrete σ-finite transition measure $W^{\otimes n}$ from $\mathcal{Y}$ into $\mathcal{Y}_1^n$ and discrete transition measure W^n from $\mathcal{Y}$ into $\mathcal{Y}$ are defined via their values for singletons as

$$W^{\otimes n}(y_0;y_1^n):=\prod\nolimits_{j=1}^{n}W(y_{j-1};y_j)\qquad\forall y_0^n\in\mathcal{Y}_0^n, \tag{122a}$$

$$W^{n}(y_0;y_n):=\sum_{y_1^{n-1}\in\mathcal{Y}_1^{n-1}}W^{\otimes n}(y_0;y_1^n)\quad\forall y_0,y_n\in\mathcal{Y}. \tag{122b}$$

For σ-bounded W's, W^n's are not necessarily, even, σ-finite. If W is bounded, then W^n is also bounded and it is the n^{th} power of W as a countably infinite dimensional matrix. If W is a transition probability, then $W^{\otimes n}(y_0;\cdot)$ and $W^n(y_0;\cdot)$ specify the probabilities of n-step trajectories (i.e., paths) and n-step transitions, respectively, for the corresponding discrete Markov chain for the initial state y_0.

The equivalence classes and related concepts about discrete Markov chains, e.g., [1, §4.2], [3, §8.4], extend naturally to σ-finite discrete transition measures, [68]–[70], because they are in essence determined by the support of the transition probabilities as defined (121). Let W be a σ-finite discrete transition measure from $\mathcal{Y}$ to $\mathcal{Y}$ and $\mathfrak{S}_W$ be its support. Then a $y\in\mathcal{Y}$ is accessible from an $x\in\mathcal{Y}$, i.e., $x\to y$, iff $\exists n\in\mathbb{Z}_+$ such that $(x,y)\in\mathfrak{S}_{W^n}$. States x and y are equivalent, i.e. $x\leftrightarrow y$, iff both $x\to y$, and $y\to x$ hold. $\mathfrak{S}_W$ partitions the set of states $\mathcal{Y}$ into disjoint equivalence classes. Each equivalence class $\mathcal{S}$ is either a singleton of the form $\{x\}$ satisfying $\nexists y\in\mathcal{Y}$ such that $x\leftrightarrow y$, or it is a set satisfying for each of its elements x the following condition: $y\in\mathcal{S}$ iff $x\leftrightarrow y$. An equivalence class $\mathcal{S}$ is closed (i.e., essential according to [70]) iff $(x,y)\notin\mathfrak{S}_W$ for all $x\in\mathcal{S}$ and $y\notin\mathcal{S}$. An equivalence class $\mathcal{S}$ is periodic iff $\exists n_0\geq 2$ such that $(x,x)\notin\mathfrak{S}_{W^n}$ for all $x\in\mathcal{S}$ unless n is an intiger multiple of n_0. An equivalence class $\mathcal{S}$ is aperiodic iff it is not periodic. A discrete transition measure W is said to be irreducible if it has a unique equivalence class and indecomposable if it has a unique closed equivalence class.

Both for the simple hypothesis testing problem and lossless source coding problem Markov transition measures of interest have typically been the ones associated with a time-homogeneous irreducible aperiodic Markov chains on a finite state-space. Such Markov chains have a unique stationary distributions and the convergence of n-step transition probabilities to the stationary distribution is geometric [1, Theo-

rem 4.3.5]. For Markov chains with countably infinite state-spaces the convergence of n-step transition probabilities to the stationary distribution is ensured for all initial states for positive-recurrent irreducible aperiodic Markov chains, see [1, 6.26], [3, Theorem 3, §8.6]. Even then, the convergence to the stationary distribution is neither necessarily uniform over the states [1, §6.3.3], nor geometric. For transition probabilities both Dobrushin–Doeblin condition and Doeblin's minorization condition are necessary-and-sufficient conditions for the uniform geometric convergence of n-step transition probabilities to the stationary distribution of the Markov chain in total variation distance over the initial states, see [66]–[68]; we state these conditions for the discrete case. Dobrushin–Doeblin condition immediately leads to specific convergence parameters. Doeblin's minorization condition, on the other hand, can be extended without modification to transition measures that are not transition probabilities.

A discrete transition probability W from $\mathcal{Y}$ to $\mathcal{Y}$ satisfies Dobrushin–Doeblin condition for (k_0, δ_0) iff constants $k_0 \in \mathbb{Z}_+$ and $\delta_0 \in (0,1)$ satisfy, see [66, (1.1)–(1.4)],

$$\left\| W^{k_0}(x;\cdot) - W^{k_0}(y;\cdot) \right\| \le 2\delta_0 \qquad \forall x, y \in \mathcal{Y}. \tag{123}$$

If Dobrushin–Doeblin condition is satisfied, then there exist constants $\gamma \in (0,\infty)$ and $\rho \in (0,1)$ satisfying

$$\|p_W(\cdot) - W^n(y;\cdot)\| < 2\gamma\rho^n \qquad \forall y \in \mathcal{Y}, n \in \mathbb{Z}_+, \tag{124}$$

where $p_W(\cdot)$ is the stationary distribution of W, e.g., (124) holds for $\gamma = 1/\delta_0$ and $\rho = \delta_0^{1/k_0}$, see [67, (0.2)].

A discrete transition measure W from $\mathcal{Y}$ to $\mathcal{Y}$ satisfies Doeblin's minorization condition for (m, γ, ν) iff $m \in \mathbb{Z}_+$, $\gamma \in (0,\infty)$, and ν is a probability measure on $\mathcal{Y}$ satisfying, see [68, Theorem 6.15, Remarks 6.1],

$$W^m(x;y) \ge \gamma\nu(y) \qquad \forall x, y \in \mathcal{Y}. \tag{125}$$

Evidently, (125) holds iff the following condition holds

$$\hbar\Big(\gamma \Big| {}^{W^m(x;\cdot)}_{\nu}\Big) = \gamma \qquad \forall x \in \mathcal{Y}. \tag{126}$$

For any discrete transition measure W from $\mathcal{X}$ to $\mathcal{Y}$ and σ-finite measure p on $\mathcal{X}$, the measure $p \otimes W$ on $\mathcal{X}\times\mathcal{Y}$ is defined via its value on singletons as

$$p \otimes W(\{(x,y)\}) = p(x)\, W(x;y) \qquad \forall x \in \mathcal{X}, y \in \mathcal{Y}. \tag{127}$$

Evidently, if p is σ-finite and W is σ-bounded, then $p \otimes W$ is σ-finite. If p is a probability measure and W is a transition probability, then $p \otimes W$ is a probability measure.

The functional Berry–Esseen Theorem of Nagaev in [67] implies the following lemma through a standard state-space enlargement argument and Jensen's inequality.

Lemma 8. *Let W be a discrete transition probability satisfying* (123) *with the stationary distribution p_W and Q be a σ-finite discrete transition measure satisfying $m_4 < \infty$, where*

$$m_k = \sup_x \sum_y \left| \ln \tfrac{W(x;y)}{Q(x;y)} \right|^k W(x;y) \qquad \forall k \in \mathbb{Z}_+. \tag{128a}$$

For all positive integers n, let the random vector Y_0^{n+1} be distributed as $p_W \otimes W^{\otimes n+1}$. Then the following expectations and limit, defining $\overline{\xi}$ and $\overline{\sigma}$, exist and are finite and S_n is defined as

$$\overline{\xi} = \mathbf{E}\Big[\ln \tfrac{W(\mathsf{Y}_0;\mathsf{Y}_1)}{Q(\mathsf{Y}_0;\mathsf{Y}_1)}\Big], \tag{128b}$$

$$\overline{\sigma}^2 = \lim_{n\to\infty} \mathbf{E}\left[\tfrac{1}{n}\left(\ln \tfrac{W^{\otimes n}(\mathsf{Y}_0;\mathsf{Y}_1^n)}{Q^{\otimes n}(\mathsf{Y}_0;\mathsf{Y}_1^n)} - n\overline{\xi}\right)^2\right], \tag{128c}$$

$$\mathsf{S}_n = \tfrac{1}{\overline{\sigma}\sqrt{n}}\left[\ln \tfrac{W^{\otimes n+1}(\mathsf{Y}_0;\mathsf{Y}_1^{n+1})}{Q^{\otimes n+1}(\mathsf{Y}_0;\mathsf{Y}_1^{n+1})} - (n+1)\overline{\xi}\right]. \tag{128d}$$

If $\overline{\sigma} > 0$, then S_n satisfies

$$|\mathbf{P}(\mathsf{S}_n \le \tau | \mathsf{Y}_0 = y) - \Phi(\tau)| \le \overline{\delta} \qquad \forall \tau \in \mathbb{R}, y \in \mathcal{Y}, \tag{128e}$$

where $\overline{\delta}$ is defined in terms of $\rho = (\delta_0)^{\frac{1}{1+k_0}}$, $M(\rho)$, $c(\rho)$, as

$$M(\rho) = 3\tfrac{3^{k_0}(1+k_0)(2+4\rho)+1}{1-\rho}, \tag{128f}$$

$$c(\rho) = 136(M(\rho))^3 + \tfrac{32}{3}(M(\rho))^2 + 1, \tag{128g}$$

$$\overline{\delta} = c_1 \tfrac{m_3 c(\rho)}{\overline{\sigma}^3\sqrt{n}} + c_2 \tfrac{m_1}{\overline{\sigma}\sqrt{n}} + c_3\big(\tfrac{1+2\rho}{3}\big)^n\big(1 + \tfrac{m_1}{\overline{\sigma}}\big), \tag{128h}$$

for universal constants c_1, c_2, c_3 that do not depend on W or Q. Furthermore, $\exists(n_0, A)$ determined by (k_0, δ_0) such that

$$\overline{\delta} \le \tfrac{A}{\sqrt{n}}\big(1 + \tfrac{m_1}{\overline{\sigma}} + \tfrac{m_3}{\overline{\sigma}^3}\big) \qquad \forall n \ge n_0. \tag{128i}$$

Constants c_1, c_2, c_3 are universal, but Lemma 8 is not universal because the coefficients of the ratios of the moments depend on W via δ_0 and k_0.

Proof of Lemma 8. Note that m_1, m_2, and m_3 are finite as a result of Jensen's inequality and the convexity of $|\cdot|^\alpha$ for $\alpha \ge 1$ because m_4 is finite by assumption. To establish the lemma, we cannot invoke [67, Theorem 1] to the Markov chain of Y_t's because summands of S_n are $\ln \frac{W(\mathsf{Y}_{t-1};\mathsf{Y}_t)}{Q(\mathsf{Y}_{t-1};\mathsf{Y}_t)}$ hence not of the the form $f(\mathsf{Y}_t)$. We will invoke [67, Theorem 1] to the Markov chain of $(\mathsf{Y}_{t-1}, \mathsf{Y}_t)$ pairs.

Let U be a discrete transition measure from $\mathcal{Y}\times\mathcal{Y}$ into $\mathcal{Y}\times\mathcal{Y}$, defined as $U((x_1,x_2);(y_1,y_2)) = \mathbb{1}_{\{x_2=y_1\}} W(x_2;y_2)$. Then

$$U^{\otimes n}(z_0;z_1^n) = \mathbb{1}_{\{z_{0,2}^{n-1} = z_{1,1}^n\}} W^{\otimes n}\big(z_{0,2};z_{1,2}^n\big),$$
$$U^n(z_0;z_n) = W^{n-1}(z_{0,2};z_{n,1})\, W(z_{n,1};z_{n,2}).$$

Thus $p_U = p_W \otimes W$. Furthermore, if W satisfies Dobrushin–Doeblin condition for (k_0, δ_0), then U satisfies Dobrushin–Doeblin condition for (k_0+1, δ_0). If $\mathsf{Z}_1^n \sim p \otimes U^{n-1}$ for $p = \mathbb{1}_{\{y=\cdot\}} \otimes W$, and $\mathsf{Y}_t = \mathsf{Z}_{t,2}$, then $\mathsf{Y}_1^n \sim W^{\otimes n}(y;\cdot)$. Then for m_{11} defined below, rather than m_1, Lemma 8 follows from [67, (0.4) and Theorem 1] for initial distribution $p(z) = \mathbb{1}_{\{y=y_1\}} W(y;y_2)$, transition probability U, and function $f(z) = \ln \frac{W(y_1;y_2)}{Q(y_1;y_2)}$, where $z = (y_1, y_2)$.

$$m_{11} = \mathbf{E}\Big[\Big|\ln \tfrac{W(\mathsf{Y}_0;\mathsf{Y}_1)}{Q(\mathsf{Y}_0;\mathsf{Y}_1)}\Big| \Big| \mathsf{Y}_0 = y\Big].$$

Lemma 8 holds for m_1 because $m_{11} \le m_1$ by definition. □

The conditional Kullaback-Leibler divergence is defined as

$$D(W\|Q|p) := D(p \otimes W \| p \otimes Q). \tag{129}$$

Thus $\overline{\xi}$ defined in (128b) is equal to $D(W\|Q|p_W)$.

Assumption 4. w and q are σ-finite measures on $\mathcal{Y}_0^{n+1}$ satisfying $w = w_0 \otimes W^{\otimes n+1}$ and $q = q_0 \otimes Q^{\otimes n+1}$ for some

countable set $\mathcal{Y}$, σ-finite measures w_0 and q_0 on $\mathcal{Y}$, and σ-finite transition measures W and Q from $\mathcal{Y}$ to $\mathcal{Y}$.

Remark 10. Note that any k-step discrete transition measure W from $\mathcal{Y}_1^k$ to $\mathcal{Y}$ can be described via 1-step discrete transition measure U from $\mathcal{Y}_1^k$ to $\mathcal{Y}_1^k$ of form

$$U\big(a_1^k;b_1^k\big) = \mathbb{1}_{\{a_2^k=b_1^{k-1}\}}\, W(a_k;b_k)$$

Thus, we only consider one-step discrete transition measures in Corollaries 3 and 4 without loss of generality.

Corollary 3. *Let (w,q) be a pair of measures satisfying Assumption 4, for a probability measure w_0 and transition probability W satisfying* (123)*, for which m_4, defined in* (128a)*, and $\mathbf{E}_{w_0}\left[\left(\ln \frac{\mathrm{d}w_0}{\mathrm{d}q_0}\right)^2\right]$ are finite. Let ξ, σ, and δ be*

$$\xi = (n+1)D(W\|Q|p_W) + D(w_0\|q_0), \tag{130a}$$
$$\sigma = \overline{\sigma}\sqrt{n}, \tag{130b}$$
$$\delta = \overline{\delta} + \tfrac{1}{\overline{\sigma}^2 n\sqrt{8\pi e}}\mathbf{E}_{w_0}\Big[\ln^2 \tfrac{\mathrm{d}w_0}{\mathrm{d}q_0}\Big], \tag{130c}$$

for $\overline{\sigma}$ and $\overline{\delta}$ are defined (128c) *and* (128h)*. Then* (99b)*,* (99c)*, and* (99d) *of Theorem 3 hold for ξ, σ, and δ defined above whenever $\sigma>0$ and $\delta\in(0,1/2)$.*

If $w_0 = q_0$, then Corollary 3 follows from Lemma 8 and Theorem 3. To prove Corollary 1 for $w_0 \neq q_0$ case without assuming $|\ln w_0(y) - \ln q_0(y)|$ to be w_0-a.s. bounded, we invoke the following lemma.

Lemma 9. *Let X and Z be a random variable satisfying*

$$\sup_\tau |\mathbf{P}(\mathsf{Z}\le\tau|\mathsf{X}) - \Phi(\tau/\sigma)| \le \overline{\delta} \qquad \text{a.s.}.$$

Let f be a function such that $\mathbf{E}[f(\mathsf{X})]=0$. Then

$$\sup_\tau |\mathbf{P}(\mathsf{Z}+f(\mathsf{X})\le\tau) - \Phi(\tau/\sigma)| \le \overline{\delta} + \tfrac{1}{\sigma^2\sqrt{8\pi e}}\mathbf{E}\big[(f(\mathsf{X}))^2\big].$$

Proof of Lemma 9 is presented in Appendix D.

Proof of Corollary 3. If $\mathsf{Y}_0^{n+1} \sim w$, then by Lemma 8, the hypothesis of Lemma 9 holds for $\mathsf{X} = \mathsf{Y}_0$, $\overline{\delta}$ defined in (128h), $\sigma = \overline{\sigma}\sqrt{n}$, $f(x) = \ln\frac{w_0(x)}{q_0(x)} - D(w_0\|q_0)$, and

$$\mathsf{Z} = \ln\frac{W^{\otimes n+1}\big(\mathsf{Y}_0;\mathsf{Y}_1^{n+1}\big)}{Q^{\otimes n+1}\big(\mathsf{Y}_0;\mathsf{Y}_1^{n+1}\big)} - (n+1)D(W\|Q|p_W).$$

Then the hypothesis of Theorem 3 follows from Lemma 9 for ξ, σ, and δ defined in (130a), (130b), and (130c), respectively, because

$$\mathsf{L}_{w,q} = \mathsf{Z} + f(\mathsf{Y}_0) + D(w_0\|q_0).$$

Then Corollary 3 follows from Theorem 3. □

For the simple hypothesis testing problem, Markov structure —like memorylessness— is hereditary under the tilting operation in the sense that if (w,q) satisfies Assumption 4, then both (v_α, q) and (v_α, w) satisfy Assumption 4 for the same integer n. To demonstrate this property more explicitly, let us introduce the concept of tilted transition measure. Let W and Q be σ-finite discrete transition measures from $\mathcal{X}$ to $\mathcal{Y}$. Then for any $\alpha\in\mathbb{R}$ the order α tilted transition measure $V_\alpha^{W,Q}$ from $\mathcal{X}$ to $\mathcal{Y}$ is defined as

$$\mathfrak{S}_{V_\alpha^{W,Q}} := \mathfrak{S}_W \cap \mathfrak{S}_Q, \tag{131a}$$
$$V_\alpha^{W,Q}(x;y) := W(x;y)^\alpha Q(x;y)^{1-\alpha} \quad \forall (x,y)\in\mathfrak{S}_{V_\alpha^{W,Q}}. \tag{131b}$$

Then for any (w,q) pair satisfying Assumption 4, the tilted measure $v_\alpha^{w,q}$ defined in (77a) satisfies

$$v_\alpha^{w,q} = v_\alpha^{w_0,q_0} \otimes V_\alpha^{W,Q,\otimes(n+1)}. \tag{132}$$

Unlike memorylessness, however, the Markov structure is not preserved with the scaling of $v_\alpha^{w,q}$ with $\|v_\alpha^{w,q}\|$. Nevertheless, one can construct a probability measure with the Markov structure using Doob's h-transform.

Let V be a σ-finite discrete transition measure from $\mathcal{Y}$ to $\mathcal{Y}$ and $h:\mathcal{Y}\mapsto(0,\infty)$ be an eigenfunction of V for a positive eigenvalue λ satisfying $Vh = \lambda h$. Then Doob's h-transform of V, denoted by V_h, is defined pointwise as

$$V_h(x;y) := \tfrac{V(x;y)h(y)}{\lambda h(x)} \qquad \forall x,y\in\mathcal{Y}. \tag{133}$$

Note that V_h is a discrete transition probability from $\mathcal{Y}$ to $\mathcal{Y}$ by construction. Furthermore, corresponding probabilities of n-step trajectories (i.e., paths) and n-step transitions satisfy

$$V_h^{\otimes n}(y_0;y_1^n) := \tfrac{V^{\otimes n}(y_0;y_1^n)h(y_n)}{\lambda^n h(y_0)} \quad \forall y_0^n\in\mathcal{Y}_0^n,\, n\in\mathbb{Z}_+, \tag{134a}$$
$$V_h^n(y_0;y_n) := \tfrac{V^n(y_0;y_n)h(y_n)}{\lambda^n h(y_0)} \quad \forall y_0,y_n\in\mathcal{Y},\, n\in\mathbb{Z}_+. \tag{134b}$$

In the following, we will invoke (122a), (132), and (134) to express $v_\alpha^{w,q}$ for a (w,q) pair satisfying Assumption 4 in terms of $V_h^{\otimes n+1}$ for the case when $V = V_\alpha^{W,Q}$. Thus the existence of Doob's h-transform —at least for the approriate closed equivalence class of $V_\alpha^{W,Q}$— is necessary for our analysis to go through.

If V is irreducible, aperiodic, and finite and $\mathcal{Y}$ is a finite set, then there exists a unique h, of the form $h:\mathcal{Y}\mapsto(0,\infty)$, satisfying $Vh=\lambda h$ for the maximum modulus eigen value λ by Perron-Frobenius theorem, [97, Theorem 8.4.4], as previously noted in [71], [73]–[75], [98]. If V is irreducible, aperiodic, and σ-finite and $\mathcal{Y}$ is a countably infinite set, then a more nuanced picture emerges. For the case when V^n is σ-finite for all positive integers n, e.g., when V is bounded, R-recurrence of V, as defined in [69], implies the existence of a unique h, of the form $h:\mathcal{Y}\mapsto(0,\infty)$, satisfying $Vh=\lambda h$ by [69, Theorem 4.1]. Let us summarize some of the fundamental findings of the spectral theory of infinite non-negative matrices from [69], [70].

Lemma 10. *For any irreducible aperiodic bounded discrete transition measure V from $\mathcal{Y}$ to $\mathcal{Y}$ there exists a $\lambda\in(0,\infty)$ satisfying*

$$\lim_{n\to\infty} \sqrt[n]{V^n(x;y)} = \lambda \qquad \forall x,y\in\mathcal{Y}. \tag{135}$$

Either $\sum_{n=1}^\infty (1/\lambda)^n V^n(x;y) < \infty$ for all x and y in $\mathcal{Y}$ and V is said to be R-transient, or $\sum_{n=1}^\infty (1/\lambda)^n V^n(x;y) = \infty$ for all x and y in $\mathcal{Y}$ and V is said to be R-recurrent. Similarly, for an R-recurrent V, either $\lim_{n\to\infty}(1/\lambda)^n V^n(x;y) = 0$ for all x and y in $\mathcal{Y}$ and V is said to be R-null, or $\lim_{n\to\infty}(1/\lambda)^n V^n(x;y) > 0$ for all x and y in $\mathcal{Y}$ and V is said to be R-positive. Furthermore, if V is R-recurrent, then

there exists a unique (upto a scaling) R-invariant function $h : \mathcal{Y} \mapsto (0,\infty)$ satisfying $Vh = \lambda h$ and a unique (upto a scaling) R-invariant measure $\pi : \mathcal{Y} \mapsto (0,\infty)$ satisfying $\pi V = \lambda\pi$.

Proof of Lemma 10. There exists a non-negative λ satisfying (135) by[18] [69, Theorem A]. λ is positive because $\mathfrak{S}_{V^n} = \mathcal{Y}\times\mathcal{Y}$ for all n large enough and $V^{m+n}(x;x) \geq V^m(x;x)\,V^n(x;x)$ for all $m, n \in \mathbb{Z}_+$ and $x \in \mathcal{Y}$. Furthermore, λ is finite because boundedness of V implies the existence of a real number M satisfying $V\mathbf{1} \leq \mathbf{1}M$; hence satisfying $V^n\mathbf{1} \leq \mathbf{1}M^n$ for all positive integers n. Dichotomies about $\sum_{n=1}^{\infty}(1/\lambda)^n\,V^n(x;y)$ and $\lim_{n\to\infty}(1/\lambda)^n\,V^n(x;y)$ are established in [69, Theorem B]. The existence of a unique invariant function h and measure π for R-recurrent V are established in [69, Theorem 4.1]. □

Evidently, if V_h is the Doob's h-transform of V for the λ in (135), then V is R-transient/R-null/R-positive iff V_h is transient/null-recurrent/positive-recurrent as characterized in [1, Ch. 6], [3, §8.5 Theorem 5]. Since we are building on [67], via Lemma 8, we will only discuss the case when V_h satisfies the Doeblin's minorization condition. Lemma 11 in the following, presents the characterization of R-positivity in terms of R-invariant measure and R-invariant function of [68]–[70], recalled in Lemma 10, and investigates implications of this characterization for Doeblin's minorization condition.

Lemma 11. *Let V be an irreducible aperiodic bounded discrete transition measure from $\mathcal{Y}$ to $\mathcal{Y}$, then V is R-positive and* (135) *holds iff there exist $h : \mathcal{Y} \mapsto (0,\infty)$ and $\pi : \mathcal{Y} \mapsto (0,\infty)$ satisfying $Vh = \lambda h$, $\pi V = \lambda\pi$, $\pi h = \sum_x \pi(x)h(x) < \infty$ for some $\lambda \in (0,\infty)$. Furthermore, let V be R-positive, then V satisfies Doeblin's minorization condition and $\sup_y h(y) < \infty$ iff V_h satisfies Doeblin's minorization condition and $\inf_y h(y) > 0$.*

Proof of Lemma 11. The characterization of R-positivity is established in [69, Criterion III], see[19] also [70, Theorem 6.4].

For any R-positive V, the R-invariant function h satisfies $h = \lambda^{-m}\,V^m h$ for any $m \in \mathbb{Z}_+$. Then Doeblin's minorization condition for V implies $h \geq (\nu h)\lambda^{-m}\gamma\mathbf{1}$. Thus $\inf_y h(y) > 0$. If $\sup_y h(y) < \infty$, then V_h satisfies Doeblin's minorization condition as a result of (122b) and (134).

For any R-positive V, let us define $g : \mathcal{Y} \mapsto (0,\infty)$ to be the pointwise multiplicative inverse of h, i.e., $g(y) = 1/h(y)$ for all $y \in \mathcal{Y}$. Then the Doeblin's minorization condition for V_h and (134) implies $g(y_0)\,V^m(y_0;y_m) \geq \lambda^m\gamma\nu(y_m)g(y_m)$ for all y_0, y_m in $\mathcal{Y}$. On the other hand, $V\mathbf{1} < M\mathbf{1}$ and hence $V^m\mathbf{1} < M^m\mathbf{1}$ hold for a finite M because V is bounded. Thus $\inf_y g(y) \geq (\nu g)(\lambda/M)^m\gamma > 0$. Hence $\sup_y h(y) < \infty$. If we also have $\inf_y h(y) > 0$, then V satisfies Doeblin's minorization condition as a result of (122b) and (134). □

Remark 11. Common limit value λ in (135) is 1 for any recurrent transition probability V. Thus R-positivity of V does not imply Doeblin's minorization condition for V, because there are positive-recurrent transition probabilities that do not satisfy Doeblin's minorization, e.g., birth–death Markov chains [1, §6.4]. One might be tempted to think that Doeblin's minorization condition for V implies the R-positivity of V, because Doeblin's minorization condition for transition probabilities implies positive-recurrence. This, however, is not true there are R-transient[20] and R-null[21] discrete transition measures satisfying Doeblin's minorization condition. A set of sufficient conditions for R-positivity of discrete substochatic transition measures is presented in [100, Theorem 1]; corresponding conditions for bounded discrete transition measures can be obtained using that fact that any bounded transition measure V satisfies $V = \gamma U$ for a substochastic transition measure U and $\gamma \in [1,\infty)$.

Doeblin–Dobbrushin condition assumed for Corollary 3 implies the uniqueness of the closed equivalence class, i.e., indecomposablity, for the Markov chain. The tilting operation described in (131) can, in principle, change the equivalence classes, i.e., the equivalence classes of $V_\alpha^{W,Q}$ and W may be different. One can refrain from coping with the subtleties arising from this issue explicitly and ensure that supports of $V_\alpha^{W,Q}$ and W are the same, i.e., $\mathfrak{S}_{V_\alpha^{W,Q}} = \mathfrak{S}_W$, by assuming either $\mathfrak{S}_W = \mathfrak{S}_Q$, see [54], [71], [101], or $\mathfrak{S}_W = \mathfrak{S}_Q = \mathcal{Y}\times\mathcal{Y}$, see [73], [74]. Evidently, this issue does not show up in the lossless source coding problem considered in [75], [76] because $\mathfrak{S}_Q = \mathcal{Y}\times\mathcal{Y}$. One can, however, avoid making such assumptions in the simple hypothesis testing problem as well by considering the restrictions of the transition measures.

For any discrete transition measure V from $\mathcal{Y}$ to $\mathcal{Y}$ and $\mathcal{S} \subset \mathcal{Y}$, the restriction $V_{|\mathcal{S}}$ of V to $\mathcal{S}$ is a discrete transition measure from $\mathcal{S}$ to $\mathcal{S}$ defined pointwise as

$$V_{|\mathcal{S}}(x;y) = V(x;y) \qquad \forall x, y \in \mathcal{S}. \tag{136}$$

If either $\mathcal{S}$ or its complement $\mathcal{Y}\setminus\mathcal{S}$ is closed, i.e., $(x,y) \notin \mathfrak{S}_V$ either for all $x \in \mathcal{S}$ and $y \notin \mathcal{S}$ or for all $x \notin \mathcal{S}$ and $y \in \mathcal{S}$, then

$$V^n(x;y) = V^n_{|\mathcal{S}}(x;y) \qquad \forall x, y \in \mathcal{S}. \tag{137}$$

Thus if either $\mathcal{S}$ or its complement is closed then the restriction of V^n to $\mathcal{S}$ is equal to the n^{th} power of $V_{|\mathcal{S}}$, i.e., $V^n{}_{|\mathcal{S}} = V^n_{|\mathcal{S}}$.

Corollary 4. *Let (w,q) be a pair of measures satisfying Assumption 4, $\mathcal{S}$ be a closed equivalence class of $\mathfrak{S}_W \cap \mathfrak{S}_Q$ satisfying $w_0(y)\,q_0(y) = 0$ for all $y \notin \mathcal{S}$ and $\|v_\alpha^{w_0,q_0}\| \in (0,\infty)$ for some $\alpha \in (0,\infty)\setminus\{1\}$. If the restriction of $V_\alpha^{W,Q}$ to $\mathcal{S}$ is a bounded R-positive transition measure satisfying Doeblin's minorization condition with a bounded R-invariant function $h : \mathcal{S} \mapsto (0,\infty)$, i.e., $\sup_{y\in\mathcal{S}} h(y) < \infty$, then $\inf_{y\in\mathcal{S}} h(y) > 0$, and there exist $\lambda \in (0,\infty)$, $k_0 \in \mathbb{Z}_+$, $\delta_0 \in (0,1)$ satisfying*

$$\lim_{n\to\infty} \sqrt[n]{V_\alpha^{W,Q,n}(x;y)} = \lambda \qquad \forall x, y \in \mathcal{S}, \tag{138a}$$

$$\left\| V_h^{k_0}(x;\cdot) - V_h^{k_0}(y;\cdot) \right\| \leq 2\delta_0 \qquad \forall x, y \in \mathcal{S}, \tag{138b}$$

[18] The existence of positive, possibly infinite, λ satisfying (135) also follows from [70, Theorem 6.1] and Hadamard formula, [99, Lemma 1.1.8].

[19] The same characterization extends to the general transition measures for $\pi h = \int h\,\mathrm{d}\pi$, as well, see [68, Definition 5.3] and [68, Proposition 5.8].

[20] The transition measure $V(x;y) = \frac{1}{8}\mathbb{1}_{\{y=1\}} + \frac{x^2+2x}{2(x+2)^2}\mathbb{1}_{\{y=x+1\}}$ from $\mathbb{Z}_+$ to $\mathbb{Z}_+$ satisfies $Vh = (1/2)h$ for $h(x) = (x+1)^2$ and V is R-transient because $V_h(x;y) = \frac{1}{(x+1)^2}\mathbb{1}_{\{y=1\}} + \frac{x^2+2x}{(x+1)^2}\mathbb{1}_{\{y=x+1\}}$ is transient.

[21] The transition measure $V(x;y) = \frac{1}{4}\mathbb{1}_{\{y=1\}} + \frac{x}{2(x+2)}\mathbb{1}_{\{y=x+1\}}$ from $\mathbb{Z}_+$ to $\mathbb{Z}_+$ satisfies $Vh = (1/2)h$ for $h(x) = x+1$ and V is R-null because $V_h(x;y) = \frac{1}{x+1}\mathbb{1}_{\{y=1\}} + \frac{x}{x+1}\mathbb{1}_{\{y=x+1\}}$ is null-recurrent.

where V_h is a transition probability from $\mathcal{S}$ to $\mathcal{S}$ defined as

$$V_h(x;y) = \tfrac{h(y)}{\lambda h(x)}\, V_\alpha^{W,Q}(x;y) \qquad \forall x, y \in \mathcal{S}, \tag{138c}$$

$$\varsigma = \left(\sup\nolimits_{x,y\in\mathcal{S}} \tfrac{h(y)}{h(x)}\right)^{\frac{1}{|1-\alpha|}}. \tag{138d}$$

If m_4, defined below in (138e)*, is finite, then the expectations and limit, defining $\overline{\xi}$ and $\overline{\sigma}$ exist and are finite*

$$m_k = \sup_{x\in\mathcal{S}} \sum_{y\in\mathcal{S}} \left|\ln \tfrac{V_h(x;y)}{W(x;y)}\right|^k V_h(x;y) \qquad \forall k \in \mathbb{Z}_+, \tag{138e}$$

$$\overline{\xi} = \mathbf{E}\Big[\ln \tfrac{V_h(\mathsf{Y}_0;\mathsf{Y}_1)}{W(\mathsf{Y}_0;\mathsf{Y}_1)}\Big], \tag{138f}$$

$$\overline{\sigma}^2 = \lim_{n\to\infty} \mathbf{E}\left[\tfrac{1}{n}\left(\ln \tfrac{V_h^{\otimes n}(\mathsf{Y}_0;\mathsf{Y}_1^n)}{W^{\otimes n}(\mathsf{Y}_0;\mathsf{Y}_1^n)} - n\overline{\xi}\right)^2\right], \tag{138g}$$

where Y_0^{n+1} is distributed as $p_{V_h} \otimes V_h^{\otimes n+1}$ for all n. If $\overline{\sigma} > 0$ $\mathbf{E}_{v_0}\Big[\ln^2 \tfrac{\mathrm{d}v_0}{\mathrm{d}w_0}\Big] < \infty$ for $v_0 = \tfrac{v_\alpha^{w_0,q_0}}{\|v_\alpha^{w_0,q_0}\|}$, and $\sqrt{n} \geq \tfrac{|1-\alpha|\varsigma}{\alpha Q^{-1}(0.4)\overline{\sigma}}$ then (103e) *of Theorem 4 hold for all $\tilde{\eta} \in [\varsigma\eta(3/5), \eta^{(2/5)}/\varsigma]$ for a Δ satisfying $|\ln\Delta| \leq \tfrac{0.43}{1\wedge\alpha} + (\sigma+2.3)15\sqrt{2\pi}\delta + \varsigma$ for γ, ξ, σ, δ defined in the following whenever $\delta \in (0, 1/10)$:*

$$\gamma = 1/(\lambda^{n+1}\|v_\alpha^{w_0,q_0}\|) \tag{138h}$$

$$\xi = (n+1)D(V_h\|W|p_{V_h}) + D(v_0\|w_0), \tag{138i}$$

$$\sigma = \overline{\sigma}\sqrt{n}, \tag{138j}$$

$$\delta = \overline{\delta} + \tfrac{1}{\overline{\sigma}^2 n\sqrt{8\pi e}} \mathbf{E}_{v_0}\Big[\ln^2 \tfrac{\mathrm{d}v_0}{\mathrm{d}w_0}\Big], \tag{138k}$$

where $\overline{\delta}$ is given in (128h) *for m_k in* (138e)*, $M(\rho)$ in* (128f)*, $c(\rho)$ in* (128g)*, and $\rho = (\delta_0)^{\frac{1}{1+k_0}}$ for δ_0 and k_0 in* (138b)*.*

It is worth mentioning that if closed equivalence class $\mathcal{S}$ of Corollary 4 is a finite subset of $\mathcal{Y}$, then the restriction of $V_\alpha^{W,Q}$ to $\mathcal{S}$ is always an R-positive, bounded transition measure satisfying Doeblin's minorization condition and $\mathbf{E}_{v_0}\Big[(\ln v_0 - \ln w_0)^2\Big]$ is always finite.

Proof of Corollary 4. Note that $\mathcal{S}$ is a closed equivalence class of the order α tilted transition measure $V_\alpha^{W,Q}$, as well. As a result $V_{\alpha|\mathcal{S}}^{W,Q}$, i.e., the restriction of $V_\alpha^{W,Q}$ to $\mathcal{S}$, is irreducible. The restriction is also aperiodic because it satisfies Doeblin's minorization condition. Since the equivalence class $\mathcal{S}$ is closed,

$$V_{\alpha|\mathcal{S}}^{W,Q,n}(x;y) = V_\alpha^{W,Q,n}(x;y) \qquad \forall x, y \in \mathcal{S}, n \in \mathbb{Z}_+.$$

Then (138a) holds for a $\lambda \in (0,\infty)$ by Lemma 10, and V_h defined in (138c) is transition measure satisfying Doeblin's minorization condition and R-invariant function h satisfies $\inf_{y\in\mathcal{S}} h(y) > 0$ by Lemma 11. Then V_h satisfies Dobrushin–Doeblin condition and there exist $k_0 \in \mathbb{Z}_+$ and $\delta_0 \in (0,1)$ satisfying (138b) because these two conditions are equivalent for transition probabilities.

Note that $w_\sim\big(\mathcal{Y}_0^{n+1} \setminus \mathcal{S}_0^{n+1}\big) = 0$. Let the measure $s \prec w_\sim$ be

$$s\big(y_0^{n+1}\big) = q\big(y_0^{n+1}\big)\left(\tfrac{h(y_{n+1})}{h(y_0)}\right)^{\frac{1}{1-\alpha}} \qquad \forall y_0^{n+1} \in \mathcal{S}_0^{n+1}.$$

Then

$$s/\varsigma \leq q \leq \varsigma s \qquad w_\sim\text{-a.e..} \tag{139}$$

Note that $\gamma v_\alpha^{w,s} = v_0\, V_h^{\otimes(n+1)}$. If $\mathsf{Y}_0^{n+1} \sim \gamma v_\alpha^{w,s}$, then by Lemma 8, the hypothesis of Lemma 9 holds for $\mathsf{X} = \mathsf{Y}_0$, $\overline{\delta}$ defined in (128h), $\sigma = \overline{\sigma}\sqrt{n}$, $f(x) = \ln\frac{v_0(x)}{w_0(x)} - D(v_0\|w_0)$, and

$$\mathsf{Z} = \ln \tfrac{V_h^{\otimes n+1}(\mathsf{Y}_0;\mathsf{Y}_1^{n+1})}{W^{\otimes n+1}(\mathsf{Y}_0;\mathsf{Y}_1^{n+1})} - (n+1)D(V_h\|W|p_{V_h}).$$

Then for (w,s) the hypothesis of Theorem 4 follows from Lemma 9 for γ, ξ, σ, and δ defined in (138h), (138i), (138j), and (138k), respectively, because

$$\mathsf{L}_{\gamma v_\alpha^{w,s},w} = \mathsf{Z} + f(\mathsf{Y}_0) + D(v_0\|w_0).$$

Thus (103c), (103d), and (103e) of Theorem 4 hold for $\beta(\cdot|_s^w)$, rather than $\beta\big(\cdot\big|_q^w\big)$. On the other hand as result of (57b), (57d), of Lemma 5 and (139), we have

$$\beta\Big(\cdot\Big|_{1/\varsigma}^{w}\Big) = \beta\Big(\cdot\Big|_{s/\varsigma}^{w}\Big) \leq \beta\big(\cdot\big|_q^w\big) \leq \beta\big(\cdot\big|_{\varsigma s}^w\big) = \beta\big(\tfrac{\cdot}{\varsigma}\big|_s^w\big).$$

Then (103e) of Theorem 4 hold for all $\tilde{\eta} \in [\varsigma\eta(3/5), \eta^{(2/5)}/\varsigma]$ for a Δ satisfying $|\ln\Delta| \leq \frac{0.43}{1\wedge\alpha} + (\sigma+2.3)15\sqrt{2\pi}\delta + \varsigma$ as a result of (103e) of Theorem 4 for $\beta(\cdot|_s^w)$. Note that $\sqrt{n} \geq \frac{|1-\alpha|\ln\varsigma}{\alpha Q^{-1}(0.4)\overline{\sigma}}$ ensures $\eta^{(2/5)}/\varsigma \geq \varsigma\eta(3/5)$ because

$$\begin{aligned}\ln\tfrac{\eta(2/5)}{\varsigma^2\eta(3/5)} &= \tfrac{2\alpha\sigma Q^{-1}(0.4)}{|1-\alpha|} - 2\ln\varsigma + \ln\tfrac{\Psi\left(\frac{\alpha\sigma}{|1-\alpha|} - Q^{-1}(2/5)\right)}{\Psi\left(\frac{\alpha\sigma}{|1-\alpha|} + Q^{-1}(2/5)\right)} \\ &\geq \tfrac{2\alpha Q^{-1}(0.4)}{|1-\alpha|}\left[\overline{\sigma}\sqrt{n} - \tfrac{|1-\alpha|\ln\varsigma}{\alpha Q^{-1}(0.4)}\right].\end{aligned}$$

□

C. Asymptotic Case

To pose the asymptotic simple hypothesis testing problem and describe the asymptotic bounds in a generality that is commensurate with the framework we have employed thus far, let us first define the transition measures (i.e., the measure kernels), [5, §10.7], [102, §I.6].

Definition 6. Let $(\mathcal{X},\mathcal{A})$ and $(\mathcal{Y},\mathcal{Y})$ be measurable spaces and W be a mapping from $\mathcal{X}\times\mathcal{Y}$ to $[0,\infty]$. Then W is called a transition measure from $(\mathcal{X},\mathcal{A})$ into $(\mathcal{Y},\mathcal{Y})$ iff the following two conditions hold:

(i) For every $x \in \mathcal{X}$, the function $W(x;\cdot)$ is a measure on $\mathcal{Y}$.
(ii) For every $\mathcal{E} \in \mathcal{Y}$, the function $W(\cdot;\mathcal{E})$ is $\mathcal{A}$-measurable.

If $W(x;\cdot)$ is a probability measure for all $x \in \mathcal{X}$, then W is called a transition probability (probability kernel).

A transition measure W from $(\mathcal{X},\mathcal{A})$ into $(\mathcal{Y},\mathcal{Y})$ is said to be finite if $W(x;\mathcal{Y}) < \infty$ for all $x \in \mathcal{X}$ and σ-finite if $W(x;\cdot)$ is σ-finite for all $x \in \mathcal{X}$. Similarly, W is said to be bounded if $W(\cdot;\mathcal{Y})$ is a bounded function and σ-bounded if there exists a (countable) $\mathcal{Y}$-measurable partition $\{\mathcal{E}_n\}_{n\in\mathbb{Z}}$ of $\mathcal{Y}$ such that $W(\cdot;\mathcal{E}_n)$ is a bounded[22] function for each n.

Lemma 12. *Let W be a σ-bounded transition measure from $(\mathcal{X},\mathcal{A})$ into $(\mathcal{Y},\mathcal{Y})$ and ν be a σ-finite measure on $\mathcal{A}$. Then there exists a unique σ-finite measure μ on $\mathcal{A}\otimes\mathcal{Y}$ with*

$$\mu(\mathcal{E}_x \times \mathcal{E}_y) = \int_{\mathcal{E}_x} W(x;\mathcal{E}_y)\nu(\mathrm{d}x) \quad \forall \mathcal{E}_x \in \mathcal{A}, \mathcal{E}_y \in \mathcal{Y}. \tag{140}$$

[22] We refrain from introducing concepts of Σ-boundedness and Σ-finiteness introduced in [102, §I.6] for conceptual parsimony. Note that every σ-finite transition measure is also a Σ-finite transition measure and every σ-bounded transition measure is also a Σ-bounded transition measure for the definitions employed in in [102, §I.6].

In addition for any non-negative $\mathcal{X}\otimes\mathcal{Y}$-measurable function f the mapping $x\mapsto\int_{\mathcal{Y}} f(x,y)\,W(x;\mathrm{d}y)$ defines a $\mathcal{X}$-measurable non-negative function satisfying

$$\int_{\mathcal{X}\times\mathcal{Y}} f(x,y)\mu(\mathrm{d}(x,y))=\int_{\mathcal{X}}\int_{\mathcal{Y}} f(x,y)\,W(x;\mathrm{d}y)\nu(\mathrm{d}x). \quad (141)$$

Lemma 12 is a generalization of [5, Theorem 10.7.2] for probability measures to σ-finite measures, which follows from [102, Proposition 6.9 and Theorem 6.11].

Remark 12. With the construction given in Assumption 4, w and q are guaranteed to be σ-finite for σ-finite discrete transition measures W and Q because $\mathcal{Y}$ is countable. When we apply the same construction to transitions measures of the form described in Definition 6, we rely on [102, Theorem 6.11] to ensure that w and q are σ-finite; hence we will assume W and Q to be σ-bounded rather merely σ-finite.

Invoking the sign based decomposition $f=f^+-f^-$, we see that (141) holds for any μ-integrable function f. Furthermore, the mapping $x\mapsto\int_{\mathcal{Y}} f(x,y)\,W(x;\mathrm{d}y)$ defines a $\mathcal{X}$-measurable function that is finite ν-a.e. for any μ-integrable function f. We invoke these assertions about μ-integrable functions only for the case of when W is a transition probability and ν is probability measure, which is covered by [5, Theorem 10.7.2].

We adopt the notation similar to [102, (6.23)] and denote $\mu(\cdot)$ defined in (140) by $\nu\otimes W$. Now we are ready to pose the asymptotic simple hypothesis testing problem.

For all positive integers n, let $(\mathcal{Y}_n,\mathcal{Y}_n)$ be a measurable space, and let $\mathcal{Y}_1^n$ and $\mathcal{Y}_1^n$ be corresponding (ordered) Cartesian product set and (ordered) product σ-algebra,

$$\mathcal{Y}_1^n=\mathcal{Y}_1^{n-1}\times\mathcal{Y}_n \qquad \forall n\in\{2,3,\ldots\}, \quad (142a)$$
$$\mathcal{Y}_1^n=\mathcal{Y}_1^{n-1}\otimes\mathcal{Y}_n \qquad \forall n\in\{2,3,\ldots\}. \quad (142b)$$

Let w_1 and q_1 be σ-finite measures on $\mathcal{Y}_1$ and for each n in $\{2,3,\ldots\}$, let W_n and Q_n be σ-bounded transition measures from $(\mathcal{Y}_1^{n-1},\mathcal{Y}_1^{n-1})$ into $(\mathcal{Y}_n,\mathcal{Y}_n)$, then w_1^n and q_1^n are σ-finite measures on $\mathcal{Y}_1^n$ defined as

$$w_1^n=w_1^{n-1}\otimes W_n \qquad \forall n\in\{2,3,\ldots\}, \quad (142c)$$
$$q_1^n=q_1^{n-1}\otimes Q_n \qquad \forall n\in\{2,3,\ldots\}. \quad (142d)$$

The asymptotic simple hypothesis testing problem is the study of the asymptotic behavior of $\beta\left(\cdot\middle|\begin{smallmatrix}q_1^n\\ w_1^n\end{smallmatrix}\right)$ as n tends to infinity, under various structural assumptions on $\{(w_1^n,q_1^n)\}_{n\in\mathbb{Z}_+}$.

Definition 7. Asymptotic simple hypothesis testing problem for $\{(w_1^n,q_1^n)\}_{n\in\mathbb{Z}_+}$ described in (142) is called standard iff w_1 is a probability measure and W_n is a transition probability for all $n\in\{2,3,\ldots\}$.

In the following, we will only be interested in the standard case, for which probability measures $\{w_1^n\}_{n\in\mathbb{Z}_+}$ uniquely extend to a probability measure on $\mathcal{Y}_1^\infty$, as demonstrated by Ionescu Tulcea, see [5, Theorem 10.7.4]. We denote the resulting probability measure and corresponding expectation by $\mathbf{P}_w(\cdot)$ and $\mathbf{E}_w[\cdot]$, respectively.[23] Furthermore, $\{\mathcal{Y}_1^n\}_{n\in\mathbb{Z}_+}$ form a filtration in $\mathcal{Y}_1^\infty$ and $\{(\mathsf{L}_{w_1^n,q_1^n},\mathcal{Y}_1^n)\}_{n\in\mathbb{Z}_+}$ form stochastic sequence in $\mathcal{Y}_1^\infty$ by construction, see [3, §7.1]. Thus

$$\mathbf{E}_w\big[\mathsf{L}_{w_1^n,q_1^n}\big]=D(w_1^n\|q_1^n) \qquad \forall n\in\mathbb{Z}_+,$$

whenever the expectation exists. For a standard asymptotic simple hypothesis testing problem, when the defining limit exists, we define $\overline{D}$ as

$$\overline{D}:=\lim_{n\to\infty}\tfrac{1}{n}\mathbf{E}_w\big[\mathsf{L}_{w_1^n,q_1^n}\big]. \quad (143)$$

The log-likelihood-ratio of a standard asymptotic simple hypothesis testing problem is said to satisfy the weak law of large numbers iff[24] $\overline{D}\in\mathbb{R}$, and

$$\lim_{n\to\infty}\mathbf{P}_w\big(\big|\tfrac{1}{n}\mathsf{L}_{w_1^n,q_1^n}-\overline{D}\big|>\varsigma\big)=0 \qquad \forall\varsigma>0. \quad (144)$$

Lemma 13 (The First Order Asymptotic Stein's Lemma)**.** *If the log-likelihood-ratio of a standard asymptotic simple hypothesis testing problem satisfies weak law of large numbers, i.e.* (144)*, then*

$$\lim_{n\to\infty}\tfrac{1}{n}\ln\beta\left(\varepsilon\middle|\begin{smallmatrix}q_1^n\\ w_1^n\end{smallmatrix}\right)=-\overline{D} \qquad \forall\varepsilon\in(0,1). \quad (145)$$

Lemma 13 follows from (86), i.e., (8), see Appendix C for a proof.

The log-likelihood-ratio of a standard asymptotic simple hypothesis testing problem is said to satisfy the central limit theorem iff $\overline{D}\in\mathbb{R}$, and there exist a $\overline{\sigma}>0$ satisfying

$$\lim_{n\to\infty}\mathbf{P}_w\left(\frac{\mathsf{L}_{w_1^n,q_1^n}-n\overline{D}}{\sqrt{n}}\le\overline{\sigma}x\right)=1-Q(x) \quad \forall x\in\mathbb{R}. \quad (146)$$

Theorem 5 (The Second Order Asymptotic Stein's Lemma)**.** *If the log-likelihood-ratio of a standard asymptotic simple hypothesis testing problem satisfies Central Limit Theorem, i.e.,* (146)*, then*

$$\lim_{n\to\infty}\tfrac{1}{\overline{\sigma}\sqrt{n}}\left(\ln\beta\left(\varepsilon\middle|\begin{smallmatrix}q_1^n\\ w_1^n\end{smallmatrix}\right)+n\overline{D}\right)=Q^{-1}(\varepsilon) \quad \forall\varepsilon\in(0,1). \quad (147)$$

Proof of Theorem 5. First note that (146) is nothing but the convergence of distribution functions of the sequence random variables $\left\{\frac{1}{\overline{\sigma}\sqrt{n}}(\mathsf{L}_{w_1^n,q_1^n}-n\overline{D})\right\}_{n\in\mathbb{Z}_+}$ to standard Gaussian cumulative distribution function, see [2, Definition 1, §3.1]. This is equivalent to the weak convergence of corresponding sequence of probability measures to the standard Gaussian measure on $\mathbb{R}$ by [2, Theorem 2]. On the other hand the weak convergence convergence to the Gaussian measure is metrized by the Kolmogorov distance by [103, Theorem 6] because Gaussian measure is absolutely continuous in the Lebesgue measure. Thus (146) implies that for all $\delta>0$ there exists an n_0 such that

$$\sup_{x\in\mathbb{R}}\left|\mathbf{P}_w\left(\frac{\mathsf{L}_{w_1^n,q_1^n}-n\overline{D}}{\overline{\sigma}\sqrt{n}}>x\right)-Q(x)\right|\le\delta \quad \forall n\ge n_0.$$

Consequently, for all $\delta>0$ there exists an n_0 such that for all $n\ge n_0$, probability measure w_1^n and σ-finite measure q_1^n

[23] We do not need the existence of the unique extension $\mathbf{P}_w(\cdot)$, either to define $\overline{D}$ and $\overline{\sigma}$, or to obtain the results discussed in the following. However, $\mathbf{P}_w(\cdot)$ makes the connection to classical limit theorems for sums of random variables more apparent than it would be otherwise.

[24] Note that $\overline{D}\in\mathbb{R}$ excludes both $\overline{D}=\infty$ and $\overline{D}=-\infty$. Since $\|q_1^n\|$ can be infinite, not only $\overline{D}$, but also $D(w_1^n\|q_1^n)$'s can be $-\infty$, in principle.

satisfies (99a) for $\xi = n\overline{D}$, $\sigma = \overline{\sigma}\sqrt{n}$, and δ. Then for all $\varepsilon\in(0,1)$, $\delta \in (0, \varepsilon \wedge (1-\varepsilon))$, Theorem 3 and (92) imply

$$\limsup_{n\to\infty} \tfrac{1}{\overline{\sigma}\sqrt{n}}\left(\ln\beta\left(\varepsilon\middle|\substack{q_1^n\\w_1^n}\right)+n\overline{D}\right) \le Q^{-1}(\varepsilon)+\tfrac{\delta\sqrt{2\pi}}{(1-\delta-\varepsilon)\wedge(\varepsilon-\delta)},$$
$$\liminf_{n\to\infty} \tfrac{1}{\overline{\sigma}\sqrt{n}}\left(\ln\beta\left(\varepsilon\middle|\substack{q_1^n\\w_1^n}\right)+n\overline{D}\right) \ge Q^{-1}(\varepsilon)-\tfrac{\delta\sqrt{2\pi}}{(1-\delta-\varepsilon)\wedge(\varepsilon-\delta)}.$$

Thus $\liminf$ expression is greater than or equal to $Q^{-1}(\varepsilon)$, and $\limsup$ expression less than or equal to $Q^{-1}(\varepsilon)$. Hence (147) holds for all $\varepsilon\in(0,1)$. □

It is worth mentioning that Theorem 5 is instructive even in the stationary and memoryless case because the absolute third moment of a random variable can be divergent even when Lindenberg condition's are satisfied and Central Limit Theorem holds. Likewise in the time-homogeneous Markovian case and non-stationary memoryless case there are incidences in which hypotheses of Corollaries 1 and 3 fail but the hypothesis of Theorem 5 holds. Theorem 5 demonstrates that Central Limit Theorem implies a second order asymptotic Stein's lemma. Using Theorem 3 one can also establish third order order asymptotic Stein's lemma tying the rate of convergence to normal distribution in the Central Limit Theorem to the third order term in asymptotic Stein's lemma.

VI. Discussion

Non-asymptotic formulations of information transmission problems enjoyed a considerable attention in the recent past [18]–[28]. One-shot bounds, such as those in (13), (14), (17), and (18), were previously interpreted merely as preliminary results for the ultimate asymptotic results. With the surge of interest in non-asymptotic formulations, one-shot bounds have become ends in themselves, heavily studied and cataloged as part of recent efforts to obtain non-asymptotic bounds. However, the one-shot quantities appearing in these bounds —such as $\beta(\cdot|_w^q)$, $\hbar(\cdot|_w^q)$, $F_\ell\big(\cdot\big|_q^w\big)$, $F_\ell^\wedge\big(\cdot\big|_q^w\big)$, and $F_\ell^\vee\big(\cdot\big|_q^w\big)$— are rarely interpreted explicitly as functions of their real parameters; instead, they are customarily studied as functions of the measures w and q for a fixed value of the real parameter. Analyzing these quantities as functions of their real parameters is the paradigm shift that led to Theorems 1 and 2. Studying these one-shot quantities as functions of their real parameters is significantly more tractable than one might initially assume. The convexity of $\beta(\cdot|_w^q)$, which is well-known, proves to be the key tool in relating $\beta(\cdot|_w^q)$, $\hbar(\cdot|_w^q)$, $F_\ell\big(\cdot\big|_q^w\big)$, $F_\ell^\wedge\big(\cdot\big|_q^w\big)$, and $F_\ell^\vee\big(\cdot\big|_q^w\big)$ to one another as functions.

Theorems 1 and 2 can be interpreted as non-asymptotic information spectrum results that establish relationships between one-shot information spectrum quantities for the simple hypothesis testing problem. These relationships have immediate consequences for known one-shot bounds, such as the meta-converse in (13), the maximum-code-theorem in (14), and dependence testing bounds in (17) and (18). More importantly, we believe similar characterizations can be established for other information transmission problems using the non-asymptotic information spectrum method. Following the initial arXiv submission of this work [45] and its subsequent conference appearances [46], [47], the authors of [44] proposed a pairwise error probability framework for one-shot information theory in three arXiv submissions [104]–[106].

Exact characterizations of optimal performance for an information transmission problem are of limited use if they are difficult to compute. Theorems 3, 4, and 5 establish tractable and tight approximations by assuming the Kolmogorov distance to normality for the log-likelihood ratio distribution is bounded by δ. This approach allows us not only to derive the non-asymptotic approximations presented in the corollaries for the non-stationary memoryless and the time-homogeneous Markov cases, but also to demonstrate the distinction between the implications of the Central Limit Theorem and the Berry–Esseen Theorem for the simple hypothesis testing problem.

Bounded Kolmogorov distance hypothesis for the log-likelihood ratio distributions naturally complements the characterization given in (86). Furthermore, the approximating distribution need not be Gaussian, as pointed out in [44, §III A]. One can obtain $o(1)$ approximations for $\ln\beta(\cdot|_w^q)$, such as those in [6, Theorem 1.1], [22, Theorem 18], [54], by using more accurate approximations for $F_{\mathsf{L}}\big(\cdot\big|_q^w\big)$ in (86). Applying these more accurate approximation with tilting, the corresponding improvements for the large deviations regime in [7], [56] can also be recovered.

Han's Information Spectrum Method [11] is a powerful, robust, and unifying approach for studying information transmission problems. Its primary drawback is that it is an asymptotic method by design (e.g., compare the present manuscript with [107]). Non-asymptotic applications of information spectrum ideas have proliferated following the surge of non-asymptotic formulations sparked by [18]–[20]. By interpreting one-shot quantities as functions of their real parameters, we are able to demonstrate previously unknown relations and significantly streamline known ones. We believe these results are not isolated. A non-asymptotic variant of the Information Spectrum Method may prove more analytically tractable than previously assumed.

The broader implications of one-shot spectral representations remain largely unexplored. We conclude by outlining several specific research problems raised by this work. Although we confined our analysis to discrete Markov chains in §V-B, our results can be readily extended to Markov chains with abstract state-spaces. The results from [69], [70] that we employed have already been established for general transition measures in [68], and the Dobrushin–Doeblin condition was originally proposed for Markov chains on measurable spaces without the discreteness assumption [66, (1.1)], [67, (0.1)]. Such a generalization would remove some of the superficial assumptions employed in [72].

An arguably more interesting generalization of the results in §V-B is the removal of Doeblin's minorization condition. While a straightforward proof relying on Lemma 5 and (139) would not be possible in that case, similar results could likely be obtained using Bolthausen's functional Berry–Esseen Theorem [108]. Extending this to a time-inhomogeneous Markov framework is also of interest, though likely more challenging.

Our primary focus in this work was $\beta(\cdot|_w^q)$. When $\beta(\cdot\|_w^q)$ and $\beta(\cdot|_w^q)$ differ due to a probability mass in $w_\sim$ (and hence in $q_\sim$), the analysis of $\beta(\cdot\|_w^q)$ becomes far more nuanced

and likelihood-ratio tests are not necessarily optimal for all ε values for $\beta(\cdot\|{}^{q}_{w})$. Unlike, likelihood-ratio tests, general detectors lack an inherent total order and are only partially ordered, see [19, Appendix B]. Semicontinuous inverses of monotonic functions appear to be the appropriate mathematical tool to analyze $\beta(\cdot\|{}^{q}_{w})$ in this setting. However, it remains unclear to what extent the results in [19, Appendix B] can be improved.

ACKNOWLEDGMENT

Author would like to thank Hao-Chung Cheng for informing him about [43, Proposition 3.2] and [48, Proposition 6.1].
LLM use disclosure: The author primarily used Google's Gemini 3.1-Pro (and, to a lesser extent, free versions of DeepSeek, OpenAI's ChatGPT, and Anthropic's Claude) to explore new concepts, paraphrase text, conduct literature searches, construct counterexamples, and test proof ideas regarding Markov chains. The only direct use of an LLM for a proof was Lemma 9. Gemini generated this proof from a single prompt containing the statement of the lemma, which the author then verified and translated into the paper's notation. The author independently verified every mathematical statement, proof, and reference in the final manuscript and assumes full responsibility for its contents.

APPENDIX

A. Proofs Of The Lemmas About LRQs

Proof of Lemma 1. First, note that $\mathcal{A}^{\wedge}_{\ell}(\cdot|{}^{w}_{q})$ and $\mathcal{A}^{\vee}_{\ell}(\cdot|{}^{w}_{q})$, defined in (40), are both non-increasing in ε by (41), i.e.,

$$\mathcal{A}^{-1}_{\ell}(\varepsilon|{}^{w}_{q}) \subset \mathcal{A}^{-1}_{\ell}(\tau|{}^{w}_{q}) \qquad \forall \varepsilon,\tau\in\mathbb{R}:\varepsilon>\tau.$$

In (37), as ε increases the set that we are taking the infimum over, i.e., $\mathcal{A}^{-1}_{\ell}(\varepsilon|{}^{w}_{q})$, can only become smaller or stay the same. Thus both $F^{\wedge}_{\ell}(\cdot|{}^{w}_{q})$ and $F^{\vee}_{\ell}(\cdot|{}^{w}_{q})$ are non-decreasing on $\mathbb{R}$.

Furthermore, as an immediate consequence of (40) and (41),

$$\mathcal{A}^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q}) = \bigcap\nolimits_{\tau:\tau<\varepsilon} \mathcal{A}^{-1}_{\ell}(\tau|{}^{w}_{q}), \tag{148a}$$

$$\mathcal{A}^{\vee}_{\ell}(\varepsilon|{}^{w}_{q}) = \bigcup\nolimits_{\tau:\tau>\varepsilon} \mathcal{A}^{-1}_{\ell}(\tau|{}^{w}_{q}), \tag{148b}$$

where $\mathcal{A}^{-1}_{\ell}(\cdot|{}^{w}_{q})$ in (148) can either be[25] $\mathcal{A}^{\vee}_{\ell}(\cdot|{}^{w}_{q})$ or $\mathcal{A}^{\wedge}_{\ell}(\cdot|{}^{w}_{q})$.

$$\begin{aligned} F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q}) &\overset{(a)}{=} \sup[0,\infty) \setminus \Big(\bigcap\nolimits_{\tau:\tau<\varepsilon} \mathcal{A}^{-1}_{\ell}(\tau|{}^{w}_{q})\Big),\\ &\overset{(b)}{=} \sup \bigcup\nolimits_{\tau:\tau<\varepsilon} \big([0,\infty) \setminus \mathcal{A}^{-1}_{\ell}(\tau|{}^{w}_{q})\big),\\ &\overset{(c)}{=} \sup\nolimits_{\tau:\tau<\varepsilon} \sup\big([0,\infty) \setminus \mathcal{A}^{-1}_{\ell}(\tau|{}^{w}_{q})\big),\\ &\overset{(d)}{=} \sup\nolimits_{\tau:\tau<\varepsilon} F^{-1}_{\ell}(\tau|{}^{w}_{q}) \end{aligned}$$

where (a) follows from (32), (37a), (40a), (42a), and (148a); (b) follows from De Morgan's law; (c) holds because $\sup\big(\bigcup_{\tau\in\mathcal{T}} \mathcal{B}_{\tau}\big) = \sup_{\tau\in\mathcal{T}}(\sup \mathcal{B}_{\tau})$ for any collection $\{\mathcal{B}_{\tau}\}_{\tau\in\mathcal{T}}$ of sets of the real numbers; (d) follows from (37), (40), and (42). Then (44a) holds because both $F^{\vee}_{\ell}(\tau|{}^{w}_{q})$ and $F^{\wedge}_{\ell}(\tau|{}^{w}_{q})$ non-decreasing in τ.

$$\begin{aligned} F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q}) &\overset{(a)}{=} \inf \bigcup\nolimits_{\tau:\tau>\varepsilon} \mathcal{A}^{-1}_{\ell}(\tau|{}^{w}_{q}),\\ &\overset{(b)}{=} \inf\nolimits_{\tau:\tau>\varepsilon} \inf \mathcal{A}^{-1}_{\ell}(\tau|{}^{w}_{q}),\\ &\overset{(c)}{=} \inf\nolimits_{\tau:\tau>\varepsilon} F^{-1}_{\ell}(\tau|{}^{w}_{q}), \end{aligned}$$

where (a) follows from (32), (37b), (40b), and (148b); (b) holds because $\inf\big(\bigcup_{\tau\in\mathcal{T}} \mathcal{B}_{\tau}\big) = \inf_{\tau\in\mathcal{T}}(\inf \mathcal{B}_{\tau})$ for any collection $\{\mathcal{B}_{\tau}\}_{\tau\in\mathcal{T}}$ of sets of the real numbers; (c) follows from (37) and (40). Then (44b) holds because both $F^{\vee}_{\ell}(\cdot|{}^{w}_{q})$ and $F^{\wedge}_{\ell}(\cdot|{}^{w}_{q})$ are non-decreasing.

We are left with confirming (44c), (44d), (44e), and (44f). Note that (44d) for $\varepsilon > \|w_{\sim}\|_1$ and (44c) for $\varepsilon < 0$ follow from the definition of $\ell_{w,q}$ in (32). Thus we need to establish (44c) and (44d) only for $\varepsilon \in [0, \|w_{\sim}\|_1]$. Let us consider (44c), first. If $F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q}) = \infty$, then (44c) follows from $w_{\sim}(\ell_{w,q}\leq\infty) = \|w_{\sim}\|_1$. If $F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q}) < \infty$ and $w_{\sim}\big(\ell_{w,q}\leq F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})\big) = \infty$, then (44c) holds trivially. If both $F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q}) < \infty$ and $w_{\sim}\big(\ell_{w,q}\leq F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})\big) < \infty$ hold, then $\exists\{\gamma_n\}_{n\in\mathbb{Z}_+} \subset (0,\infty)$ satisfying $\gamma_n \downarrow F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})$ and the continuity of $w_{\sim}$ from above (i.e., at $\emptyset$), implies

$$\begin{aligned} w_{\sim}\big(\ell_{w,q}\leq F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})\big) &= \lim\nolimits_{n\to\infty} w_{\sim}(\ell_{w,q}\leq\gamma_n),\\ &\geq \varepsilon, \end{aligned}$$

where the inequality holds because $w_{\sim}(\ell_{w,q}\leq\gamma)\geq\varepsilon$ for all γ in $(F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q}),\infty)$ by (37a). Thus (44c) holds for the case when both $F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})$ and $w_{\sim}\big(\ell_{w,q}\leq F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})\big)$ are finite, as well.

Let us proceed with establishing (44d). If $F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q}) = -\infty$, then $w\big(\ell_{w,q} < F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q})\big) = 0$ and (44d) holds trivially. If $F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q}) > -\infty$, then $\exists\{\gamma_n\}_{n\in\mathbb{Z}_+} \subset \mathbb{R}$ satisfying $\gamma_n \uparrow F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q})$ and the continuity of w from below (i.e., σ-additivity) implies

$$\begin{aligned} w\big(\ell_{w,q} < F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})\big) &= \lim\nolimits_{n\to\infty} w(\ell_{w,q}\leq\gamma_n),\\ &\leq \varepsilon, \end{aligned}$$

where the inequality holds because $w(\ell_{w,q}\leq\gamma)\leq\varepsilon$ for all γ in $(-\infty, F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q}))$ by (43b). Thus (44d) holds for $F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q}) > -\infty$ case, as well.

$F^{\wedge}_{\ell}(\cdot|{}^{w}_{q})$ and $F^{\vee}_{\ell}(\cdot|{}^{w}_{q})$ are the right and left limits of $F^{\wedge}_{\ell}(\cdot|{}^{w}_{q})$ by (44a) and (44b). Those left and right limits are equal and $F^{\wedge}_{\ell}(\cdot|{}^{w}_{q})$ is continuous everywhere except for a countable set because monotonic functions have at most countably many discontinuities by [86, Theorem 4.30] and $F^{\wedge}_{\ell}(\cdot|{}^{w}_{q})$ is a non-decreasing function. The points of discontinuity should lie between zero and $\|w_{\sim}\|_1$ because $F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q}) = F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q}) = -\infty$ for all $\varepsilon<0$ and $F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q}) = F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q}) = \infty$ for all $\varepsilon > \|w_{\sim}\|_1$. Thus (44e) holds for some countable subset $\mathcal{D}$ of the closed interval $[0, \|w_{\sim}\|_1]$.

(44f) is void for $w \perp q$ case because $w \perp q$ iff $\|w_{\sim}\|_1 = 0$. If $w \not\perp q$, i.e., $\|w_{\sim}\|_1 > 0$, then:

- $F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})$ —and hence $F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q})$— is positive for any ε in $(0, \|w_{\sim}\|_1)$ by (44c), because $w_{\sim}(\ell_{w,q}\leq 0) = 0$.
- $F^{\vee}_{\ell}(\varepsilon|{}^{w}_{q})$ —and hence $F^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})$— is finite for any ε in $(0, \|w_{\sim}\|_1)$ by (44d), because $w(\ell_{w,q}<\infty) = \|w_{\sim}\|_1$.

□

[25]Note that $\{\mathcal{A}^{-1}_{\ell}(\varepsilon_n|{}^{w}_{q})\}_{n\in\mathbb{Z}_+}$ is a non-decreasing sequence of set whenever $\varepsilon_n \downarrow \varepsilon$ and $\{\mathcal{A}^{-1}_{\ell}(\varepsilon_n|{}^{w}_{q})\}_{n\in\mathbb{Z}_+}$ is a non-increasing sequence of set whenever $\varepsilon_n \uparrow \varepsilon$, by (41). Thus $\mathcal{A}^{-1}_{\ell}(\varepsilon_n|{}^{w}_{q}) \uparrow \mathcal{A}^{\vee}_{\ell}(\varepsilon|{}^{w}_{q})$ whenever $\varepsilon_n \downarrow \varepsilon$ and $\mathcal{A}^{-1}_{\ell}(\varepsilon_n|{}^{w}_{q}) \downarrow \mathcal{A}^{\wedge}_{\ell}(\varepsilon|{}^{w}_{q})$ whenever $\varepsilon_n \uparrow \varepsilon$.

B. Proofs Of The Lemmas About PES

Proof of Lemma 3. Both $\ell_{w,q}\wedge\gamma\to 0$ as $\gamma\downarrow 0$ and $\ell_{w,q}\wedge\gamma\leq \ell_{w,q}\wedge 1$ hold q-a.e. for $\gamma\in(0,1)$ and $\int(\ell_{w,q}\wedge 1)\mathrm{d}q<\infty$, i.e., $\hbar\big(1\big|{}^{w}_{q}\big)$ is finite, by hypothesis. Then (52a) follows from the dominated convergence theorem [5, Corollary 2.8.6]. On the other hand $w(\ell_{w,q}\leq\gamma)\leq\hbar\big(\gamma\big|{}^{w}_{q}\big)\leq\|w_{\sim}\|_1$ for all $\gamma\in(0,\infty)$ by (9) and $\lim_{\gamma\uparrow\infty} w(\ell_{w,q}\leq\gamma)=\|w_{\sim}\|_1$ by the countable additivity, i.e., the continuity from below, of w.

Let us proceed with establishing (52c). First using (32), (51a), and the definition of left derivative; then invoking the additivity of the measure q for disjoint events $\{y:\ell_{w,q}\geq\gamma\}$ and $\{y:\gamma>\ell_{w,q}>\tau\}$ for $\tau\in(0,\gamma)$, we get

$$\begin{aligned}\mathrm{D}_{-}\hbar\big(\gamma\big|{}^{w}_{q}\big)&=\lim_{\delta\downarrow 0}\tfrac{1}{\delta}\int_{\gamma-\delta}^{\gamma} q(\ell_{w,q}>\tau)\mathrm{d}\tau,\\&=q(\ell_{w,q}\geq\gamma)+\lim_{\delta\downarrow 0}\int_{\gamma-\delta}^{\gamma}\tfrac{q(\gamma>\ell_{w,q}>\tau)}{\delta}\mathrm{d}\tau.\end{aligned}\tag{149}$$

On the other hand for all $\delta\in[0,\gamma)$ and $\tau\in[\gamma-\delta,\gamma]$ we have

$$0\leq q(\gamma>\ell_{w,q}>\tau)\leq q(\gamma>\ell_{w,q}>\gamma-\delta).\tag{150}$$

The continuity of w from above (i.e., at $\emptyset$) imply

$$\lim_{\delta\downarrow 0} q(\gamma>\ell_{w,q}>\gamma-\delta)=0.\tag{151}$$

(52c) follows from (32), (149), (150), and (151).

Let us proceed with establishing (52d). First using (32), (51a), and the definition of right derivative; then invoking the additivity of the measure q for disjoint events $\{y:\ell_{w,q}>\tau\}$ and $\{y:\tau\geq\ell_{w,q}>\gamma\}$ for $\tau\in(\gamma,\infty)$, we get

$$\begin{aligned}\mathrm{D}_{+}\hbar\big(\gamma\big|{}^{w}_{q}\big)&=\lim_{\delta\downarrow 0}\tfrac{1}{\delta}\int_{\gamma}^{\gamma+\delta} q(\ell_{w,q}>\tau)\mathrm{d}\tau,\\&=q(\ell_{w,q}>\gamma)-\lim_{\delta\downarrow 0}\int_{\gamma}^{\gamma+\delta}\tfrac{q(\tau\geq\ell_{w,q}>\gamma)}{\delta}\mathrm{d}\tau.\end{aligned}\tag{152}$$

On the other hand for all $\delta\in[0,\infty)$ and $\tau\in[\gamma,\gamma+\delta]$ we have

$$0\leq q(\tau\geq\ell_{w,q}>\gamma)\leq q(\gamma+\delta\geq\ell_{w,q}>\gamma).\tag{153}$$

The continuity of w from above (i.e., at $\emptyset$) imply

$$\lim_{\delta\downarrow 0} q(\gamma+\delta\geq\ell_{w,q}>\gamma)=0.\tag{154}$$

(52c) follows from (32), (152), (153), and (154). □

Proof of Lemma 4. $\hbar\big(\gamma\big|{}^{w}_{q}\big)$ is non-decreasing and concave in the pair $(w,\gamma q)$ because the expression in side the infimum in (48b) is a non-decreasing and linear (hence concave) function in the pair $(w,\gamma q)$ for all $\mathcal{E}\in\mathcal{Y}$. Then both (53a) and (53b) hold because pointwise infimum of non-decreasing functions is nondecreasing and pointwise infimum of concave functions is concave. (53c) and (53d) follow from the definition of primitive entropy spectrum in (9). □

C. Proofs of First-Order Stein Lemmas

Proof of Lemma 7. First note that $\mathbf{E}_w[|\mathsf{L}_{w,q}|]$ is finite because $\mathbf{E}_w\big[(\mathsf{L}_{w,q})^2\big]$ is finite and $\mathbf{E}[|\mathsf{X}|]^2\leq\mathbf{E}\big[\mathsf{X}^2\big]$ for any extended real valued random variable X. Then $\mathbf{E}_w[\mathsf{L}_{w,q}]$, and hence $D(w\|q)$ exists and it is a real number.

Chebyshev-Cantelli inequality implies

$$F_{\mathsf{L}}\big(\tau\big|{}^{w}_{q}\big)\leq\tfrac{\sigma^2}{\sigma^2+(D(w\|q)-\tau)^2}\qquad\forall\tau\leq D(w\|q).$$

Then $F_{\mathsf{L}}\big(\tau\big|{}^{w}_{q}\big)\leq\varepsilon$ for all $\tau\leq D(w\|q)-\sigma\sqrt{\frac{1-\varepsilon}{\varepsilon}}$. Thus as a result of (86)

$$\beta(\varepsilon|{}^{q}_{w})=\int_{D(w\|q)-\sigma\sqrt{\frac{1-\varepsilon}{\varepsilon}}}^{\infty} e^{-\tau}\big|F_{\mathsf{L}}\big(\tau\big|{}^{w}_{q}\big)-\varepsilon\big|^{+}\mathrm{d}\tau.$$

Then the upper bound in (87a) follows from $F_{\mathsf{L}}\big(\tau\big|{}^{w}_{q}\big)\leq 1$. Note that Chebyshev-Cantelli inequality also implies

$$1-F_{\mathsf{L}}\big(\tau\big|{}^{w}_{q}\big)\leq\tfrac{\sigma^2}{\sigma^2+(\tau-D(w\|q))^2}\qquad\forall\tau\geq D(w\|q).$$

Thus $1-F_{\mathsf{L}}\big(\tau\big|{}^{w}_{q}\big)\leq\frac{1-\varepsilon}{1+\Delta}$ for all $\tau\geq D(w\|q)+\sigma\sqrt{\frac{\varepsilon+\Delta}{1-\varepsilon}}$. Thus as a result of (86)

$$\begin{aligned}\beta(\varepsilon|{}^{q}_{w})&\geq\int_{D(w\|q)+\sigma\sqrt{\frac{\varepsilon+\Delta}{1-\varepsilon}}}^{\infty} e^{-\tau}\big|F_{\mathsf{L}}\big(\tau\big|{}^{w}_{q}\big)-\varepsilon\big|^{+}\mathrm{d}\tau\\&\geq\int_{D(w\|q)+\sigma\sqrt{\frac{\varepsilon+\Delta}{1-\varepsilon}}}^{\infty} e^{-\tau}\Big|1-\tfrac{1-\varepsilon}{1+\Delta}-\varepsilon\Big|^{+}\mathrm{d}\tau\\&=e^{-D(w\|q)-\sigma\sqrt{\frac{\varepsilon+\Delta}{1-\varepsilon}}}\tfrac{\Delta(1-\varepsilon)}{1+\Delta}.\end{aligned}$$

□

Proof of Lemma 13. As a result of (144) for any $\varsigma>0$ and $\delta>0$ there exists a $n_0(\varsigma,\delta)$ such that

$$\mathbf{P}_w\big(\mathsf{L}_{w_1^n,q_1^n}\leq(\overline{D}-\varsigma)n\big)\leq\delta\qquad\forall n\geq n_0(\varsigma,\delta).$$

Thus as a result of (86) for any $n\geq n_0(\varsigma,\varepsilon)$ have

$$\begin{aligned}\beta\Big(\varepsilon\Big|{}^{q_1^n}_{w_1^n}\Big)&=\int_{(\overline{D}-\varsigma)n}^{\infty} e^{-t}\big|\mathbf{P}_w\big(\mathsf{L}_{w_1^n,q_1^n}\leq t\big)-\varepsilon\big|^{+}\mathrm{d}t,\\&\leq(1-\varepsilon)e^{-(\overline{D}-\varsigma)n}.\end{aligned}$$

Thus

$$\limsup_{n\to\infty}\tfrac{1}{n}\ln\beta\Big(\varepsilon\Big|{}^{q_1^n}_{w_1^n}\Big)\leq-\overline{D}+\varsigma\qquad\forall\varsigma>0.\tag{155}$$

As a result of (144) for any $\varsigma>0$ and $\delta>0$ there exists a $n_1(\varsigma,\delta)$ such that

$$\mathbf{P}_w\big(\mathsf{L}_{w_1^n,q_1^n}>(\overline{D}+\varsigma)n\big)\leq\delta\qquad\forall n\geq n_1(\varsigma,\delta).$$

Thus as a result of (86) for any $n\geq n_1(\varsigma,\frac{1-\varepsilon}{2})$ have

$$\begin{aligned}\beta\Big(\varepsilon\Big|{}^{q_1^n}_{w_1^n}\Big)&=\int_{-\infty}^{\infty} e^{-t}\big|1-\varepsilon-\mathbf{P}_w\big(\mathsf{L}_{w_1^n,q_1^n}>t\big)\big|^{+}\mathrm{d}t,\\&\geq\int_{-(\overline{D}+\varsigma)n}^{\infty} e^{-t}\big|1-\varepsilon-\mathbf{P}_w\big(\mathsf{L}_{w_1^n,q_1^n}>t\big)\big|^{+}\mathrm{d}t,\\&\geq\tfrac{1-\varepsilon}{2}e^{-(\overline{D}+\varsigma)n}.\end{aligned}$$

Thus

$$\liminf_{n\to\infty}\tfrac{1}{n}\ln\beta\Big(\varepsilon\Big|{}^{q_1^n}_{w_1^n}\Big)\geq-\overline{D}-\varsigma\qquad\forall\varsigma>0.\tag{156}$$

(145) follows from (155) and (156). □

D. Proof of Lemma 9

Proof of Lemma 9. Let us denote the random variable $f(\mathsf{X})$ by V. For any random variable Y, the triangle inequality for Kolmogorov distance imply

$$d\big(F_{\mathsf{V}+\mathsf{Z}}(\cdot),\Phi\big(\tfrac{\cdot}{\sigma}\big)\big)\le d(F_{\mathsf{V}+\mathsf{Z}}(\cdot),F_{\mathsf{V}+\mathsf{Y}}(\cdot))+d\big(F_{\mathsf{V}+\mathsf{Y}}(\cdot),\Phi\big(\tfrac{\cdot}{\sigma}\big)\big).$$

Let Y be a zero-mean Gaussian random variable with variance σ^2 independent of X, and hence V. Then

$$F_{\mathsf{V}+\mathsf{Y}}(\tau)=\mathbf{E}\big[\Phi\big(\tfrac{\tau-\mathsf{V}}{\sigma}\big)\big].$$

On the other hand, the law of iterated expectations asserts

$$F_{\mathsf{V}+\mathsf{Z}}(\tau)=\mathbf{E}\big[\mathbf{E}\big[\mathbb{1}_{\{\mathsf{Z}+\mathsf{V}\le\tau\}}\big|\mathsf{X}\big]\big].$$

Thus using first the triangle inequality for expectations, and then the hypothesis of the lemma we get

$$\begin{aligned}|F_{\mathsf{V}+\mathsf{Y}}(\tau)-F_{\mathsf{V}+\mathsf{Z}}(\tau)|&\le\mathbf{E}\big[\big|\,\Phi\big(\tfrac{\tau-\mathsf{V}}{\sigma}\big)-\mathbf{E}\big[\mathbb{1}_{\{\mathsf{Z}+\mathsf{V}\le\tau\}}\big|\mathsf{X}\big]\,\big|\big]\\&\le\overline{\delta}\qquad\qquad\forall\tau.\end{aligned}$$

Thus $d(F_{\mathsf{V}+\mathsf{Z}}(\cdot),F_{\mathsf{V}+\mathsf{Y}}(\cdot))\le\overline{\delta}$.

On the other hand a result of Taylor's theorem, for all $x\in\mathbb{R}$, there exists a ξ_x between 0 and x satisfying

$$\Phi\big(\tfrac{\tau-x}{\sigma}\big)=\Phi\big(\tfrac{\tau}{\sigma}\big)-\tfrac{x}{\sigma}\varphi\big(\tfrac{\tau}{\sigma}\big)-\tfrac{x^2}{2\sigma^2}\left(t\varphi(t)\right)\big|_{t=\frac{\tau-\xi_x}{\sigma}}.$$

On the other hand $\sup_t|t\varphi(t)|=\frac{1}{\sqrt{2\pi e}}$. Thus using $\mathbf{E}[\mathsf{V}]=0$, we get

$$\mathbf{E}\big[\Phi\big(\tfrac{\tau-\mathsf{V}}{\sigma}\big)\big]=\Phi\big(\tfrac{\tau}{\sigma}\big)\pm\tfrac{1}{2\sigma^2\sqrt{2\pi e}}\mathbf{E}\big[\mathsf{V}^2\big]$$

Thus $d\big(F_{\mathsf{V}+\mathsf{Y}}(\cdot),\Phi\big(\tfrac{\cdot}{\sigma}\big)\big)\le\frac{\mathbf{E}[\mathsf{V}^2]}{2\sigma^2\sqrt{2\pi e}}$ and lemma follows. □

REFERENCES

[1] R. G. Gallager, *Stochastic processes : theory for applications*. Cambridge: Cambridge University Press, 2013.

[2] A. N. Shiryaev, *Probability 1*, 3rd ed., ser. Graduate Texts in Mathematics. New York: Springer, 2016.

[3] ——, *Probability 2*, 3rd ed., ser. Graduate Texts in Mathematics. New York: Springer, 2019.

[4] P. R. Halmos, *Measure Theory*, 1st ed., ser. Graduate Texts in Mathematics. New York, NY: Springer, 1974.

[5] V. I. Bogachev, *Measure Theory*. Berlin Heidelberg: Springer-Verlag, 2007.

[6] V. Strassen, "Asymptotische abschätzungen in Shannons Informationstheorie," in *Trans. Third Prague Conf. Inf. Theory*, 1962, pp. 689–723, (https://pi.math.cornell.edu/~pmlut/strassen.pdf).

[7] I. Csiszár and G. Longo, "On the error exponent for source coding and for testing simple statistical hypotheses," *Studia Scientiarum Mathematicarum Hungarica*, vol. 6, pp. 181–191, 1971, (http://real-j.mtak.hu/id/eprint/5457).

[8] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge, UK: Cambridge University Press, 2011.

[9] R. G. Gallager, *Information theory and reliable communication*. New York, NY: John Wiley & Sons, Inc., 1968.

[10] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New York, NY: Wiley-Interscience, 2006.

[11] T. S. Han, *Information-Spectrum Methods in Information Theory*, ser. Stochastic Modelling and Applied Probability (Book 50). Springer Berlin Heidelberg, 2003.

[12] Y. Polyanskiy and Y. Wu, *Information Theory: From Coding to Learning*. Cambridge University Press, 2025.

[13] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[14] E. A. Haroutunian, "Bounds for the exponent of the probability of error for a semicontinuous memoryless channel," *Problems of Information Transmission*, vol. 4, no. 4, pp. 29–39, 1968.

[15] U. Augustin, "The capacity of the product channel," *Information and Control*, vol. 13, no. 6, pp. 594–596, 1968.

[16] E. A. Haroutunian, "Lower bound for error probability in channels with feedback," *Problems of Information Transmission*, vol. 13, no. 2, pp. 107–114, 1977.

[17] U. Augustin, "Noisy channels," Habilitation Thesis, Universität Erlangen-Nürnberg, 1978, (http://bit.ly/3bsWDgG).

[18] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4947–4966, Nov 2009.

[19] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, 2010.

[20] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[21] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.

[22] P. Moulin, "The log-volume of optimal codes for memoryless channels, asymptotically within a few nats," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2278–2313, April 2017.

[23] R. C. Yavas, V. Kostina, and M. Effros, "Third-order analysis of channel coding in the small-to-moderate deviations regime," *IEEE Transactions on Information Theory*, vol. 70, no. 9, pp. 6139–6170, 2024.

[24] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics for most discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7041–7051, Nov 2013.

[25] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends®in Communications and Information Theory*, vol. 11, no. 1-2, pp. 1–184, 2014. [Online]. Available: http://dx.doi.org/10.1561/0100000086

[26] L. Wang, R. Colbeck, and R. Renner, "Simple channel coding bounds," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, June 2009, pp. 1804–1808.

[27] L. Wang and R. Renner, "One-shot classical-quantum capacity and hypothesis testing," *Physical Review Letters*, vol. 108, no. 20, May 2012.

[28] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7693–7710, Nov 2013.

[29] I. Csiszár, "Eine Informationstheoretisehe Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten," *Publications of the Mathematical Institute of Hungarian Academy of Sciences Series A*, vol. 8, no. 1-2, pp. 85–108, 1963.

[30] ——, "Information-type measures of difference of probability distributions and indirect observations," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, no. 3-4, pp. 299–318, 1967.

[31] T. Morimoto, "Markov processes and the h-theorem," *Journal of the Physical Society of Japan*, vol. 18, no. 3, pp. 328–331, 1963.

[32] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 28, no. 1, pp. 131–142, 1966.

[33] F. Osterreicher and I. Vajda, "Statistical information and discrimination," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 1036–1039, 1993.

[34] F. Liese and I. Vajda, "On divergences and informations in statistics and information theory," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4394–4412, 2006.

[35] M. D. Reid and R. C. Williamson, "Information, divergence and risk for binary experiments," *Journal of Machine Learning Research*, vol. 12, no. 22, pp. 731–817, 2011, (arXiv:0901.0356 [stat.ML]). [Online]. Available: http://jmlr.org/papers/v12/reid11a.html

[36] F. Liese, "ϕ-divergences, sufficiency, bayes sufficiency, and deficiency," *Kybernetika*, vol. 48, no. 4, pp. 690–713, 2012. [Online]. Available: https://www.kybernetika.cz/content/2012/4/690

[37] A. Guntuboyina, S. Saha, and G. Schiebinger, "Sharp inequalities for f-divergences," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 104–121, Jan 2014.

[38] S. Chen, M. Effros, and V. Kostina, "Lossless source coding in the point-to-point, multiple access, and random access scenarios," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 6688–6722, 2020.

[39] I. Sason and S. Verdú, "f-divergence inequalities," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 5973–6006, Nov 2016.

[40] J. Liu, "Information theory from a functional viewpoint," Ph.D. dissertation, Princeton University, 2018.

[41] J. Liu, P. Cuff, and S. Verdú, "Resolvability," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2629–2658, 2017.

[42] Y. Liu, Y. Fan, and D. Gündüz, "Tighter information-theoretic generalization bounds via a novel class of change of measure inequalities," *arXiv:2602.07999v4[cs.IT]*, 2026, (arXiv:2602.07999v4[cs.IT]).

[43] K. M. R. Audenaert, M. Mosonyi, and F. Verstraete, "Quantum state discrimination bounds for finite sample size," *Journal of Mathematical Physics*, vol. 53, no. 12, p. 122205, 12 2012, (arXiv.1204.0711. [Online]. Available: https://doi.org/10.1063/1.4768252

[44] N. Elkayam and M. Feder, "Variational formulas for the power of the binary hypothesis testing problem with applications," *arXiv:1601.06810v1 [cs.IT]*, 2016.

[45] B. Nakiboğlu, "The spectral representations of the simple hypothesis testing problem," *arXiv:2601.09564v1 [cs.IT]*, 2026, (arXiv:2601.09564v1[cs.IT]).

[46] ——, "A spectral representation of the simple hypothesis testing and refined approximations," in *14th Iran Workshop on Communication and Information Theory*, 2026. [Online]. Available: http://www.aparat.com/v/sefhpq9

[47] ——, "A spectral representation of the simple hypothesis testing problem," in *2026 IEEE International Symposium on Information Theory (ISIT)*, 2026, pp. 1–6.

[48] H.-C. Cheng and P.-C. Liu, "Multiple quantum hypothesis testing: One-shot pairwise bounds and sharp asymptotics," *arXiv:2606.06246v1 [cs.IT]*, 2026, (arXiv:2606.06246v1[cs.IT]).

[49] J. Lizarribar-Carrillo, G. Vazquez-Vilar, and T. Koch, "Non-asymptotic bounds via the Nussbaum–Szkoła mapping for quantum hypothesis testing," *arXiv:2610.39784v1[quant-ph]*, 2026, (arXiv:2610.39784v1[quant-ph]).

[50] S. Verdú, "Relative information spectra with applications to statistical inference," *AIMS Mathematics*, vol. 9, no. 12, pp. 35 038–35 090, 2024. [Online]. Available: https://www.aimspress.com/article/doi/10.3934/math.20241668

[51] ——, "Correction: Relative information spectra with applications to statistical inference," *AIMS Mathematics*, vol. 10, no. 4, pp. 9322–9323, 2025. [Online]. Available: https://www.aimspress.com/article/doi/10.3934/math.2025429

[52] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: Non-asymptotics and asymptotics," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 777–795, 2014.

[53] A. A. Yushkevich, "On limit theorems connected with the concept of entropy of Markov chains," *Uspekhi Mat. Nauk*, vol. 8, no. 5(57, pp. 177–180, 1953.

[54] A. V. Nagaev, "An asymptotic formula for the neyman-pearson risk in discriminating between two Markov chains," *Journal of Mathematical Sciences*, vol. 111, no. 3, pp. 3582–3591, 2002. [Online]. Available: https://doi.org/10.1515/156939203322733282

[55] L. Gavalakis and I. Kontoyiannis, "Fundamental limits of lossless data compression with side information," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2680–2692, 2021.

[56] G. Vazquez-Vilar, A. Guillen i Fabregas, T. Koch, and A. Lancho, "Saddlepoint approximation of the error probability of binary hypothesis testing," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 2306–2310.

[57] B. Nakiboğlu, "A simple derivation of the refined sphere packing bound under certain symmetry hypotheses," *Turkish Journal Of Mathematics*, vol. 44, no. 3, pp. 919–948, 2020, (arXiv:1904.12780 [cs.IT]).

[58] ——, "A Simple Derivation of the Refined SPB for the Constant Composition Codes," in *2019 IEEE International Symposium on Information Theory (ISIT)*, Paris, France, July 2019, pp. 2659–2663.

[59] V. Lungu and I. Kontoyiannis, "The optimal finite-sample error probability in asymmetric binary hypothesis testing," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 831–836, (arXiv:2404.09605v2 [cs.IT]).

[60] A. Theocharous and I. Kontoyiannis, "Lossless data compression at pragmatic rates," in *2025 IEEE International Symposium on Information Theory (ISIT)*, 2025, pp. 1–6.

[61] A. Theocharous, L. Gavalakis, and I. Kontoyiannis, "Pragmatic lossless compression: Fundamental limits and universality," *IEEE Transactions on Information Theory*, vol. 72, no. 7, pp. 4587–4601, 2026, (arXiv:2501.10103v4 [cs.IT]).

[62] Y. Altuğ and A. B. Wagner, "Refinement of the sphere-packing bound: Asymmetric channels," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1592–1614, March 2014.

[63] N. R. Chaganty and J. Sethuraman, "Strong Large Deviation and Local Limit Theorems," *The Annals of Probability*, vol. 21, no. 3, pp. 1671 – 1690, 1993. [Online]. Available: https://doi.org/10.1214/aop/1176989136

[64] R. R. Bahadur and R. R. Rao, "On deviations of the sample mean," *The Annals of Mathematical Statistics*, vol. 31, no. 4, pp. 1015–1027, 12 1960.

[65] H.-C. Cheng and B. Nakiboğlu, "Refined strong converse for the constant composition codes," in *2020 IEEE International Symposium on Information Theory (ISIT)*, June 2020, pp. 2149–2154, (arXiv:2002.11414 [cs.IT]).

[66] S. V. Nagaev, "Some limit theorems for stationary Markov chains," *Theory of Probability & Its Applications*, vol. 2, no. 4, pp. 378–406, 1957. [Online]. Available: https://doi.org/10.1137/1102029

[67] ——, "More exact statement of limit theorems for homogeneous Markov chains," *Theory of Probability & Its Applications*, vol. 6, no. 1, pp. 62–81, 1961.

[68] E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators*, ser. Cambridge Tracts in Mathematics (83). Cambridge University Press, 1984.

[69] D. Vere-Jones, "Ergodic properties of nonnegative matrices. I." *Pacific Journal of Mathematics*, vol. 22, no. 2, pp. 361–386, 1967.

[70] S. E., *Non-negative Matrices and Markov Chains*, 2nd ed., ser. Springer Series in Statistics. New York, NY, USA: Springer, 2006.

[71] S. Watanabe and M. Hayashi, "Finite-length analysis on tail probability for Markov chain and application to simple hypothesis testing," *The Annals of Applied Probability*, vol. 27, no. 2, pp. 811–845, 2017. [Online]. Available: https://doi.org/10.1214/16-AAP1216

[72] L. H. Koopmans, "Asymptotic Rate of Discrimination for Markov Processes," *The Annals of Mathematical Statistics*, vol. 31, no. 4, pp. 982 – 994, 1960. [Online]. Available: https://doi.org/10.1214/aoms/1177705671

[73] K. Nakagawa and F. Kanaya, "On the converse theorem in statistical hypothesis testing for Markov chains," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 629–633, 1993.

[74] S. Natarajan, "Large deviations, hypotheses testing, and source coding for finite Markov chains," *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 360–365, 1985.

[75] K. Vašek, "On the error exponent for ergodic Markov source," *Kybernetika*, vol. 16, no. 4, pp. 318–329, 1980.

[76] L. Davisson, G. Longo, and A. Sgarro, "The error exponent for the noiseless encoding of finite ergodic Markov sources," *IEEE Transactions on Information Theory*, vol. 27, no. 4, pp. 431–438, 1981.

[77] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, 1st ed., ser. Grundlehren Text Editions. Heidelberg: Springer-Verlag Berlin, 2001.

[78] ——, *Convex Analysis and Minimization Algorithms I: Fundamentals*, 1st ed., ser. Grundlehren der mathematischen Wissenschaften. Heidelberg: Springer-Verlag Berlin, 2013.

[79] E. P. Klement, R. Mesiar, and E. Pap, "Quasi- and pseudo-inverses of monotone functions, and the construction of t-norms," *Fuzzy Sets and Systems*, vol. 104, no. 1, pp. 3–13, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165011498002528

[80] P. Viceník, "A note to a construction of t-norms based on pseudo-inverses of monotone functions," *Fuzzy Sets and Systems*, vol. 104, no. 1, pp. 15–18, 1999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016501149800253X

[81] C. Feng, W. H., X. Tu, and J. Kowalski, "A note on generalized inverses of distribution function and quantile transformation," *pplied Mathematics*, no. 3, pp. 2098–2100, 2012. [Online]. Available: http://dx.doi.org/10.4236/am.2012.312A289

[82] N. Falkner and G. Teschl, "On the substitution rule for lebesgue–stieltjes integrals," *Expositiones Mathematicae*, vol. 30, no. 4, pp. 412–418, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0723086912000588

[83] P. Embrechts and M. Hofert, "A note on generalized inverses," *Mathematical Methods of Operations Research*, vol. 77, no. 3, pp. 4230–432, 2013. [Online]. Available: https://doi.org/10.1007/s00186-013-0436-7

[84] A. de La Fortelle, "Generalized inverses of increasing functions and lebesgue decomposition," *Markov Processes And Related Fields*, Aug 2020. [Online]. Available: https://minesparis-psl.hal.science/hal-03113748

[85] P. Wacker, "Please, not *another* note about generalized inverses," *arXiv:2306.06989v1 [math.CA]*, 2023. [Online]. Available: https://arxiv.org/abs/2306.06989v1

[86] W. Rudin, *Principles of Mathematical Analysis*. New York, NY: McGraw-Hill, 1976.

[87] R. M. Dudley, *Real analysis and probability*. New York, NY: Cambridge University Press, 2002, vol. 74.

[88] F. Topsøe, "Entropy and equilibrium via games of complexity," *Physica A: Statistical Mechanics and its Applications*, vol. 340, no. 1, pp. 11–31, 2004, news and Expectations in Thermostatistics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437104003930

[89] M. V. Burnashev, "On stein's lemma in hypotheses testing in general non-asymptotic case," *Statistical Inference for Stochastic Processes*, vol. 26, no. 1, pp. 89–97, 2023.

[90] A. Baricz, "Mills' ratio: Monotonicity patterns and functional inequalities," *Journal of Mathematical Analysis and Applications*, vol. 340, no. 2, pp. 1362–1370, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022247X07011730

[91] A. Gasull and F. Utzet, "Approximating mills ratio," *Journal of Mathematical Analysis and Applications*, vol. 420, no. 2, pp. 1832–1853, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022247X14004764

[92] P. Borjesson and C.-E. Sundberg, "Simple approximations of the error function q(x) for communications applications," *IEEE Transactions on Communications*, vol. 27, no. 3, pp. 639–643, 1979.

[93] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. Washington, DC, USA: U.S. Government Printing Office, 1972.

[94] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, vol. 49, no. 1, pp. 122–136, 1941.

[95] C. G. Esseen, "On the Liapunoff limit of error in the theory of probability," *Arkiv för Matematik, Astronomi Och Fysik*, vol. A28, pp. 1–19, 1942.

[96] I. G. Shevtsova, "An improvement of convergence rate estimates in the Lyapunov theorem," *Doklady Mathematics*, vol. 82, no. 3, pp. 862–864, 2010.

[97] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY: Cambridge University Press, 2013.

[98] G. Wolfer and S. Watanabe, "Information geometry of Markov kernels: a survey," *Frontiers in Physics*, vol. 11-2023, 2023. [Online]. Available: https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2023.1195562

[99] S. Krantz and H. Parks, *A Primer of Real Analytic Functions*. Boston: Birkhäuser, 2002.

[100] P. A. Ferrari, H. Kesten, and S. Martínez, "R-positivity, quasi-stationary distributions and ratio limit theorems for a class of probabilistic automata," *The Annals of Applied Probability*, vol. 6, no. 2, pp. 577–616, 1996. [Online]. Available: https://doi.org/10.1214/aoap/1034968146

[101] A. V. Nagaev, "Limit theorems and testing hypotheses on Markov chains," *Discrete Mathematics and Applications*, vol. 13, no. 6, pp. 569–599, 2003. [Online]. Available: https://doi.org/10.1515/156939203322733282

[102] E. Çınlar, *Probability and Stochastics*. New York, NY: Springer, 2011.

[103] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *International Statistical Review / Revue Internationale de Statistique*, vol. 70, no. 3, pp. 419–435, 2002. [Online]. Available: http://www.jstor.org/stable/1403865

[104] N. Elkayam and M. Feder, "A pairwise-error-probability framework for one-shot information theory," *arXiv:2608.06577v1 [cs.IT]*, 2026.

[105] ——, "One-shot information theory via the pairwise error probability: Lossy, joint source-channel, erasure, and multiuser coding," *arXiv:2608.15169v1 [cs.IT]*, 2026.

[106] ——, "Universal decoding via the pairwise error probability," *arXiv:2609.27887v1 [cs.IT]*, 2026.

[107] T. S. Han, "Hypothesis testing with the general source," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2415–2427, 2000.

[108] E. Bolthausen, "The Berry–Esseen theorem for functionals of discrete Markov chains," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 54, no. 1, pp. 59–73, 1980.