A Spectral Representation Of The Simple Hypothesis Testing Problem

Barış Nakiboğlu
Department of Electrical and Electronics Engineering
Middle East Technical University
Ankara, Türkiye
0000-0001-7737-5423

2026 IEEE International Symposium on Information Theory (ISIT) | 979-8-3315-8004-9/26/$31.00 ©2026 IEEE | DOI: 10.1109/ISIT62367.2026.11653959

***Abstract*—For the simple hypothesis testing problem with randomized detectors, the convex conjugate (i.e., the Legendre transform) of the type II error probability, as a function of the type I error probability, is shown to be the reflection through the origin of the primitive entropy spectrum. By applying the tail expectation formula to the primitive entropy spectrum, the type II error probability is expressed as a Riemann integral of the cumulative distribution function of the likelihood ratio. The derivation extends to tests between σ-finite measures in all non-void cases. The resulting characterization of the type II error implies state-of-the-art bounds on the type II error in terms of the type I error via Berry–Esseen theorem through a brief analysis.**

I. Introduction

The analysis of the following elementary decision problem allows us to illustrate the principal novelty of this work and helps us to delineate our contribution relative to earlier results. Let Y be a zero mean unit variance Gaussian random variable. What is the infimum of the lengths of Borel sets $\mathcal{E}$ satisfying $\mathbf{P}(\mathsf{Y}\notin\mathcal{E})\leq\epsilon$? i.e, what is the value of $\beta(\epsilon)$?

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

where $\mathcal{B}$ is the set of all Borel subset 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 $\phi(\epsilon)$ for $\varphi(\cdot)$ such that $\mathbf{P}(\varphi(\mathsf{Y})\leq\phi(\epsilon))=\epsilon$, then $\beta(\epsilon)$ is the length of the set of all y's satisfying $\varphi(y)>\phi(\epsilon)$. The threshold $\phi(\epsilon)$ is the ϵ-quantile for the random variable $\varphi(\mathsf{Y})$, i.e., $\phi(\epsilon)=F^{-1}_{\varphi(\mathsf{Y})}(\epsilon)$.

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

where $F_{\varphi(\mathsf{Y})}(\cdot)$ 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

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

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

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

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 $\beta(\cdot)$ allow 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

$$\begin{aligned}\beta'(\epsilon)&=\tfrac{-1}{\varphi(Q^{-1}(\epsilon/2))}\\&=\tfrac{-1}{F^{-1}_{\varphi(\mathsf{Y})}(\epsilon)}&&\forall\epsilon\in(0,1).\end{aligned}\tag{5a}$$

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

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

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

$$\begin{aligned}\beta(\epsilon)&=\sup_{\gamma\geq0}\hbar(\gamma)-\gamma\epsilon&&\forall\epsilon\in(0,1),\qquad(5c)\\&=\hbar(\gamma_\epsilon)-\gamma_\epsilon\epsilon&&\forall\epsilon\in(0,1),\qquad(5d)\end{aligned}$$

where $\hbar(\gamma)=\mathbf{E}[\gamma\wedge{}^1/_{\varphi(\mathsf{Y})}]$ and $\gamma_\epsilon={}^1/_{F^{-1}_{\varphi(\mathsf{Y})}(\epsilon)}$. The expectation of any non-negative random variable is 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 the tail expectation formula, see [1, (1.30)]. Thus,

$$\begin{aligned}\hbar(\gamma)&=\int_0^\infty\mathbf{P}\Big(\Big[\gamma\wedge\tfrac{1}{\varphi(\mathsf{Y})}\Big]>\tau\Big)\mathrm{d}\tau&&\forall\gamma\geq0,\\&=\int_0^\gamma\mathbf{P}\Big(\tfrac{1}{\varphi(\mathsf{Y})}>\tau\Big)\mathrm{d}\tau&&\forall\gamma\geq0.\end{aligned}$$

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

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

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 $\beta(\epsilon)$ is the infimum type II error (i.e., missed detection) volume (length) when type I error (i.e., false alarm) probability is less than or equal to ϵ.

Authorized licensed use limited to: ULAKBIM UASL - MIDDLE EAST TECHNICAL UNIVERSITY. Downloaded on August 31,2026 at 11:51:38 UTC from IEEE Xplore. Restrictions apply.

More generally we denote the infimum type II error volume as a function of type I error volume by $\beta(\cdot\|_w^q)$, when the measures associated with the alternative and null hypotheses are q and w, respectively. If randomized detectors are permitted, then we use $\beta(\cdot|_w^q)$. That is for any pair of σ-finite measures (w,q) on the measurable space $(\mathcal{Y},\mathcal{Y})$, we define $\beta(\cdot\|_w^q):[0,\infty)\mapsto[0,\infty]$ and $\beta(\cdot|_w^q):[0,\infty)\mapsto[0,\infty]$ as

$$\beta(\epsilon\|_w^q) := \inf_{\mathsf{T}\in\mathcal{T}^d(\nu):\int \mathsf{T}\mathrm{d}w\le\epsilon} \int (1-\mathsf{T})\mathrm{d}q \qquad \forall \epsilon\ge 0, \tag{6a}$$

$$\beta(\epsilon|_w^q) := \inf_{\mathsf{T}\in\mathcal{T}(\nu):\int \mathsf{T}\mathrm{d}w\le\epsilon} \int (1-\mathsf{T})\mathrm{d}q \qquad \forall \epsilon\ge 0, \tag{6b}$$

where ν is any σ-finite reference measure in which both w and q are absolutely continuous (i.e., for which $w\prec\nu$ and $q\prec\nu$), $\mathcal{T}^d(\nu)$ is the set of all ν-measurable deterministic tests on $\mathcal{Y}$ (i.e., ν-measurable functions of the form $\mathsf{T}:\mathcal{Y}\mapsto\{0,1\}$), and $\mathcal{T}(\nu)$ is the set of all ν-measurable randomized tests on $\mathcal{Y}$ (i.e., ν-measurable functions of the form $\mathsf{T}:\mathcal{Y}\mapsto[0,1]$). Then $\beta(\cdot)=\beta(\cdot\|_w^q)$ for the case when q is the Lebesgues measure and w is the standard Gaussian measure on $\mathbb{R}$ by definition. Furthermore, $\beta(\cdot)=\beta(\cdot|_w^q)$ 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 w and q are probability measures, to the case when w and q are σ-finite measures results in a subtlety: for some σ-finite measures w and q, both $\beta(\epsilon|_w^q)$ and $\beta(\epsilon\|_w^q)$ are infinite for all finite ϵ values. We deem those cases to be the void cases. We will see that non-void cases are those cases satisfying the following condition:

$$q\Big(\tfrac{\mathrm{d}q}{\mathrm{d}\nu}<\tfrac{\mathrm{d}w}{\mathrm{d}\nu}\Big)+w\Big(\tfrac{\mathrm{d}w}{\mathrm{d}\nu}\le\tfrac{\mathrm{d}q}{\mathrm{d}\nu}\Big)<\infty. \tag{7}$$

In all non-void cases, all of the identities in (5) for $\beta(\cdot|_w^q)$, hold with appropriate modifications, see (21). 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 $-1/F_\ell^\wedge\big(\cdot\big|_q^w\big)$ and $-1/F_\ell^\vee\big(\cdot\big|_q^w\big)$, where $F_\ell^\wedge\big(\cdot\big|_q^w\big)$ and $F_\ell^\vee\big(\cdot\big|_q^w\big)$ are the lower semicontinuous and the upper semicontinuous quantiles for the likelihood ratio, see (12) and (15) 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 $F_\ell^\wedge\big(\cdot\big|_q^w\big)$ and $F_\ell^\vee\big(\cdot\big|_q^w\big)$. Whenever $F_\ell\big(\cdot\big|_q^w\big)$ is invertible $F_\ell^{-1}\big(\cdot\big|_q^w\big)=F_\ell^\wedge\big(\cdot\big|_q^w\big)=F_\ell^\vee\big(\cdot\big|_q^w\big)$ 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

$$h(\gamma|_w^q) := \int \Big[\tfrac{\mathrm{d}q}{\mathrm{d}\nu}\wedge\big(\gamma\tfrac{\mathrm{d}w}{\mathrm{d}\nu}\big)\Big]\mathrm{d}\nu \qquad \forall\gamma\in[0,\infty), \tag{8}$$

where ν is any σ-finite reference measure in which both q and w are absolutely continuous, [2, §31]. To obtain the spectral representation of $h(\cdot|_w^q)$, we use the following consequence of integration by parts, see [3, Theorem 2.9.3]: for any measure μ and μ-measurable function f satisfying $\operatorname{ess\,inf}_\mu f\ge 0$

$$\int f\mathrm{d}\mu=\int_0^\infty \mu(f>\tau)\mathrm{d}\tau. \tag{9}$$

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

In §II, we first provide a very brief overview of some of the related prior work, then describe our main contributions. In §III, we define the likelihood ratio, its distribution, and its quantiles formally for arbitrary ordered pairs of σ-finite measures on the same measurable space and review some elementary properties of $h(\cdot|_w^q)$ defined in (8). In §IV, we state Theorem 1 and discuss some of its immediate implications. In §V, we first express $\beta(\cdot|_w^q)$ as a Riemann integral of the distribution function of the log-likelihood ratio, i.e. $F_{\mathsf{L}}\big(\cdot\big|_q^w\big)$ using Theorem 1, see (26). Assuming the Kolmogorov distance of $F_{\mathsf{L}}\big(\cdot\big|_q^w\big)$ to a Gaussian distribution is less than one, we approximate $\ln\beta(\cdot|_w^q)$ using (26), see Theorem 2. This leads to a state-of-the-art non-asymptotic approximation to $\ln\beta(\cdot|_w^q)$ for the memoryless case via the Berry–Esseen theorem, see Proposition 1. In §VI, we discuss the implications of these results and the non-asymptotic spectral approach for information transmission more broadly.

II. Prior Work and Main Contributions

The importance of hypothesis testing problem, and of the $\beta(\cdot|_w^q)$ function in particular, was already clear for information transmission problems by the time of Strassen's seminal work [4]. Furthermore, as pointed out in [4] and employed in [5], the lossless source coding problem provides a concrete operational justification for considering $\beta(\cdot|_w^q)$ and $\beta(\cdot\|_w^q)$ for the case when q is not probability measure. 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 [4], but also at other rates [6]. The hypothesis testing problem is now a key tool for analyzing and understanding information transmission problems, [7, Ch. 1], [8, Ch. 3 and Appendix 5.A], [9, Ch. 5 and 11], [10, Ch. 1 and 4], [11, Ch. 10, 11, and Part III]. The interest in finite-length (i.e., non-asymptotic) analysis and exact asymptotic analysis in information transmissionproblems revived the interest in $\beta(\cdot|_w^q)$ and $\beta(\cdot\|_w^q)$ functions [12]–[17], be it with the notation $\beta(\epsilon|_w^q)=\beta_{1-\epsilon}(w\|q)$.

The upper-semicontinuous likelihood ratio quantile $F_\ell^\vee\big(\cdot\big|_q^w\big)$ and its natural logarithm (i.e., $\ln F_\ell^\vee\big(\cdot\big|_q^w\big)$, the log-likelihood quantile) have been introduced and analyzed by many authors under various names. $F_\ell^\vee\big(\cdot\big|_q^w\big)$ is denoted by μ in [4, (2.5)].

Authorized licensed use limited to: ULAKBIM UASL - MIDDLE EAST TECHNICAL UNIVERSITY. Downloaded on August 31,2026 at 11:51:38 UTC from IEEE Xplore. Restrictions apply.

Inspired by [10, Ch. IV], $\ln F_\ell^\vee \cdot \frac{w}{q}$ is called information spectrum divergence in [17, (2.9)], the classical entropic information spectrum in [18, (2)], the relative entropy information spectrum (and the divergence spectrum) in [19, p. 7044].

For probability measures w and q, the primitive entropy spectrum of $h\big(\cdot\big|\frac{w}{q}\big)$ is related to the primitive divergence spectrum $d\big(\cdot\big|\frac{w}{q}\big)$, commonly used in the spectral representation of f-divergences reported in [20]–[23]:

$$d\big(\gamma\big|\tfrac{w}{q}\big) = 1 \wedge \gamma - h\big(\gamma\big|\tfrac{w}{q}\big) \qquad \forall \gamma \in [0,\infty). \tag{10}$$

The quantity $\frac{1}{1+\gamma} h\big(\gamma\big|\frac{w}{q}\big)$ is called Bayes loss in [21, (66)] and a posteriori loss in [22, (2.25)]. The primitive entropy spectrum $h\big(\cdot\big|\frac{w}{q}\big)$ is denoted by $\psi_{w,q}(\cdot)$ in [23, §III.A].

Motivated by an expression equivalent to [13, (76)] and inspired by f-divergences [24], the divergence $E_\gamma(w\|q)$ was defined in [12, (2.141)] for all probability measures w and q such that w is absolutely continuous in q, i.e., $w \prec q$, see also [25], [26]. One can show for all probability measures w and q satisfying the hypothsis $w \prec q$, that

$$E_\gamma(w\|q) = 1 - h\big(\gamma\big|\tfrac{w}{q}\big) \qquad \forall \gamma \in [0,\infty). \tag{11}$$

The main contributions of this work are the extension of the hypothesis testing problem to σ-finite measures, the identities in (21) and (26), and non-asymptotic approximations for $\ln \beta(\cdot|\frac{q}{w})$. The extension to σ-finite measures yields a unified treatment of closely related problems.

To the best of our knowledge, the identities in (21) and (26) have not previously appeared in this particular form. Nevertheless, they do follow from the Neyman–Pearson lemma by employing elementary properties of convex functions and (9). Furthermore, for probability measures w and q, expressions equivalent to (21a), (21b), and (21c) were reported before with different parametrizations and at times with additional hypotheses, see [12, (2.147) of Theorem 21], [27, Theorem 21], [28, Lemma 1 and Lemma 2], [29, (4.1) and (8.35)], [30]. Irrespective of their prior appearance, the significance of the identities in (21) and (26) lies in the spectral interpretation and the proofs they enable.

Most prior works approximate $\beta(\cdot|\frac{q}{w})$ by first approximating $F_\ell^\vee\big(\cdot\big|\frac{w}{q}\big)$, primarily because $F_\ell^\vee\big(\cdot\big|\frac{w}{q}\big)$ can be readily approximated using existing bounds on $F_\ell\big(\cdot\big|\frac{w}{q}\big)$. The identity (21g) provides a straightforward yet tight method for converting bounds on $F_\ell^\vee\big(\cdot\big|\frac{w}{q}\big)$ into bounds on $\beta(\cdot|\frac{q}{w})$. Theorem 2 can be established using this method by invoking (25) and (27a) to bound $F_\ell^\vee\big(\cdot\big|\frac{w}{q}\big)$. More importantly, one can avoid discussing likelihood ratio quantiles completely, and convert bounds on $F_\ell\big(\cdot\big|\frac{w}{q}\big)$ into bounds on $\beta(\cdot|\frac{q}{w})$ directly by invoking (21d). This is how Theorem 2 will be proved.

For any $\delta \in (0, 1/2)$, Proposition 1 provides a non-asymptotic approximation for $\ln \beta(\cdot|\frac{q}{w})$ on $(\delta, 1-\delta)$ with a uniform $O(1)$ error for the memoryless case via (27d). This is qualitatively equivalent to the corresponding results in [4], [12], [31]. For the memoryless case with γ, σ, and δ values given in Proposition 1, (27b) and (27c) were previously reported for the case when q is a probability measure in [28, (9) and (13)], under the assumption that $\mathcal{Y}$ is countable. We became aware of [28], after the initial submission of this work to ISIT 2026.

III. The Likelihood Ratio, Its Distribution, Its Quantiles, and The Primitive Entropy Spectrum

Recall that the Lebesgue decomposition theorem holds not only for finite measures but also for σ-finite measures, see [32, Theorem 5.5.3] and [2, §32, Theorem C]. As a consequence, Radon–Nikodym theorem for finite measures generalizes to the σ-finite measures, see [3, §3.2], [2, §31, Theorem B]. Thus for any two σ-finite measures w and ν satisfying $w \prec \nu$, there exists 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 1. For any pair of σ-finite measures (w, q) on the same measurable space, the likelihood ratio $\ell_{w,q}(\cdot)$ and the (cumulative) distribution function $F_\ell\big(\cdot\big|\frac{w}{q}\big)$ are defined as

$$\ell_{w,q} := \begin{cases} \infty & \text{if } \frac{\mathrm{d}q}{\mathrm{d}\nu} = 0, \\ \frac{\mathrm{d}w_\sim}{\mathrm{d}q} & \text{if } \frac{\mathrm{d}q}{\mathrm{d}\nu} > 0, \end{cases} \tag{12a}$$

$$F_\ell\big(\tau\big|\tfrac{w}{q}\big) := \int \mathbb{1}_{\{\ell_{w,q} \le \tau\}} \mathrm{d}w \qquad \forall \tau \in \mathbb{R}, \tag{12b}$$

where ν is any σ-finite reference measure satisfying $w \prec \nu$ and $q \prec \nu$, and $w_\sim$ is the q-absolutely continuous component of w.

Note that $\ell_{w,q} = \widetilde{\ell_{w,q}}$ holds $(w+q)$-a.e., for any likelihood ratio $\widetilde{\ell_{w,q}}$ associated with another reference measure $\widetilde{\nu}$ satisfying $w \prec \widetilde{\nu}$ and $q \prec \widetilde{\nu}$. Thus $F_\ell\big(\cdot\big|\frac{w}{q}\big)$ does not depend on the choice of ν. Furthermore, with the understanding that $\frac{1}{0}$ and $\frac{1}{\infty}$ stand for ∞ and 0, respectively, we have

$$\ell_{q,w} = \tfrac{1}{\ell_{w,q}} \qquad (w+q)\text{-a.e.}. \tag{13}$$

If w is a probability measure, then $F_\ell\big(\cdot\big|\frac{w}{q}\big)$ is the cumulative distribution function of 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 distribution function $F_\ell\big(\cdot\big|\frac{w}{q}\big)$ depends only on the absolutely continuous components of w and q in each other, as an immediate consequence of Definition 1:

$$F_\ell\big(\cdot\big|\tfrac{w}{q}\big) = F_\ell\big(\cdot\big|\tfrac{w_\sim}{q_\sim}\big) \tag{14}$$

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

Definition 2. For any pair of σ-finite measures (w, q) on the same measurable space, the likelihood ratio quantiles (LRQs) $F_\ell^\wedge\big(\cdot\big|\frac{w}{q}\big) : \mathbb{R} \mapsto \overline{\mathbb{R}}$ and $F_\ell^\vee\big(\cdot\big|\frac{w}{q}\big) : \mathbb{R} \mapsto \overline{\mathbb{R}}$ are defined as

$$F_\ell^\wedge\big(\epsilon\big|\tfrac{w}{q}\big) := \inf\big\{\gamma \in \mathbb{R} : F_\ell\big(\gamma\big|\tfrac{w}{q}\big) \ge \epsilon\big\} \qquad \forall \epsilon \in \mathbb{R}, \tag{15a}$$

$$F_\ell^\vee\big(\epsilon\big|\tfrac{w}{q}\big) := \inf\big\{\gamma \in \mathbb{R} : F_\ell\big(\gamma\big|\tfrac{w}{q}\big) > \epsilon\big\} \qquad \forall \epsilon \in \mathbb{R}. \tag{15b}$$

As a result of (14), we have

$$F_\ell^\vee\big(\cdot\big|\tfrac{w}{q}\big) = F_\ell^\vee\big(\cdot\big|\tfrac{w_\sim}{q_\sim}\big) \quad \text{and} \quad F_\ell^\wedge\big(\cdot\big|\tfrac{w}{q}\big) = F_\ell^\wedge\big(\cdot\big|\tfrac{w_\sim}{q_\sim}\big). \tag{16}$$

The LRQs are the semicontinuous inverses of $F_\ell\big(\cdot\big|\frac{w}{q}\big)$, this is why we imitate the notation in [33], [34] and denote them with $F_\ell^\wedge\big(\cdot\big|\frac{w}{q}\big)$ and $F_\ell^\vee\big(\cdot\big|\frac{w}{q}\big)$. A brief description of semicontinuous inverses of monotonic functions and properties of likelihood ratio quantiles that are relevant for our purposes can be found in [35, §2].

Authorized licensed use limited to: ULAKBIM UASL - MIDDLE EAST TECHNICAL UNIVERSITY. Downloaded on August 31,2026 at 11:51:38 UTC from IEEE Xplore. Restrictions apply.

For σ-finite measures w and q on the same measurable space, $h(\cdot|_w^q):[0,\infty)\mapsto[0,\infty]$, defined in (8), satisfies

$$h(0|_w^q)=0, \tag{17a}$$
$$h(\gamma|_w^q)=h\big(\gamma\big|_{w_\sim}^{q_\sim}\big) \qquad \forall\gamma\in(0,\infty), \tag{17b}$$
$$h(\gamma|_w^q)=q_\sim\Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w}\le\gamma\Big)+\gamma w\Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w}>\gamma\Big) \qquad \forall\gamma\in(0,\infty), \tag{17c}$$
$$h(\gamma|_w^q)=\gamma h\big(1/\gamma\big|_q^w\big) \qquad \forall\gamma\in(0,\infty), \tag{17d}$$

Note that $h(\gamma|_w^q)$ is non-decreasing in γ. Thus (17d) implies

$$(1\vee\gamma)h(1|_w^q)\ge h(\gamma|_w^q)\ge(1\wedge\gamma)h(1|_w^q) \quad \forall\gamma\in(0,\infty). \tag{18}$$

Thus for any pair of σ-finite measures (w,q) and $\forall\gamma\in(0,\infty)$,

$$\text{“}h\big(1\big|_q^w\big)\in\mathcal{A}_\jmath \quad \text{iff} \quad h\big(\gamma\big|_q^w\big)\in\mathcal{A}_\jmath\text{”} \qquad \forall\jmath\in\{1,2,3\}, \tag{19}$$

where $\mathcal{A}_1=\{0\}$, $\mathcal{A}_2=(0,\infty)$, and $\mathcal{A}_3=\{\infty\}$. We call the case when $h(\gamma|_w^q)=\infty$ for all $\gamma\in\mathbb{R}_+$ to be the void case. As a result of (19), the condition in (7) is a necessary and sufficient condition for the non-void cases for $h(\cdot|_w^q)$, as well. A more detail discussion of primitive entropy spectrum and its properties can be found in [35, §2].

IV. A Spectral Neyman–Pearson Lemma

Note that (6) and (8) imply $h(1|_w^q)\le\beta(\epsilon|_w^q)+\epsilon$ for all $\epsilon\ge0$. Thus whenever the condition in (7) is violated, i.e., whenever $h(1|_w^q)=\infty$, we are in the void case for $\beta(\cdot|_w^q)$, as well. When (7) holds, i.e., $h(1|_w^q)<\infty$, the function $\beta(\cdot|_w^q)$ is finite on $(0,\infty)$, see Theorem 1. Similar to what we have observed for $F_\ell\big(\cdot\big|_q^w\big)$ in (14), LRQs in (16), and $h(\cdot|_w^q)$ in (17b), we have

$$\beta(\cdot|_w^q)=\beta\big(\cdot\big|_{w_\sim}^{q_\sim}\big), \qquad \text{and} \qquad \beta(\cdot\|_w^q)=\beta\big(\cdot\big\|_{w_\sim}^{q_\sim}\big). \tag{20}$$

Consequently, $\beta(\epsilon\|_w^q)=\beta(\epsilon|_w^q)=0$ for all $\epsilon\ge\|w_\sim\|_1$. The behavior of the function $\beta(\cdot|_w^q)$ is characterized in terms of $h(\cdot|_w^q)$, $F_\ell^\wedge\big(\cdot\big|_q^w\big)$, and $F_\ell^\vee\big(\cdot\big|_q^w\big)$ by Theorem 1 whose proof can be found in [35, §3]. Theorem 1 is nothing but the Neyman–Pearson lemma strengthened with various spectral relations and their implications.

Theorem 1. *If (w,q) is a pair of σ-finite measures satisfying (7), then $\beta(\cdot|_w^q)$ is a non-increasing, convex, and continuous function satisfying*

$$\beta(\epsilon|_w^q)=\sup_{\gamma\ge0}h(\gamma|_w^q)-\gamma\epsilon \qquad \forall\epsilon\ge0, \tag{21a}$$
$$=h(\gamma_\epsilon|_w^q)-\gamma_\epsilon\epsilon \qquad \forall\epsilon>0, \tag{21b}$$
$$=\int_0^{\gamma_\epsilon}\Big[w\Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w}>\tau\Big)-\epsilon\Big]\mathrm{d}\tau \qquad \forall\epsilon\ge0, \tag{21c}$$
$$=\int_0^{\infty}\Big|w\Big(\tfrac{\mathrm{d}q_\sim}{\mathrm{d}w}>\tau\Big)-\epsilon\Big|^+\mathrm{d}\tau \qquad \forall\epsilon\ge0, \tag{21d}$$
$$\mathrm{D}_-\beta(\epsilon|_w^q)=\tfrac{-1}{F_\ell^\wedge(\epsilon|_q^w)} \qquad \forall\epsilon>0, \tag{21e}$$
$$\mathrm{D}_+\beta(\epsilon|_w^q)=\tfrac{-1}{F_\ell^\vee(\epsilon|_q^w)} \qquad \forall\epsilon\ge0, \tag{21f}$$

for any $\gamma_\epsilon\in\Big[\frac{1}{F_\ell^\vee(\epsilon|_q^w)},\frac{1}{F_\ell^\wedge(\epsilon|_q^w)\vee0}\Big]$. Furthermore, the restriction of $\beta(\cdot|_w^q)$ to $(0,\|w_\sim\|_1)$ is an invertible function with the image $(0,\|q_\sim\|_1)$ and there exists a $\mathsf{T}_\epsilon\in\mathcal{T}(\nu)$ satisfying both $\int\mathsf{T}_\epsilon\mathrm{d}w=\epsilon$ and $\int(1-\mathsf{T}_\epsilon)\mathrm{d}q=\beta(\epsilon|_w^q)$ for all $\epsilon\in[0,\|w_\sim\|_1]$.

$$\beta(\epsilon|_w^q)=\int_\epsilon^{\|w_\sim\|_1}g(\tau)\mathrm{d}\tau \qquad \forall\epsilon\in[0,\|w_\sim\|_1], \tag{21g}$$
$$\partial\beta(\epsilon|_w^q)\subset(-\infty,0) \qquad \forall\epsilon\in(0,\|w_\sim\|_1), \tag{21h}$$

for any $g:(0,\|w_\sim\|_1)\to(0,\infty)$ s.t. $\frac{1}{F_\ell^\vee(\cdot|_q^w)}\le g(\cdot)\le\frac{1}{F_\ell^\wedge(\cdot|_q^w)}$. If $q_\sim$ is a finite measure, i.e., $\|q_\sim\|_1<\infty$, then

$$\beta(\epsilon|_w^q)=\|q_\sim\|_1-\int_0^\epsilon g(\tau)\mathrm{d}\tau \quad \forall\epsilon\ge0. \tag{21j}$$

Remark 1. In (21c) and (21d) the ‘$>$’ signs can be replaced by ‘$\ge$’ signs because monotonic functions have at most countably many discontinuities by [38, Theorem 4.30] and any countable set has zero Lebesgue measure. Hence (12) and (13) imply

$$\beta(\epsilon|_w^q)=\int_0^{\gamma_\epsilon}\big[F_\ell\big(\tfrac{1}{\tau}\big|_q^w\big)-\epsilon\big]\mathrm{d}\tau \qquad \forall\epsilon\ge0, \tag{22a}$$
$$=\int_0^{\infty}\big|F_\ell\big(\tfrac{1}{\tau}\big|_q^w\big)-\epsilon\big|^+\mathrm{d}\tau \qquad \forall\epsilon\ge0. \tag{22b}$$

Furthermore, the convex conjugate $\beta^*(\cdot|_w^q)$ of $\beta(\cdot|_w^q)$ can be determined using (21e) and (21f) to be

$$\beta^*(\lambda|_w^q):=\sup_{x\ge0}\lambda x-\beta(x|_w^q) \qquad \forall\lambda\in\mathbb{R}, \tag{23a}$$
$$=\begin{cases}-h(-\lambda|_w^q) & \text{if } \lambda\le0,\\ \infty & \text{if } \lambda>0.\end{cases} \tag{23b}$$

V. Approximation for $\beta(\cdot|_w^q)$

To employ classical limit theorems and finite-sample bounds for sums of random variables in approximating $\beta(\cdot|_w^q)$, we need to express (22b), i.e., (21d), in terms of the distribution function $F_{\mathsf{L}}\big(\cdot\big|_q^w\big)$ of the log-likelihood-ratio.

$$\mathsf{L}_{w,q}:=\ln\ell_{w,q}, \tag{24a}$$
$$F_{\mathsf{L}}\big(\tau\big|_q^w\big):=\int\mathbb{1}_{\{\mathsf{L}_{w,q}\le\tau\}}\mathrm{d}w \qquad \forall\tau\in\mathbb{R}. \tag{24b}$$

Then for any pair of σ-finite measure w and q we have

$$F_{\mathsf{L}}\big(\tau\big|_q^w\big)=F_\ell\big(\exp(\tau)\big|_q^w\big) \qquad \forall\tau\in\mathbb{R}. \tag{25}$$

If w and q also satisfy (7), applying the change of variable $\tau=e^{-s}$ in the integral in (22b) and invoking (25), we get

$$\beta(\epsilon|_w^q)=\int_{-\infty}^{\infty}e^{-s}\big|F_{\mathsf{L}}\big(s\big|_q^w\big)-\epsilon\big|^+\mathrm{d}s \qquad \forall\epsilon\ge0. \tag{26}$$

Theorem 2. *If w is a probability measure and q is a σ-finite measure satisfying*

$$\big|F_{\mathsf{L}}\big(\gamma-\sigma x\big|_q^w\big)-Q(x)\big|\le\delta \qquad \forall x\in\mathbb{R}, \tag{27a}$$

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

$$\beta(\epsilon|_w^q)\le e^{-\gamma}\tfrac{Q\left(\sigma-Q^{-1}(\epsilon-\delta)\right)}{\varphi(\sigma)\sqrt{2\pi}} \qquad \forall\epsilon\in(\delta,1), \tag{27b}$$
$$\beta(\epsilon|_w^q)\ge e^{-\gamma}\tfrac{Q\left(\sigma-Q^{-1}(\epsilon+\delta)\right)}{\varphi(\sigma)\sqrt{2\pi}} \qquad \forall\epsilon\in(0,1-\delta), \tag{27c}$$
$$\beta(\epsilon|_w^q)=\tfrac{\Psi(\sigma-\tau)}{\sqrt{2\pi}}e^{-\gamma+\sigma\tau-\frac{\tau^2}{2}+\Delta} \quad \forall\epsilon\in(\delta,1-\delta-Q(\sigma)), \tag{27d}$$

where $\tau=Q^{-1}(\epsilon)$, $\Psi(x)=\frac{Q(x)}{\varphi(x)}$, and $|\Delta|\le\frac{\delta(\sigma+0.5)\sqrt{2\pi}}{(1-\delta-\epsilon)\wedge(\epsilon-\delta)}$.

Authorized licensed use limited to: ULAKBIM UASL - MIDDLE EAST TECHNICAL UNIVERSITY. Downloaded on August 31,2026 at 11:51:38 UTC from IEEE Xplore. Restrictions apply.

Proof of Theorem 2.

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

where (a) follows from (26) for $s=\gamma-\sigma x$, (b) follows from the hypothesis (27a), (c) follows from the fact that $Q(\cdot)$ is an invertable and decreasing function. Then (27b) follows from (28) established below for $\lambda=Q^{-1}(\epsilon-\delta)$.

$$\begin{aligned}\int_{-\infty}^{\lambda} e^{\sigma x}(Q(x)-Q(\lambda))\mathrm{d}x &= \int_{-\infty}^{\lambda} e^{\sigma x}\int_x^{\lambda}\varphi(\tau)\mathrm{d}\tau\mathrm{d}x,\\ &= \int_{-\infty}^{\lambda}\varphi(\tau)\int_{-\infty}^{\tau} e^{\sigma x}\mathrm{d}x\mathrm{d}\tau,\\ &= \int_{-\infty}^{\lambda}\tfrac{\varphi(\tau)e^{\sigma\tau}}{\sigma}\mathrm{d}\tau,\\ &= e^{\frac{\sigma^2}{2}}\int_{-\infty}^{\lambda}\tfrac{\varphi(\tau-\sigma)}{\sigma}\mathrm{d}\tau,\\ &= \tfrac{Q(\sigma-\lambda)}{\sigma\varphi(\sigma)\sqrt{2\pi}}.\end{aligned}\tag{28}$$

(27c) is established using the lower bound on $F_{\mathsf{L}}\big(\gamma-\sigma x\big|_q^w\big)$ in (27a) by following a similar analysis.

To establish (27d), first note that as a result of (27b) and (27c), there exists a $\phi\in(\epsilon-\delta,\epsilon+\delta)$ such that

$$\beta(\epsilon|_w^q)=e^{-D(w\|q)+g(\phi)}\quad\text{where}\quad g(\phi):=\ln\tfrac{Q\left(\sigma-Q^{-1}(\phi)\right)}{\varphi(\sigma)\sqrt{2\pi}}.$$

Since $g(\cdot)$ is continuously differentiable, as a result of Taylor's theorem there exists ϕ_0 between ϕ and ϵ such that

$$g(\phi)=g(\epsilon)+(\phi-\epsilon)g'(\phi_0).\tag{29}$$

On the other hand $g'(\phi)=\frac{-\varphi\left(\sigma-Q^{-1}(\phi)\right)}{Q(\sigma-Q^{-1}(\phi))}\frac{1}{\varphi(Q^{-1}(\phi))}$, and

$$\begin{aligned}|g'(\phi_0)| &\overset{(a)}{\le} \tfrac{1}{\Psi(\sigma-Q^{-1}(\phi_0))}\tfrac{\sqrt{\pi/2}}{(1-\phi_0)\wedge\phi_0},\\ &\overset{(b)}{\le} \tfrac{1}{\Psi(\sigma-Q^{-1}(\epsilon+\delta))}\tfrac{\sqrt{\pi/2}}{(1-\phi_0)\wedge\phi_0},\\ &\overset{(c)}{\le} \tfrac{1}{\Psi(2\sigma)}\tfrac{\sqrt{\pi/2}}{(1-\phi_0)\wedge\phi_0},\\ &\overset{(d)}{\le} (\sigma+\sqrt{\sigma^2+1})\tfrac{\sqrt{\pi/2}}{(1-\phi_0)\wedge\phi_0},\\ &\overset{(e)}{\le} (\sigma+0.5)\tfrac{\sqrt{2\pi}}{(1-\phi_0)\wedge\phi_0},\end{aligned}\tag{30}$$

where (a) holds because $Q(\tau)\le\varphi(\tau)\sqrt{\pi/2}$ for all $\tau>0$, by [39, 7.1.13], (b) holds because $\Psi(\cdot)$ and $Q^{-1}(\cdot)$ are both decreasing functions and $\phi_0\in(\epsilon-\delta,\epsilon+\delta)$, (c) holds because $\Psi(\cdot)$ and $Q^{-1}(\cdot)$ are both decreasing functions, $\epsilon+\delta\le 1-Q(\sigma)$ by hypothesis, and $Q^{-1}(1-\varepsilon)=-Q^{-1}(\varepsilon)$ for all $\varepsilon\in(0,1)$, (d) follows from [39, 7.1.13], (e) follows from $\sqrt{\sigma^2+1}\le\sigma+1$. Thus (27d) follows from (29), (30), and $\phi_0\in(\epsilon-\delta,\epsilon+\delta)$. □

Proposition 1. *Let w_t be a probability measure, and q_t be a σ-finite measure satisfying $|D(w_t\|q_t)|<\infty$ for all $t\in\{1,\dots,n\}$ for an $n\in\mathbb{Z}_+$, w be $w=\otimes_{t=1}^n w_t$, q be $q=\otimes_{t=1}^n q_t$, and T be*

$$T=\sum\nolimits_{t=1}^{n}\int\left|\ln\tfrac{\mathrm{d}w_t}{\mathrm{d}q_t}-D(w_t\|q_t)\right|^3\mathrm{d}w_t.$$

If $\int(\mathsf{L}_{w,q}-D(w\|q))^2\mathrm{d}w>0$, i.e., the variance of $\mathsf{L}_{w,q}$, which is defined in (24a)*, under w is positive, then* (27b)*,* (27c)*, and* (27d) *hold for $\gamma=D(w\|q)$ and $\sigma=\sqrt{\int(\mathsf{L}_{w,q}-D(w\|q))^2\mathrm{d}w}$, and $\delta=0.56^T/\sigma^3$ whenever $\delta\in[0,1)$.*

Proof of Proposition 1. (27a) holds as a result of Berry-Esseen theorem [40]–[42] because $\mathsf{L}_{w,q}=\sum_{t=1}^n\ln\frac{\mathrm{d}w_t}{\mathrm{d}q_t}$. We calculate the first moment and variance of $\mathsf{L}_{w,q}$ using Fubini's theorem [3, Theorem 3.4.4], giving us the expression for γ and σ^2. Then proposition follows from Theorem 2. □

VI. Discussion

One of the primary motivations for considering σ-finite measures is the fixed-length ϵ-lossless source coding problem which aims to characterize $\beta\big(\epsilon\big\|_w^\sharp\big)$ for the counting measure $\sharp$. Since $\beta\big(\epsilon\big\|_w^\sharp\big)=\lceil\beta\big(\epsilon\big|_w^\sharp\big)\rceil$ for all non-negative ϵ and probability mass functions w (see [36, Remark 1], [37]), Theorem 2 can be invoked to approximate $\beta\big(\epsilon\big\|_w^\sharp\big)$. Specifically, by applying a non-universal (Markov chain specific) Berry-Esseen theorem [31, Theorem 21], Theorem 2 improves[1] the best known upper bound [31, Theorem 19] and recovers the best known lower bound [31, Theorem 20] with better constants for time-homogeneous, irreducible, aperiodic Markov sources, see [36, Proposition 3], [37]. Furthermore, substituting [31, Theorem 21] with Nagaev's result [43] yields similar approximations for more general abstract time-homogeneous Markov measures w and q, see [37].

The characterization in (26) facilitates the application of Watson's Lemma alongside classical limit theorems for sums of random variables to derive sharp approximations for $\beta(\cdot|_w^q)$, see [37]. Analogous to the proof of Theorem 2, this approach decouples the approximation of $\beta(\cdot|_w^q)$ from hypothesis about the underlying structure of w and q.

The characterization $\beta(\cdot|_w^q)$ in Theorem 1 enables exact measure change identities for β functions, see [35, Lemma 7], [36, Theorem 2], [37]. For the memoryless case, when invoked together with Proposition 1 the measure change identities recovers non-asymptotic approximations for $\beta\big(e^{-D(v_\alpha\|q)+b_n}\big\|_q^w\big)$ under the hypothesis $|b_n|=O(\sqrt{n})$, that are as tight as the state-of-the-art non-asymptotic bounds in [44]–[48], with much less effort, see [35, Theorem 3] [36, Proposition 2], [37]. The analysis extends to the Markovian case [37], obtaining such results for the first time.

Non-asymptotic spectral representations similar to Theorem 1 for other information transmission problems are likely to lead to not only new and sharper approximations, but also to new connections such as the ones between $\beta(\cdot|_w^q)$ and $h(\cdot|_w^q)$ established in Theorem 1.

[1] [31] points out that [31, Theorems 19] can be improved but only by "extending the involved proof of" [31, Theorems 17].

Authorized licensed use limited to: ULAKBIM UASL - MIDDLE EAST TECHNICAL UNIVERSITY. Downloaded on August 31,2026 at 11:51:38 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

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

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

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

[4] 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).

[5] 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).

[6] 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.

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

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

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

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

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

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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

[18] 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.

[19] 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.

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

[21] 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.

[22] 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

[23] 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.

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

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

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

[27] 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

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

[29] 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

[30] ——, "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

[31] 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.

[32] R. M. Dudley, *Real analysis and probability*. New York, NY: Cambridge University Press, 2002, vol. 74.

[33] 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

[34] 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

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

[36] ——, "A spectral representation of the simple hypothesis testing and refined approximations," in *14th Iran Workshop on Communication and Information Theory (submitted to)*, 2026.

[37] ——, "A spectral representation of the simple hypothesis testing problem," *arXiv:2601.09564v2 [cs.IT]*, 2026.

[38] W. Rudin, *Principles of Mathematical Analysis*. New York, NY: McGraw-Hill, 1976.

[39] 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.

[40] 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.

[41] 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.

[42] I. G. Shevtsova, "An improvement of convergence rate estimates in the Lyapunov theorem," *Doklady Mathematics*, vol. 82, no. 3, pp. 862–864, 2010.

[43] S. V. Nagaev, "More exact statement of limit theorems for homogeneous markov chains," *Theory of Probability & Its Applications*, vol. 6, no. 1, pp. 62–81, 1961.

[44] B. Nakiboğlu, "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.

[45] ——, "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]).

[46] 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]).

[47] A. Theocharous and I. Kontoyiannis, "Lossless data compression at pragmatic rates," in *2025 IEEE International Symposium on Information Theory (ISIT)*, 2025, pp. 1–6.

[48] 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]).

Authorized licensed use limited to: ULAKBIM UASL - MIDDLE EAST TECHNICAL UNIVERSITY. Downloaded on August 31,2026 at 11:51:38 UTC from IEEE Xplore. Restrictions apply.