



Article

Weighted Chernoff Information and Optimal Loss Exponent in Context-Sensitive Hypothesis Testing

Mark Kelbert ^{1,2,*}  and El'mira Yu. Kalimulina ^{3,4} 

¹ Laboratory of Stochastic Analysis and Its Applications, Department of Statistics and Data Analysis, National Research University Higher School of Economics, 101000 Moscow, Russia

² Department of Mathematics, Swansea University, Swansea SA2 8PP, UK

³ Institute for Information Transmission Problems, Russian Academy of Sciences (IITP RAS), 127051 Moscow, Russia; eyk@iitp.ru

⁴ Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, 119991 Moscow, Russia

* Correspondence: mkelbert@hse.ru

Abstract

We study binary hypothesis testing for i.i.d. observations under a multiplicative context weight. For the optimal weighted total loss, defined as the sum of weighted type-I and type-II losses, we prove the logarithmic asymptotic $L_n^* = \exp\{-nD_C^w(\mathbb{P}, \mathbb{Q}) + o(n)\}$, $n \rightarrow \infty$, where D_C^w is the weighted Chernoff information. The single-letter form of the exponent relies on a structural assumption that the weight factorises across observations, $\varphi(x_1^n) = \prod_{i=1}^n \varphi(x_i)$; this restriction is essential for the single-letter representation and should be distinguished from the weaker qualitative description “multiplicative context weight”. The proof embeds the weighted geometric mixtures $\varphi p^\alpha q^{1-\alpha}$ into a likelihood-ratio exponential family and identifies the rate through its log-normaliser. We also derive concentration bounds for the tilted weighted log-likelihood, obtain closed forms for Gaussian, Poisson, and exponential models, and extend the exponent characterisation to finitely many hypotheses.

Keywords: hypothesis testing; weighted Chernoff information; weighted Bhattacharyya coefficient; exponential family; information geometry; context-sensitive loss

MSC: 62F03; 60F10

1. Introduction

Let \mathcal{X} be a Polish space with its Borel σ -algebra and let $X_1^n = (X_1, \dots, X_n)$ be i.i.d. \mathcal{X} -valued observations. We consider the simple hypotheses

$$H_0 : X_1^n \sim \mathbb{P}^{\otimes n} \quad \text{versus} \quad H_1 : X_1^n \sim \mathbb{Q}^{\otimes n},$$

where \mathbb{P} and \mathbb{Q} are probability measures on \mathcal{X} dominated by a reference measure μ . Without loss of generality, one may take $\mu = \frac{1}{2}(\mathbb{P} + \mathbb{Q})$ and write $p = \frac{d\mathbb{P}}{d\mu}$ and $q = \frac{d\mathbb{Q}}{d\mu}$. In the unweighted setting, the optimal sum of type-I and type-II error probabilities is characterized by $\text{TV}(\mathbb{P}^{\otimes n}, \mathbb{Q}^{\otimes n})$ and can be written as

$$\int_{\mathcal{X}^n} \min\{p(x_1^n), q(x_1^n)\} d\mu^{\otimes n}(x_1^n), \quad p(x_1^n) = \prod_{i=1}^n p(x_i), \quad q(x_1^n) = \prod_{i=1}^n q(x_i). \quad (1)$$



Academic Editor: Nikolai Leonenko

Received: 20 March 2026

Revised: 30 April 2026

Accepted: 2 May 2026

Published: 8 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

In the standard (unweighted) Bayesian setting, the decay rate of the optimal total error probability is governed by the Chernoff information [1,2]:

$$\begin{aligned} \rho_\alpha(p, q) &:= \int_{\mathcal{X}} p(x)^\alpha q(x)^{1-\alpha} d\mu(x), \quad \alpha \in [0, 1], \\ \rho(p, q) &:= \inf_{\alpha \in [0, 1]} \rho_\alpha(p, q), \\ D_C(\mathbb{P}, \mathbb{Q}) &:= -\ln \rho(p, q) = \max_{\alpha \in [0, 1]} [-\ln \rho_\alpha(p, q)]. \end{aligned} \tag{2}$$

Here ρ_α is usually called the α -skewed Bhattacharyya affinity coefficient, and $\rho(p, q) = \inf_{\alpha \in [0, 1]} \rho_\alpha(p, q)$ is the affinity coefficient. In view of Hölder’s inequality, $\rho_\alpha(p, q) \in [0, 1]$.

Chernoff also introduced an asymptotic efficiency notion for comparing two experimental designs $e = \frac{\ln \rho_1}{\ln \rho_2}$ such that n observations on one test are equivalent (i.e., they give asymptotically the same total loss as $n \rightarrow \infty$) to en observations on another test; see [1].

The paper studies a context-sensitive (weighted) analogue of this criterion and the logarithmic asymptotics of the optimal total loss as $n \rightarrow \infty$, in the framework of [3,4]. In the weighted setting, a nonnegative weight function $\varphi(x_1^n)$ reweights the loss of a wrong decision according to the realised sample. Thus, φ acts as a context factor that changes the relevance of different observations for the statistical task.

Weights of this form arise naturally whenever observations are not equally informative for the inference task. Two canonical mechanisms produce such φ . In *importance-type reweighting*, samples drawn under a proposal density g are used to perform inference with respect to a target h , and the Radon–Nikodym factor $\varphi(x) = h(x)/g(x)$ enters the loss as a strictly positive (non-indicator) tilt; this is the mechanism underlying the context-sensitive framework of [3,4].

In applications, the informational value of an observation often depends on the underlying channel state. A canonical example, directly relevant to multiple hypothesis testing of transmission regimes, is a mobile communication channel modulated by a multi-zone coverage process (e.g., strong / weak / outage) along the receiver trajectory: samples acquired in outage carry little information about the regime and are weighted accordingly. Such reliability-weighted aggregation in multi-state channels was studied within multi-valued frameworks in [5,6].

Under the standard assumption that the modulating state at time i is determined by X_i alone, the resulting weight is a strictly positive bounded function $\varphi(x)$ and extends multiplicatively to X_1^n . The weighted Chernoff information $D_C^w(\mathbb{P}, \mathbb{Q})$ then quantifies the effective discrimination rate under channel-dependent reliability and reduces to the classical rate $D_C(\mathbb{P}, \mathbb{Q})$ in the limit $\varphi \equiv 1$. Further parametric instances (Gaussian, Poisson, exponential) are worked out in Section 4.

Throughout we assume that the weight is compatible with the i.i.d. structure and factorises across observations; by abuse of notation, φ denotes both the one-step weight and its product extension.

Assumption 1 (Factorised weight). *The weight function $\varphi(x_1^n)$ satisfies*

$$\varphi(x_1^n) = \prod_{i=1}^n \varphi(x_i), \quad \varphi \geq 0. \tag{3}$$

Assumption 1 is the key single-letter hypothesis. It yields the weighted affinities

$$\rho_\alpha^w(p, q) = \int_{\mathcal{X}} \varphi(x) p(x)^\alpha q(x)^{1-\alpha} d\mu(x),$$

hence an additive logarithmic rate. For one observation and equal priors, the weighted Bayes risk equals

$$\frac{1}{2} \int_{\mathcal{X}} \varphi(x) \min\{p(x), q(x)\} d\mu(x).$$

Since $\min\{a, b\} \leq a^\alpha b^{1-\alpha}$ for every $\alpha \in [0, 1]$,

$$\int_{\mathcal{X}} \varphi(x) \min\{p(x), q(x)\} d\mu(x) \leq \rho_\alpha^w(p, q),$$

and therefore

$$\int_{\mathcal{X}} \varphi(x) \min\{p(x), q(x)\} d\mu(x) \leq \exp\{-D_C^w(\mathbb{P}, \mathbb{Q})\},$$

where $D_C^w(\mathbb{P}, \mathbb{Q}) := \max_{\alpha \in [0,1]} [-\ln \rho_\alpha^w(p, q)]$ (see Definition 2). Under Assumption 1, the same bound factorises over n observations and yields the exponential scale $\exp\{-nD_C^w(\mathbb{P}, \mathbb{Q})\}$. Theorem 1 shows that this scale is exact on the logarithmic level.

1.1. Main Result and Contributions

Let L_n^* denote the optimal total context-sensitive loss (sum of weighted type-I and type-II losses, minimised over decision rules) for n i.i.d. observations under Assumption 1. Our main theorem (Theorem 1) proves the single-letter logarithmic asymptotic

$$L_n^* = \exp\{-nD_C^w(\mathbb{P}, \mathbb{Q}) + o(n)\}, \quad n \rightarrow \infty, \tag{4}$$

where the rate is the *weighted Chernoff information*

$$D_C^w(\mathbb{P}, \mathbb{Q}) = \max_{\alpha \in [0,1]} \left[-\ln \int_{\mathcal{X}} \varphi(x) p(x)^\alpha q(x)^{1-\alpha} d\mu(x) \right]. \tag{5}$$

For $\varphi \equiv 1$, (5) reduces to the classical Chernoff information.

We also extend the exponent characterisation to a finite family of simple hypotheses: the optimal M -ary rate is the minimum pairwise weighted Chernoff information (cf. [7] in the unweighted case). A central technical device is an exponential-family representation of the weighted geometric mixtures $\alpha \mapsto \varphi p^\alpha q^{1-\alpha}$. This embeds the mixtures into a likelihood-ratio exponential family and identifies the exponent through the corresponding log-normaliser. We further derive concentration bounds for tilted weighted log-likelihood ratios and closed-form expressions for D_C^w in several parametric models; see Section 4.

1.2. Contributions

Items (N1)–(N4) below indicate new results; items (A1)–(A3) summarise definitions, geometric context, and tools adopted from the existing literature.

- (N1) (*New.*) Theorem 1 establishes the logarithmic asymptotic (4) for the optimal weighted total loss under the factorised weight of Assumption 1, with rate given by the weighted Chernoff information (5).
- (N2) (*New.*) The exponential-family representation of the weighted geometric mixtures $\alpha \mapsto \varphi p^\alpha q^{1-\alpha}$ (Section 3.2) and the resulting uniqueness of the optimal skewing parameter α^* .
- (N3) (*New.*) Concentration bounds for the tilted weighted log-likelihood and the finite- n tail bound of Theorem 2 (Section 3.4).
- (N4) (*New.*) Closed-form expressions for D_C^w in the Gaussian, Poisson, and exponential models (Section 4), and the M -ary extension showing that the optimal rate equals the minimum pairwise weighted Chernoff information.

- (A1) (*Adapted definitions.*) The definitions of the weighted Bhattacharyya affinities and the weighted Chernoff information generalise the classical unweighted quantities of [1,2] and follow the context-sensitive framework of [3,4]; their asymptotic and information-geometric consequences developed below are new.
- (A2) (*Geometric context.*) The information-geometric identities of Section 3.3 are derived in the spirit of the Chentsov–Amari–Nielsen framework [8–11] but are stated and proved for the tilted log-normaliser $\hat{F}(\theta) = \ln \int \varphi(x) e^{\theta^T t(x) + k(x)} d\mu(x)$; the unweighted limit $\varphi \equiv 1$ recovers the classical statements of [11,12].
- (A3) (*Standard tool.*) The concentration argument uses the Azuma–Hoeffding/McDiarmid inequality [13,14]; the novelty lies in its application to the tilted weighted log-likelihood.

1.3. Related Work

The exponential theory of testing errors goes back to Chernoff [1] and Hoeffding [2]. The context-sensitive framework and the weighted information quantities used here were developed in [3,4]. The information-geometric viewpoint on Chernoff information originates with Chentsov [9]; the dually flat structure of exponential and mixture families and the associated α -divergences are developed in [8,10], and the Chernoff point is characterised as the intersection of an exponential geodesic with the Kullback–Leibler bisector in [11]. For $\varphi \equiv 1$, the likelihood-ratio exponential family description is given in [12]; the present paper extends this picture to the tilted integrand $\varphi p^\alpha q^{1-\alpha}$. The minimum-pairwise principle for multiple testing is due to [7]. Weighting mechanisms for covariate-dependent relevance have also been studied outside the asymptotic error-exponent framework, e.g., adaptive-kernel conditional-independence testing [15].

1.4. Structure of the Paper

Section 2 introduces the weighted Bhattacharyya affinities and the weighted Chernoff information. Section 3 proves the main asymptotic result (4) and develops the exponential-family and information-geometric identities. Subsection 3.4 studies the tilted weighted log-likelihood and derives finite- n concentration bounds. Section 4 examines Gaussian, Poisson, and exponential models and includes the M -ary extension. Auxiliary computations are collected in the appendices.

2. Problem Set-Up and Weighted Divergences

2.1. Context-Sensitive Losses and Weighted Total Variation

We keep the binary i.i.d. model from Section 1 and work under Assumption 1. In particular,

$$p(x_1^n) = \prod_{i=1}^n p(x_i), \quad q(x_1^n) = \prod_{i=1}^n q(x_i), \quad \varphi(x_1^n) = \prod_{i=1}^n \varphi(x_i).$$

Define the φ -tilted (reweighted) densities

$$p^*(x) := \frac{\varphi(x)p(x)}{E_\varphi(p)}, \quad q^*(x) := \frac{\varphi(x)q(x)}{E_\varphi(q)}, \quad E_\varphi(p) := \int \varphi(x)p(x) d\mu(x),$$

(and similarly for $E_\varphi(q)$). Throughout this section, we assume that $E_\varphi(p), E_\varphi(q) \in (0, \infty)$. (Equivalently, $\rho_0^W(p, q), \rho_1^W(p, q) \in (0, \infty)$.) Then p^*, q^* are probability densities and, under $\varphi(x_1^n) = \prod_{i=1}^n \varphi(x_i)$, we have $\varphi(x_1^n)p(x_1^n) = E_\varphi(p)^n (p^*)^{\otimes n}(x_1^n)$ (and similarly for q).

Under Assumption 1, we have

$$\int_{\mathcal{X}^n} \varphi(x_1^n) p(x_1^n) d\mu^{\otimes n} = (E_\varphi(p))^n, \quad \int_{\mathcal{X}^n} \varphi(x_1^n) q(x_1^n) d\mu^{\otimes n} = (E_\varphi(q))^n.$$

Let \mathcal{D} denote the class of (possibly randomised) decision rules $D : \mathcal{X}^n \rightarrow [0, 1]$, where $D(x_1^n)$ is the probability of deciding in favour of H_1 after observing x_1^n . (Deterministic rules correspond to $D \in \{0, 1\}$.)

For $D \in \mathcal{D}$, define the context-sensitive type-I and type-II losses by

$$\alpha_\varphi(D) := \mathbb{E}_{\mathbb{P}^{\otimes n}}[\varphi(X_1^n) D(X_1^n)] = \int_{\mathcal{X}^n} \varphi(x_1^n) D(x_1^n) p(x_1^n) d\mu^{\otimes n}(x_1^n), \tag{6}$$

$$\beta_\varphi(D) := \mathbb{E}_{\mathbb{Q}^{\otimes n}}[\varphi(X_1^n) (1 - D(X_1^n))] = \int_{\mathcal{X}^n} \varphi(x_1^n) (1 - D(x_1^n)) q(x_1^n) d\mu^{\otimes n}(x_1^n), \tag{7}$$

and the corresponding total loss

$$L_n(D) := \alpha_\varphi(D) + \beta_\varphi(D), \quad L_n^* := \inf_{D \in \mathcal{D}} L_n(D).$$

Proposition 1 (Pointwise form of the optimal total loss). *For each $n \geq 1$,*

$$L_n^* = \int_{\mathcal{X}^n} \varphi(x_1^n) \min\{p(x_1^n), q(x_1^n)\} d\mu^{\otimes n}(x_1^n). \tag{8}$$

Moreover, an optimal (deterministic) decision rule is given by the likelihood-ratio test

$$D_n^*(x_1^n) = \mathbf{1}\{q(x_1^n) \geq p(x_1^n)\}$$

(with any measurable tie-breaking on $\{p = q\}$).

Proof. Fix x_1^n . The integrand in $L_n(D)$ equals

$$\varphi(x_1^n) (p(x_1^n) D(x_1^n) + q(x_1^n) (1 - D(x_1^n))) = \varphi(x_1^n) (q(x_1^n) + D(x_1^n) (p(x_1^n) - q(x_1^n))).$$

Minimising pointwise over $D(x_1^n) \in [0, 1]$ yields $D_n^*(x_1^n) = 1$ when $p(x_1^n) \leq q(x_1^n)$ and $D_n^*(x_1^n) = 0$ when $p(x_1^n) > q(x_1^n)$, giving (8). \square

We also use the weighted total variation distance

$$\text{TV}_\varphi(\mathbb{P}^{\otimes n}, \mathbb{Q}^{\otimes n}) := \frac{1}{2} \int_{\mathcal{X}^n} \varphi(x_1^n) |p(x_1^n) - q(x_1^n)| d\mu^{\otimes n}(x_1^n). \tag{9}$$

Remark 1. For $\varphi \equiv 1$, this reduces to the usual total variation distance. If φ vanishes on a non-negligible set, TV_φ is, in general, a pseudo-distance; this is sufficient for our purposes since it characterises the weighted losses.

Using $\min\{a, b\} = \frac{1}{2}(a + b - |a - b|)$ in (8) and the definition of TV_φ yields

$$L_n^* = \frac{1}{2} \left((E_\varphi(p))^n + (E_\varphi(q))^n \right) - \text{TV}_\varphi(\mathbb{P}^{\otimes n}, \mathbb{Q}^{\otimes n}). \tag{10}$$

2.2. Weighted Affinities and Chernoff Information

We introduce the weighted Bhattacharyya affinities and the weighted Chernoff information. Assume that $\rho_\alpha^w(p, q) \in (0, \infty)$ for all $\alpha \in [0, 1]$.

Definition 1 (Weighted Bhattacharyya coefficient and distance). For $\alpha \in [0, 1]$ define the weighted α -skewed Bhattacharyya affinity coefficient

$$\rho_\alpha^w(p, q) := \int_{\mathcal{X}} \varphi(x) p(x)^\alpha q(x)^{1-\alpha} d\mu(x), \tag{11}$$

and the corresponding weighted Bhattacharyya distance

$$D_{B,\alpha}^w(p, q) := -\ln \rho_\alpha^w(p, q). \tag{12}$$

Definition 2 (Weighted Chernoff information). The weighted Chernoff information divergence between \mathbb{P} and \mathbb{Q} is

$$D_C^w(\mathbb{P}, \mathbb{Q}) = \max_{\alpha \in [0,1]} \left[-\ln \int_{\mathcal{X}} \varphi(x) p(x)^\alpha q(x)^{1-\alpha} d\mu(x) \right] = \max_{\alpha \in [0,1]} D_{B,\alpha}^w(p, q). \tag{13}$$

A maximiser $\alpha^* = \alpha^*(p, q)$ in (13) is called the optimal Chernoff parameter.

Remark 2. The weighted Chernoff information is symmetric: $D_C^w(\mathbb{P}, \mathbb{Q}) = D_C^w(\mathbb{Q}, \mathbb{P})$, since $\rho_\alpha^w(p, q) = \rho_{1-\alpha}^w(q, p)$. In general, however, D_C^w does not satisfy the triangle inequality and is therefore a divergence rather than a metric.

Remark 3. Under Assumption 1, for every $\alpha \in [0, 1]$,

$$\int_{\mathcal{X}^n} \varphi(x_1^n) p(x_1^n)^\alpha q(x_1^n)^{1-\alpha} d\mu^{\otimes n}(x_1^n) = (\rho_\alpha^w(p, q))^n.$$

Consequently, the weighted Bhattacharyya distances are additive in n and the corresponding Chernoff exponent is of single-letter form.

3. Asymptotics and Information-Geometric Identities

Before stating the main theorem, we separate the two optimisations that appear throughout this section and should not be conflated. The total loss $L_n(D) = \alpha_\varphi(D) + \beta_\varphi(D)$ is a non-negative functional of the decision rule and is *minimised* over $D \in \mathcal{D}$, giving the optimal loss $L_n^* = \inf_{D \in \mathcal{D}} L_n(D)$. The map $\alpha \mapsto -\ln \rho_\alpha^w(p, q)$ is a non-negative concave functional of the skewing parameter and is *maximised* over $\alpha \in [0, 1]$, giving the weighted Chernoff information $D_C^w(\mathbb{P}, \mathbb{Q}) = \sup_{\alpha \in [0,1]} [-\ln \rho_\alpha^w(p, q)]$. Theorem 1 below connects the two via the single-letter asymptotic $L_n^* = \exp\{-nD_C^w(\mathbb{P}, \mathbb{Q}) + o(n)\}$.

3.1. Asymptotics of the Optimal Sum of Losses

Recall from Section 2 that

$$L_n^* := \inf_{D \in \mathcal{D}} [\alpha_\varphi(D) + \beta_\varphi(D)],$$

and that Proposition 1 yields (8). The next theorem identifies its exact logarithmic asymptotic rate with the weighted Chernoff information from Definition 2.

Theorem 1 (Optimal sum of context-sensitive losses). Consider the binary hypotheses $H_0 : X_1^n \sim \mathbb{P}^{\otimes n}$ versus $H_1 : X_1^n \sim \mathbb{Q}^{\otimes n}$, under Assumption 1. Assume also

$$\sup_{\alpha \in [0,1]} \int \varphi(x) \left| \ln \frac{p(x)}{q(x)} \right| p(x)^\alpha q(x)^{1-\alpha} d\mu(x) < \infty. \tag{14}$$

Let \mathcal{D} be the class of (possibly randomised) decision rules $D : \mathcal{X}^n \rightarrow [0, 1]$, and let $\alpha_\varphi(D)$, $\beta_\varphi(D)$ be defined by (6) and (7). Assume that $p, q > 0$ μ -a.e. and that $\rho_\alpha^w(p, q) \in (0, \infty)$ for all $\alpha \in [0, 1]$.

Then, as $n \rightarrow \infty$,

$$L_n^* = \exp\{-nD_C^w(\mathbb{P}, \mathbb{Q}) + o(n)\}. \tag{15}$$

Equivalently,

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \ln L_n^* = D_C^w(\mathbb{P}, \mathbb{Q}).$$

Proof. By Proposition 1,

$$L_n^* = \int_{\mathcal{X}^n} \varphi(x_1^n) \min\{p(x_1^n), q(x_1^n)\} d\mu^{\otimes n}(x_1^n).$$

For any $\alpha \in [0, 1]$, $\min\{a, b\} \leq a^\alpha b^{1-\alpha}$, hence, by factorisation,

$$L_n^* \leq (\rho_\alpha^w(p, q))^n.$$

Taking the infimum over $\alpha \in [0, 1]$ gives

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \ln L_n^* \geq D_C^w(\mathbb{P}, \mathbb{Q}).$$

Now fix $\alpha^* \in \arg \min_{\alpha \in [0, 1]} \rho_\alpha^w(p, q)$ and define

$$r_{\alpha^*}(x) := \frac{\varphi(x)p(x)^{\alpha^*}q(x)^{1-\alpha^*}}{\rho_{\alpha^*}^w(p, q)}, \quad S_n := \sum_{i=1}^n \ln \frac{p(X_i)}{q(X_i)}.$$

A direct change of measure yields

$$L_n^* = (\rho_{\alpha^*}^w(p, q))^n \mathbb{E}_{r_{\alpha^*}^{\otimes n}} \left[e^{-\alpha^* S_n} \mathbf{1}\{S_n > 0\} + e^{(1-\alpha^*) S_n} \mathbf{1}\{S_n \leq 0\} \right].$$

The bracket is bounded above by 1.

Let $F(\alpha) := \ln \rho_\alpha^w(p, q)$, it is easy to check that $F(\alpha)$ is a convex function. Under the regularity assumption of (14),

$$F'(\alpha) = \mathbb{E}_{r_\alpha} \left[\ln \frac{p(X)}{q(X)} \right].$$

If $\alpha^* \in (0, 1)$, then $F'(\alpha^*) = 0$. Hence, in view of LLN, $S_n/n \rightarrow 0$ in r_{α^*} -probability. Therefore, for every $\varepsilon > 0$,

$$\mathbb{E}_{r_{\alpha^*}^{\otimes n}} \left[e^{(1-\alpha^*) S_n} \mathbf{1}\{S_n \leq 0\} + e^{-\alpha^* S_n} \mathbf{1}\{S_n > 0\} \right] \geq e^{-\varepsilon n} r_{\alpha^*}^{\otimes n}(|S_n| \leq \varepsilon n) = \exp\{o(n)\}.$$

Combined with the upper bound by 1, this implies

$$L_n^* = \exp\{-nD_C^w(\mathbb{P}, \mathbb{Q}) + o(n)\}.$$

In the boundary case $\alpha^* = 0$, we have the mean value $m > 0$, and $\mathbf{1}(S_n > 0)e^{-\alpha^* S_n}|_{\alpha^*=0} \rightarrow 1$ a.s. as $n \rightarrow \infty$ by the strong LLN. Similarly, in the case $\alpha^* = 1$, we have the mean value $m \leq 0$, and $\mathbf{1}(S_n \leq 0)e^{(1-\alpha^*) S_n}|_{\alpha^*=1} \rightarrow 1$ a.s. as $n \rightarrow \infty$. This completes the proof. \square

Corollary 1 (Asymptotics of the weighted total variation). *Under the assumptions of Theorem 1, the weighted total variation satisfies*

$$\text{TV}_\varphi(\mathbb{P}^{\otimes n}, \mathbb{Q}^{\otimes n}) = \frac{1}{2} \left((E_\varphi(p))^n + (E_\varphi(q))^n \right) - \exp\{-nD_C^w(\mathbb{P}, \mathbb{Q}) + o(n)\}, \quad n \rightarrow \infty, \quad (16)$$

where $E_\varphi(r) = \int_{\mathcal{X}} \varphi(x)r(x) \, d\mu(x)$.

Proof. Combine the identity (10) with (15). \square

Remark 4. *The weighted α -skewed Bhattacharyya distance (12) appears in many papers, see e.g., [12]. Definition 2 shows that the weighted Chernoff information divergence is the maximally skewed weighted Bhattacharyya distance.*

3.2. Exponential-Family Representation and Uniqueness of α^*

In order to develop an effective computational procedure and to connect the weighted Chernoff information to information geometry, we embed the weighted geometric mixtures of p and q into a one-parameter likelihood-ratio exponential family. For $\alpha \in [0, 1]$ define

$$Z_{pq}(\alpha) := \int_{\mathcal{X}} \varphi(x) p(x)^\alpha q(x)^{1-\alpha} \, d\mu(x) = \rho_\alpha^w(p, q), \quad (17)$$

and the corresponding normalised density

$$\mathcal{E}_{pq} = \left\{ (pq)_\alpha(x) := \frac{\varphi(x) p(x)^\alpha q(x)^{1-\alpha}}{Z_{pq}(\alpha)} : \alpha \in [0, 1] \right\}. \quad (18)$$

By assumption $Z_{pq}(\alpha) \in (0, \infty)$, so $(pq)_\alpha$ is well-defined as a probability density w.r.t. μ .

Set $t(x) := \ln \frac{p(x)}{q(x)}$ and $k_{pq}(x) := \ln \varphi(x) + \ln q(x)$. Then, $(pq)_\alpha$ admits the exponential-family form

$$(pq)_\alpha(x) = \exp\{\alpha t(x) - F_{pq}(\alpha) + k_{pq}(x)\}, \quad (19)$$

$$F_{pq}(\alpha) := \ln Z_{pq}(\alpha) = -D_{B,\alpha}^w(p, q). \quad (20)$$

In particular, $t(X)$ is a sufficient statistic for the family \mathcal{E}_{pq} .

The log-normaliser F_{pq} is convex on $[0, 1]$; if $\ln \frac{p}{q}$ is not μ -a.e. constant on $\{\varphi > 0\}$, then F_{pq} is strictly convex and the maximiser α^* in (13) is unique. By Hölder’s inequality,

$$Z_{pq}(\alpha) = \int_{\mathcal{X}} (\varphi p)^\alpha (\varphi q)^{1-\alpha} \, d\mu \leq (E_\varphi(p))^\alpha (E_\varphi(q))^{1-\alpha}.$$

Finally, note that $Z_{pq}(1) = E_\varphi(p)$ and $Z_{pq}(0) = E_\varphi(q)$; hence,

$$(pq)_1(x) = \frac{\varphi(x)p(x)}{E_\varphi(p)}, \quad (pq)_0(x) = \frac{\varphi(x)q(x)}{E_\varphi(q)},$$

so \mathcal{E}_{pq} is an exponential arc between the tilted versions of \mathbb{P} and \mathbb{Q} . By this definition, the following identities hold:

$$D_C^w(p, q) = D_{B,\alpha^*(p,q)}^w(p, q) = D_{B,\alpha^*(q,p)}^w(q, p) = D_C^w(q, p). \quad (21)$$

3.3. Weighted Bregman Divergence and Information-Geometric Identities

3.3.1. Weighted KL Divergence and Weighted Bregman Divergence

This subsection collects information-geometric identities useful for analysing ρ_α^w and for computing the optimal Chernoff parameter α^* . We follow [3] for weighted Bregman divergences.

Let $\mathcal{E} = \{p_\theta : \theta \in \Theta \subset \mathbb{R}^d\}$ be a regular exponential family of densities (with respect to μ),

$$p_\theta(x) = \exp\{\theta^T t(x) - F(\theta) + k(x)\}, \quad x \in \mathcal{X}. \tag{22}$$

For a density r , set $E_\varphi(r) := \int_{\mathcal{X}} \varphi(x) r(x) d\mu(x)$ and write $E_\varphi(\theta) := E_\varphi(p_\theta)$.

Assume $E_\varphi(\theta) \in (0, \infty)$ for $\theta \in \Theta$ and define the tilted log-normaliser

$$\hat{F}(\theta) := \ln \int_{\mathcal{X}} \varphi(x) e^{\theta^T t(x) + k(x)} d\mu(x) = F(\theta) + \ln E_\varphi(\theta). \tag{23}$$

Equivalently, the tilted density is

$$p_\theta^*(x) = \frac{\varphi(x) p_\theta(x)}{E_\varphi(\theta)}. \tag{24}$$

Definition 3 (Weighted Kullback–Leibler divergence). *For densities p, q on \mathcal{X} define*

$$D_{\text{KL}}^w(p||q) := \int_{\mathcal{X}} \varphi(x) p(x) \ln \frac{p(x)}{q(x)} d\mu(x), \tag{25}$$

whenever the integral is well defined in $(-\infty, \infty]$.

Definition 4 (Weighted Bregman divergence). *The weighted Bregman divergence associated with (F, \hat{F}) is*

$$\begin{aligned} B_{\varphi, F}^w(\theta_1, \theta_2) &:= e^{\hat{F}(\theta_2) - F(\theta_2)} \left[F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla \hat{F}(\theta_2) \right] \\ &= E_\varphi(\theta_2) \left[F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^T \nabla \hat{F}(\theta_2) \right]. \end{aligned} \tag{26}$$

Proposition 2 (Weighted KL as weighted Bregman divergence). *For a regular exponential family $\mathcal{E} = \{p_\theta, \theta \in \Theta\}$, assume that the integral in (25) is well-defined in $(-\infty, \infty]$. Then for any $\theta_1, \theta_2 \in \Theta$,*

$$D_{\text{KL}}^w(p_{\theta_1}||p_{\theta_2}) = B_{\varphi, F}^w(\theta_2, \theta_1). \tag{27}$$

Proof. This identity is stated in [3] (Proposition 4.1); we give a short derivation for completeness. By (22),

$$\ln \frac{p_{\theta_1}(x)}{p_{\theta_2}(x)} = (\theta_1 - \theta_2)^T t(x) - (F(\theta_1) - F(\theta_2)).$$

Substituting this into (25) yields

$$D_{\text{KL}}^w(p_{\theta_1}||p_{\theta_2}) = (\theta_1 - \theta_2)^T \int_{\mathcal{X}} \varphi(x) t(x) p_{\theta_1}(x) d\mu(x) - (F(\theta_1) - F(\theta_2)) E_\varphi(\theta_1).$$

Using (23) and differentiation under the integral sign (regularity of \mathcal{E}),

$$\nabla \hat{F}(\theta_1) = \frac{\int_{\mathcal{X}} \varphi(x) t(x) e^{\theta_1^T t(x) + k(x)} d\mu(x)}{\int_{\mathcal{X}} \varphi(x) e^{\theta_1^T t(x) + k(x)} d\mu(x)} = \frac{\int_{\mathcal{X}} \varphi(x) t(x) p_{\theta_1}(x) d\mu(x)}{E_\varphi(\theta_1)}.$$

Hence, $\int \varphi t p_{\theta_1} d\mu = E_\varphi(\theta_1) \nabla \hat{F}(\theta_1)$, and therefore

$$D_{KL}^W(p_{\theta_1} \| p_{\theta_2}) = E_\varphi(\theta_1) \left[F(\theta_2) - F(\theta_1) - (\theta_2 - \theta_1)^T \nabla \hat{F}(\theta_1) \right] = B_{\varphi, F}^W(\theta_2, \theta_1),$$

which proves (27). \square

Proposition 3 (Primal–dual identities for (weighted) Bregman divergences). *Let F be a log-normaliser of a regular exponential family and let F^* denote its Legendre transform. Write $\theta^* = \nabla F(\theta)$ and $\theta = \nabla F^*(\theta^*)$.*

(a) *Weighted one-parameter identity. Assume $d = 1$ (one-parameter case) and let $\theta_i^* := F'(\theta_i)$. Then, the following weighted analogue of the classical Bregman duality holds:*

$$B_{\varphi, F}^W(\theta_1, \theta_2) = B_{\varphi, F^*}^W(\theta_2^*, \theta_1^*) - (\theta_1 - \theta_2) \ln E_\varphi(\theta_2) + (\theta_2^* - \theta_1^*) \ln E_\varphi(\theta_1), \tag{28}$$

where $B_{\varphi, F}^W$ is as in Definition 4 and B_{φ, F^*}^W is defined analogously (with F replaced by F^* and with the convention $E_\varphi(\theta^*) := E_\varphi(\theta)$ under $\theta = \nabla F^*(\theta^*)$).

(b) *Classical identity. For any $d \geq 1$, the (unweighted) Bregman divergence admits the standard Legendre representation*

$$B_F(\theta_0, \theta_1) = F(\theta_0) + F^*(\theta_1^*) - \theta_0^T \theta_1^*, \quad \theta_1^* = \nabla F(\theta_1), \tag{29}$$

where $B_F(\theta_0, \theta_1) = F(\theta_0) - F(\theta_1) - (\theta_0 - \theta_1)^T \nabla F(\theta_1)$ is the usual (unweighted) Bregman divergence.

Proof. Part (a) is a weighted extension of the classical duality $B_F(\theta_1, \theta_2) = B_{F^*}(\theta_2^*, \theta_1^*)$ and follows by combining the weighted representation (26) with Legendre relations; see also [3]. Part (b) is standard. \square

3.3.2. Weighted Chernoff/Bhattacharyya Quantities Inside an Exponential Family

Let $p_{\theta_1}, p_{\theta_2} \in \mathcal{E}$ and $\theta_\alpha := \alpha\theta_1 + (1 - \alpha)\theta_2$. A direct calculation yields

$$\begin{aligned} \rho_\alpha^W(p_{\theta_1}, p_{\theta_2}) &= \ln \int_{\mathcal{X}} \varphi(x) p_{\theta_1}(x)^\alpha p_{\theta_2}(x)^{1-\alpha} d\mu(x) \\ &= \hat{F}(\theta_\alpha) - \alpha F(\theta_1) - (1 - \alpha)F(\theta_2). \end{aligned} \tag{30}$$

Consequently,

$$\begin{aligned} D_{B, \alpha}^W(p_{\theta_1}, p_{\theta_2}) &= \alpha F(\theta_1) + (1 - \alpha)F(\theta_2) - \hat{F}(\theta_\alpha) \\ &= U_{F, \alpha}(\theta_1, \theta_2) - \ln E_\varphi(\theta_\alpha), \end{aligned} \tag{31}$$

where $U_{F, \alpha}(\theta_1, \theta_2) := \alpha F(\theta_1) + (1 - \alpha)F(\theta_2) - F(\theta_\alpha)$ is the (unweighted) Jensen/Burbea–Rao divergence induced by F . In particular, when $\varphi \equiv 1$ we have $\hat{F} \equiv F$ and $D_{B, \alpha}^W(p_{\theta_1}, p_{\theta_2}) = U_{F, \alpha}(\theta_1, \theta_2)$.

Remark 5 (Geometric mixtures and tilting by φ). *In particular, when $\varphi \equiv 1$, we have $\hat{F} \equiv F$ and the normalised geometric mixture $p_{\theta_1}^\alpha p_{\theta_2}^{1-\alpha}$ belongs to the same exponential family, namely, $\propto p_{\theta_\alpha}$ with $\theta_\alpha = \alpha\theta_1 + (1 - \alpha)\theta_2$.*

Proposition 4 (Optimal Chernoff parameter in an exponential family). *Assume that \hat{F} is strictly convex on the segment $[\theta_1, \theta_2]$ and that the maximiser $\alpha^* \in (0, 1)$ exists. Then, α^* is unique and satisfies*

$$(\theta_1 - \theta_2)^T \nabla \hat{F}(\theta_{\alpha^*}) = F(\theta_1) - F(\theta_2), \quad \theta_{\alpha^*} = \alpha^* \theta_1 + (1 - \alpha^*) \theta_2, \tag{32}$$

with $\nabla \hat{F}(\theta) = \mathbb{E}_{p_\theta^*}[t(X)]$.

Proof. Differentiate (31) with respect to α and use $\frac{d}{d\alpha}\theta_\alpha = \theta_1 - \theta_2$. Strict convexity of \hat{F} on $[\theta_1, \theta_2]$ implies strict concavity of $\alpha \mapsto D_{B,\alpha}^w(p_{\theta_1}, p_{\theta_2})$, hence uniqueness. \square

Proposition 5 (Chernoff information as a Jensen-type divergence and a Bregman bisector). *Let $p_{\theta_1}, p_{\theta_2} \in \mathcal{E}$ and assume that the maximiser $\alpha^* \in (0, 1)$ in Definition 2 exists and is unique. Set $\theta_\alpha = \alpha\theta_1 + (1 - \alpha)\theta_2$. Then*

$$D_C^w(p_{\theta_1}, p_{\theta_2}) = D_{B,\alpha^*}^w(p_{\theta_1}, p_{\theta_2}) = \alpha^*F(\theta_1) + (1 - \alpha^*)F(\theta_2) - \hat{F}(\theta_{\alpha^*}). \tag{33}$$

Moreover, θ_{α^*} is characterised by the weighted Bregman bisector condition

$$B_{\varphi,F}^w(\theta_1, \theta_{\alpha^*}) = B_{\varphi,F}^w(\theta_2, \theta_{\alpha^*}), \tag{34}$$

and the common value recovers the Chernoff information as

$$D_C^w(p_{\theta_1}, p_{\theta_2}) = \frac{1}{E_\varphi(\theta_{\alpha^*})} B_{\varphi,F}^w(\theta_1, \theta_{\alpha^*}) - \ln E_\varphi(\theta_{\alpha^*}) \tag{35}$$

$$= \frac{1}{E_\varphi(\theta_{\alpha^*})} B_{\varphi,F}^w(\theta_2, \theta_{\alpha^*}) - \ln E_\varphi(\theta_{\alpha^*}). \tag{36}$$

In the special case $\varphi \equiv 1$, we have $\hat{F} \equiv F$ and $E_\varphi(\theta) \equiv 1$, so that (33) reduces to the classical Jensen divergence induced by F and (35) becomes $D_C(p_{\theta_1}, p_{\theta_2}) = B_F(\theta_1, \theta_{\alpha^*}) = B_F(\theta_2, \theta_{\alpha^*})$.

Proof. Equation (33) is (31) at $\alpha = \alpha^*$. For (34), expand $B_{\varphi,F}^w(\theta_i, \theta_{\alpha^*}) = E_\varphi(\theta_{\alpha^*}) [F(\theta_i) - F(\theta_{\alpha^*}) - (\theta_i - \theta_{\alpha^*})^T \nabla \hat{F}(\theta_{\alpha^*})]$ and use (32) to see that the difference vanishes. Finally, substituting $\theta_1 - \theta_{\alpha^*} = (1 - \alpha^*)(\theta_1 - \theta_2)$ into $B_{\varphi,F}^w(\theta_1, \theta_{\alpha^*}) / E_\varphi(\theta_{\alpha^*})$ and using (32) gives

$$\frac{1}{E_\varphi(\theta_{\alpha^*})} B_{\varphi,F}^w(\theta_1, \theta_{\alpha^*}) = \alpha^*F(\theta_1) + (1 - \alpha^*)F(\theta_2) - F(\theta_{\alpha^*}).$$

Since $\hat{F}(\theta_{\alpha^*}) = F(\theta_{\alpha^*}) + \ln E_\varphi(\theta_{\alpha^*})$, this yields (35). \square

3.3.3. Derivative and Weighted KL

Recall $F_{pq}(\alpha) = \ln Z_{pq}(\alpha) = \ln \rho_\alpha^w(p, q)$. In view of (14), the differentiation under the integral sign is justified, and we have

$$F'_{pq}(\alpha) = \mathbb{E}_{(pq)_\alpha} \left[\ln \frac{p(X)}{q(X)} \right], \tag{37}$$

where $(pq)_\alpha$ is the Chernoff-tilted density from (18). In particular,

$$F'_{pq}(1) = \frac{1}{E_\varphi(p)} \int_{\mathcal{X}} \varphi(x) p(x) \ln \frac{p(x)}{q(x)} d\mu(x) = \frac{1}{E_\varphi(p)} D_{\text{KL}}^w(p \| q). \tag{38}$$

Analogously, $F'_{pq}(0) = -\frac{1}{E_\varphi(q)} D_{\text{KL}}^w(q \| p)$.

3.3.4. Chernoff–KL

Lemma 1. *Let α^* be a maximiser in Definition 2 and assume that $\alpha^* \in (0, 1)$ (so that $F'_{pq}(\alpha^*) = 0$ below). Set $F_{pq}(\alpha) := \ln \rho_\alpha^w(p, q)$, $r_\alpha := (pq)_\alpha$. Then*

$$D_C^w(p, q) = D_{\text{KL}}(r_{\alpha^*} \| r_1) - \ln E_\varphi(p) = D_{\text{KL}}(r_{\alpha^*} \| r_0) - \ln E_\varphi(q), \tag{39}$$

where $r_1(x) = \varphi(x)p(x)/E_\varphi(p)$ and $r_0(x) = \varphi(x)q(x)/E_\varphi(q)$.

Here, D_{KL} denotes the standard (unweighted) Kullback–Leibler divergence.

Proof. A direct computation yields, for $\alpha \in [0, 1]$,

$$D_{KL}(r_\alpha \| r_1) = -(1 - \alpha)F'_{pq}(\alpha) - F_{pq}(\alpha) + F_{pq}(1),$$

$$D_{KL}(r_\alpha \| r_0) = \alpha F'_{pq}(\alpha) - F_{pq}(\alpha) + F_{pq}(0).$$

At $\alpha = \alpha^*$, we have $F'_{pq}(\alpha^*) = 0$. Since $F_{pq}(1) = \ln E_\varphi(p)$ and $F_{pq}(0) = \ln E_\varphi(q)$, the claim follows from $D_C^w(p, q) = -F_{pq}(\alpha^*)$. \square

Corollary 2 (Chernoff information as a Bregman divergence on the Chernoff arc). *Let $F_{pq}(\alpha) := \ln \rho_\alpha^w(p, q)$ and define the one-dimensional Bregman divergence*

$$B_{F_{pq}}(a, b) := F_{pq}(a) - F_{pq}(b) - (a - b)F'_{pq}(b).$$

Assume that the maximiser $\alpha^* \in (0, 1)$ in Definition 2 is interior, so that $F'_{pq}(\alpha^*) = 0$. Then

$$D_C^w(p, q) = B_{F_{pq}}(1, \alpha^*) - \ln E_\varphi(p) = B_{F_{pq}}(0, \alpha^*) - \ln E_\varphi(q). \tag{40}$$

Equivalently,

$$B_{F_{pq}}(1, \alpha^*) = D_{KL}(r_{\alpha^*} \| r_1) = D_C^w(p, q) + \ln E_\varphi(p),$$

$$B_{F_{pq}}(0, \alpha^*) = D_{KL}(r_{\alpha^*} \| r_0) = D_C^w(p, q) + \ln E_\varphi(q),$$

with $r_\alpha = (pq)_\alpha$, $r_1 = \varphi p / E_\varphi(p)$ and $r_0 = \varphi q / E_\varphi(q)$.

Proof. In the Chernoff exponential family $\{r_\alpha\}$ with log-normalizer F_{pq} , the KL–Bregman identity gives $D_{KL}(r_{\alpha^*} \| r_1) = B_{F_{pq}}(1, \alpha^*)$ and $D_{KL}(r_{\alpha^*} \| r_0) = B_{F_{pq}}(0, \alpha^*)$. Since $F'_{pq}(\alpha^*) = 0$ and $F_{pq}(1) = \ln E_\varphi(p)$, $F_{pq}(0) = \ln E_\varphi(q)$, we obtain

$$B_{F_{pq}}(1, \alpha^*) = -F_{pq}(\alpha^*) + F_{pq}(1) = D_C^w(p, q) + \ln E_\varphi(p),$$

and similarly for 0. Substituting this into Lemma 1 yields (40). \square

Remark 6 (One-parameter case). Assume $d = 1$, $\theta_1 \neq \theta_2$, and that the maximiser $\alpha^* \in (0, 1)$ in Definition 2 is interior. Assume moreover that \hat{F}' is strictly increasing on Θ and set $\hat{G} = (\hat{F}')^{-1}$. Then (32) yields

$$\alpha^* = \frac{1}{\theta_1 - \theta_2} \left(\hat{G} \left(\frac{F(\theta_1) - F(\theta_2)}{\theta_1 - \theta_2} \right) - \theta_2 \right). \tag{41}$$

When $\varphi \equiv 1$, we have $\hat{F} \equiv F$ and (41) reduces to the classical formula.

To illustrate the general identities above, we provide explicit expressions for $D_{B,\alpha}^w$ and D_C^w in several parametric settings (Gaussian, Poisson, and exponential); see Section 4.

Remark 7. The Bregman representation (40) identifies $D_C^w(\mathbb{P}, \mathbb{Q})$ with a Bregman divergence on the tilted log-normaliser \hat{F} ; geometrically, the optimiser α^* marks the intersection of the exponential geodesic of the tilted family $\{\varphi p^\alpha q^{1-\alpha}\}_{\alpha \in [0,1]}$ with the weighted Kullback–Leibler bisector, generalising the unweighted characterisation of [11,12]. In the exponential-family setting, the computation of D_C^w therefore reduces to \hat{F} and its gradient: once \hat{F} is available in closed form, α^* is determined

by (41) in the one-parameter case and by a monotone equation in the natural-parameter space in general, without evaluating ρ_α^w for each α . The examples of Section 4 exemplify this reduction.

3.4. Tilted Weighted Likelihood and Concentration Bounds

Although the optimal rule in the context-sensitive problem is still the usual likelihood-ratio test q/p (cf. Section 3), it is convenient to work with the tilted ratio q^*/p^* . The factor φ cancels pointwise in q^*/p^* and enters only through the normalisation constants $E_\varphi(p)$ and $E_\varphi(q)$. We record two consequences: a large-deviation representation for L^*/n via the cumulant generating functions $\psi_{\mathbb{P}}, \psi_{\mathbb{Q}}$ and their Legendre transforms and a finite- n concentration bound based on a martingale argument.

For the tilted distributions, the log-likelihood takes the form

$$\begin{aligned} L^*(X_1^n) &= L^*(X_1, \dots, X_n) = \sum_{i=1}^n \ln \frac{q^*(X_i)}{p^*(X_i)} \\ &= \sum_{i=1}^n \ln \frac{q(X_i)}{p(X_i)} - n \ln E_\varphi(q) + n \ln E_\varphi(p). \end{aligned} \tag{42}$$

Here $E_\varphi(p) = \int \varphi(x) p(x) d\mu(x)$.

In particular, since

$$\ln \frac{q^*(x)}{p^*(x)} = \ln \frac{q(x)}{p(x)} + \ln E_\varphi(p) - \ln E_\varphi(q),$$

we may equivalently rewrite likelihood-ratio threshold rules in terms of L^* . For example,

$$\sum_{i=1}^n \ln \frac{q(X_i)}{p(X_i)} \geq 0 \iff L^*(X_1^n) \geq n(\ln E_\varphi(p) - \ln E_\varphi(q)).$$

Thus, L^* is the usual log-likelihood ratio, shifted by a constant determined by the context weight φ .

The log of the moment generating function and its Legendre transform take the form

$$\begin{aligned} \psi_{\mathbb{P}}(\alpha) &= \ln \mathbb{E}_{\mathbb{P}} \left[e^{\alpha \ln \frac{q^*(X)}{p^*(X)}} \right] \\ &= \ln \int_{\mathcal{X}} q(x)^\alpha p(x)^{1-\alpha} d\mu(x) - \alpha \ln E_\varphi(q) + \alpha \ln E_\varphi(p), \\ I_{\mathbb{P}}(r) &= \sup_{\alpha} [\alpha r - \psi_{\mathbb{P}}(\alpha)], \end{aligned} \tag{43}$$

where α ranges over the set $\{\alpha \in \mathbb{R} : \psi_{\mathbb{P}}(\alpha) < \infty\}$ (and similarly for $\psi_{\mathbb{Q}}$).

Similarly,

$$\begin{aligned} \psi_{\mathbb{Q}}(\alpha) &:= \ln \mathbb{E}_{\mathbb{Q}} \left[e^{\alpha \ln \frac{q^*(X)}{p^*(X)}} \right] = \ln \mathbb{E}_{\mathbb{P}} \left[e^{\ln \frac{q(X)}{p(X)} + \alpha \ln \frac{q^*(X)}{p^*(X)}} \right] \\ &= \ln \int_{\mathcal{X}} q(x)^{\alpha+1} p(x)^{-\alpha} d\mu(x) - \alpha \ln E_\varphi(q) + \alpha \ln E_\varphi(p). \end{aligned} \tag{44}$$

This implies the relation of Legendre transforms

$$I_{\mathbb{Q}}(r) = I_{\mathbb{P}}(r) - r + \ln E_\varphi(p) - \ln E_\varphi(q). \tag{45}$$

In particular, $I_{\mathbb{P}}(0)$ may be treated as a natural weighted version of the Chernoff divergence between q and p :

$$\begin{aligned}
 I_{\mathbb{P}}(0) &= \sup_{\alpha} \left[-\ln \int_{\mathcal{X}} q(x)^{\alpha} p(x)^{1-\alpha} d\mu(x) + \alpha \ln E_{\varphi}(q) - \alpha \ln E_{\varphi}(p) \right] \\
 &=: \hat{D}_{\mathbb{C}}^{\mathbb{W}}(q, p).
 \end{aligned}
 \tag{46}$$

Interpretation. The value $I_{\mathbb{P}}(0)$ is the Chernoff–Cramér exponent controlling the tail event $\{L^*/n \geq 0\}$ under \mathbb{P} , i.e., (under standard regularity assumptions), $\mathbb{P}(L^*(X_1^n) \geq 0) \asymp e^{-nI_{\mathbb{P}}(0)}$. In the unweighted case $\varphi \equiv 1$, we have $E_{\varphi}(p) = E_{\varphi}(q) = 1$ and $I_{\mathbb{P}}(0)$ reduces to the classical Chernoff information between p and q . We also stress that $\hat{D}_{\mathbb{C}}^{\mathbb{W}}(q, p) = I_{\mathbb{P}}(0)$ is a *tilted-likelihood* exponent and is distinct from the weighted Chernoff information $D_{\mathbb{C}}^{\mathbb{W}}(\mathbb{P}, \mathbb{Q})$ from Definition 2, which governs the optimal sum of context-sensitive losses.

Non-Asymptotic Concentration Via a Doob Martingale

The rate functions $I_{\mathbb{P}}$ and $I_{\mathbb{Q}}$ capture the exponential scale of deviations of L^*/n as $n \rightarrow \infty$. To obtain explicit *finite-n* bounds, we now apply a standard martingale method to L^* under \mathbb{Q} .

Consider the filtration $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$ and define the random variables $\{U_k, k = 0, \dots, n\}$ by

$$\begin{aligned}
 U_k &= \mathbb{E}_{\mathbb{Q}}[L^*(X_1, \dots, X_n) | \mathcal{F}_k] \\
 &= \sum_{j=1}^k \ln \frac{q(X_j)}{p(X_j)} + (n - k) D_{\text{KL}}(\mathbb{Q} \| \mathbb{P}) - n \ln E_{\varphi}(q) + n \ln E_{\varphi}(p),
 \end{aligned}
 \tag{47}$$

where $D_{\text{KL}}(\mathbb{Q} \| \mathbb{P}) = \int q(x) \ln \frac{q(x)}{p(x)} d\mu(x)$ stands for the (unweighted) Kullback–Leibler divergence of \mathbb{Q} and \mathbb{P} . Then

$$U_k - U_{k-1} = \ln \frac{q(X_k)}{p(X_k)} - D_{\text{KL}}(\mathbb{Q} \| \mathbb{P}), \quad k = 1, \dots, n.
 \tag{48}$$

Observe that $\{U_k, \mathcal{F}_k\}$ is a martingale w.r.t. \mathbb{Q} .

Assume now that $d < \infty$ and $\sigma^2 < \infty$, where $d = \sup_{x \in \mathcal{X}} \left| \ln \frac{q(x)}{p(x)} - D_{\text{KL}}(\mathbb{Q} \| \mathbb{P}) \right|$ and

$$\begin{aligned}
 \sigma^2 &= \mathbb{E}_{\mathbb{Q}} \left[(U_k - U_{k-1})^2 | \mathcal{F}_{k-1} \right] \\
 &= \int_{\mathcal{X}} q(x) \left(\ln \frac{q(x)}{p(x)} - D_{\text{KL}}(\mathbb{Q} \| \mathbb{P}) \right)^2 d\mu(x).
 \end{aligned}
 \tag{49}$$

In view of a refined Azuma–Hoeffding/McDiarmid inequality [13,14],

$$\begin{aligned}
 \mathbb{P}_{\mathbb{Q}} \left(L^*(X_1, \dots, X_n) > (D_{\text{KL}}(\mathbb{Q} \| \mathbb{P}) + \beta)n - (\ln E_{\varphi}(q) - \ln E_{\varphi}(p))n \right) \\
 \leq \exp \left\{ -nD \left(\frac{\delta + \gamma^*}{1 + \gamma^*} \parallel \frac{\gamma^*}{1 + \gamma^*} \right) \right\},
 \end{aligned}
 \tag{50}$$

where $\delta = \frac{\sigma^2}{d^2}$, $\gamma^* = \frac{\beta^*}{d}$, $\beta^* = \beta - \ln E_{\varphi}(q) + \ln E_{\varphi}(p)$ and $D(p \| q) = p \ln \frac{p}{q} + (1 - p) \ln \frac{1-p}{1-q}$ stands for the Kullback–Leibler divergence between the two Bernoulli distributions $(p, 1 - p)$ and $(q, 1 - q)$.

We use the following modified version of Azuma–Hoeffding inequality; see [14].

Lemma 2. Let $\{U_k, \mathcal{F}_k\}$ be a discrete-time real-valued martingale. Assume that, for some constants $d, \sigma > 0$, the following two requirements are satisfied a.s. for every $k \in \{1, \dots, n\}$:

$$\begin{aligned} |U_k - U_{k-1}| &\leq d, \\ \text{Var}[U_k - U_{k-1} | \mathcal{F}_{k-1}] &\leq \sigma^2. \end{aligned} \tag{51}$$

Then, for every $\beta \geq 0$,

$$\mathbb{P}(|U_n - U_0| \geq \beta n) \leq 2 \exp\left\{-nD\left(\frac{\delta + \gamma}{1 + \gamma} \parallel \frac{\gamma}{1 + \gamma}\right)\right\}, \tag{52}$$

where $\delta = \frac{\sigma^2}{d^2}$ and $\gamma = \frac{\beta}{d}$. (Note that $\delta \in [0, 1]$ automatically whenever $|U_k - U_{k-1}| \leq d$ a.s.)

In particular, the one-sided bound $\mathbb{P}(U_n - U_0 \geq \beta n) \leq \exp\{-nD(\cdot)\}$ holds.

Theorem 2. Set

$$\beta^* = \beta - \ln E_\varphi(p) + \ln E_\varphi(q), \quad \gamma^* = \frac{\beta^*}{d}.$$

Under conditions $d < \infty, \sigma^2 < \infty$, and $\beta^* \geq 0$,

$$\mathbb{P}_\mathbb{Q}(L^*(X_1, \dots, X_n) \geq \beta n) \leq \exp\left\{-nD\left(\frac{\delta + \gamma^*}{1 + \gamma^*} \parallel \frac{\gamma^*}{1 + \gamma^*}\right)\right\}. \tag{53}$$

Lemma 2 is quoted from [14]. Theorem 2 is its direct application to the tilted log-likelihood ratio (42); the only dependence on the context weight φ is through $E_\varphi(p)$ and $E_\varphi(q)$.

4. Examples and Applications

The identities of Section 3 reduce the computation of $D_{B,\alpha}^w$ and D_C^w to the single-letter weighted affinity

$$\rho_\alpha^w(p, q) = \int \varphi(x) p(x)^\alpha q(x)^{1-\alpha} d\mu(x),$$

followed by optimisation over $\alpha \in [0, 1]$. We work this out for Gaussian, Poisson, and exponential families, highlighting how the context weight φ modifies the classical formulas. When $\varphi \equiv 1$, the expressions reduce to the standard unweighted Bhattacharyya and Chernoff quantities; more involved non-exponential-family computations (such as the Cauchy location–scale family) are deferred to the Appendix A.

4.1. A Numerical Illustration

This subsection illustrates the behaviour of α^* and D_C^w under a non-trivial factorised weight and serves as a direct numerical verification of the Bregman identities of Section 3.3: for the model below the affinity ρ_α^w is available in closed form, and we check that closed-form evaluation and direct numerical integration agree to machine precision.

Consider the asymmetric Gaussian hypotheses

$$H_0 : \mathbb{P} = \mathcal{N}(\mu_0, \sigma_0^2), \quad H_1 : \mathbb{Q} = \mathcal{N}(\mu_1, \sigma_1^2), \quad (\mu_0, \sigma_0^2) = (0, 1), \quad (\mu_1, \sigma_1^2) = (3, 2),$$

with a non-indicator factorised weight

$$\varphi(x) = \exp(-\beta(x - x_0)^2), \quad x_0 = 0, \beta \geq 0. \tag{54}$$

At $\beta = 0$, one has $\varphi \equiv 1$ and the unweighted Chernoff information is recovered. For $\beta > 0$, the weight concentrates near $x_0 = \mu_0$; in particular, (54) is not an indicator-type weight, so the weighted problem does not reduce to the unweighted Chernoff information on a restricted domain.

The asymmetry $\sigma_0^2 \neq \sigma_1^2$ and the centring of φ at the H_0 mean are essential for the illustration. In a fully symmetric configuration ($\sigma_0 = \sigma_1, \mu_1 = -\mu_0, x_0 = 0$), the problem is invariant under $\alpha \leftrightarrow 1 - \alpha$, so the optimum is pinned to $\alpha^* = 1/2$ for every β and the effect of the weight on the Chernoff compromise is invisible. Asymmetric hypotheses are precisely where the weighted formalism is operationally distinct from the classical one, and it is this distinction that the numerics below is designed to expose.

Writing $A(\alpha) = \alpha/\sigma_0^2 + (1 - \alpha)/\sigma_1^2 + 2\beta$, $B(\alpha) = \alpha\mu_0/\sigma_0^2 + (1 - \alpha)\mu_1/\sigma_1^2 + 2\beta x_0$, and $C(\alpha) = \alpha\mu_0^2/\sigma_0^2 + (1 - \alpha)\mu_1^2/\sigma_1^2 + 2\beta x_0^2$, a direct Gaussian integration yields

$$\ln \rho_\alpha^w(p, q) = \frac{1}{2} \ln \frac{2\pi}{A(\alpha)} - \frac{\alpha}{2} \ln(2\pi\sigma_0^2) - \frac{1-\alpha}{2} \ln(2\pi\sigma_1^2) - \frac{1}{2} \left(C(\alpha) - \frac{B(\alpha)^2}{A(\alpha)} \right), \tag{55}$$

and maximising (55) over $\alpha \in [0, 1]$ gives $\alpha^*(\beta)$ and $D_C^w(\mathbb{P}, \mathbb{Q})(\beta)$. Table 1 reports their values at three selected β . Direct numerical integration of ρ_α^w from its definition agrees with (55) to machine precision on all tabulated entries, which confirms the Bregman identities of Section 3.3 numerically for this example.

Table 1. Optimal skewing parameter and weighted Chernoff information for the Gaussian example (54), obtained by maximising (55).

β	$\alpha^*(\beta)$	$D_C^w(\mathbb{P}, \mathbb{Q})$
0 (unweighted)	0.4153	0.8018
1/16	0.3355	0.9827
1/4	0.0963	1.4935

The monotone growth of $\beta \mapsto D_C^w$ and the leftward shift of $\alpha^*(\beta)$ illustrate a qualitative conclusion of Section 3: localising φ near μ_0 preferentially retains observations that are more probable under H_0 and thereby increases the effective discrimination rate, while simultaneously moving the optimal tilting towards the H_0 side. The classical unweighted limit is recovered at $\beta = 0$.

In the language of hypothesis testing, α^* is the parameter that balances the exponential rates of the type-I and type-II losses at the Bayes optimum: the type-I exponent equals $\alpha^* D_C^w$ and the type-II exponent equals $(1 - \alpha^*) D_C^w$ (cf. Section 3). A leftward shift of α^* therefore corresponds to reallocating the available exponential budget towards faster decay of the type-II loss at the expense of the type-I loss, which is the optimal response to a weight that concentrates mass in the region where H_0 is more plausible.

Data and Code Availability

A Jupyter/Colab notebook reproducing Table 1 and Figures 1–3, together with the direct-integration verification of (55), is archived on Zenodo [16] and mirrored on GitHub.

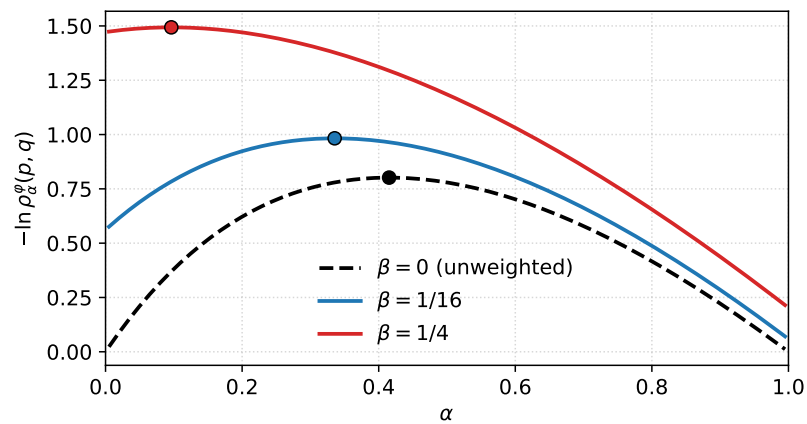


Figure 1. The map $\alpha \mapsto -\ln \rho_\alpha^w(p, q)$ for the Gaussian hypotheses $\mathcal{N}(0, 1), \mathcal{N}(3, 2)$ with weight (54), for $\beta \in \{0, 1/16, 1/4\}$. The optimum α^* is marked by a bullet on each curve and shifts to the left as β increases.

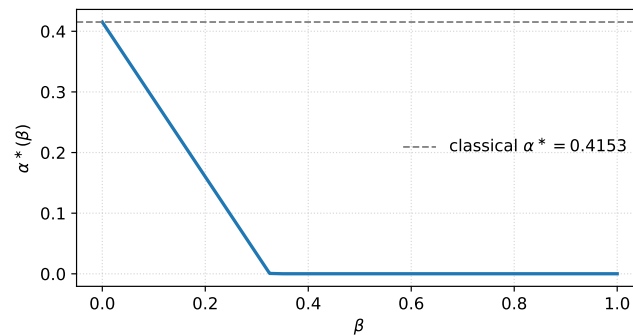


Figure 2. Optimal skewing parameter $\alpha^*(\beta)$ for the Gaussian example. The dashed line marks the unweighted value $\alpha^*(0)$.

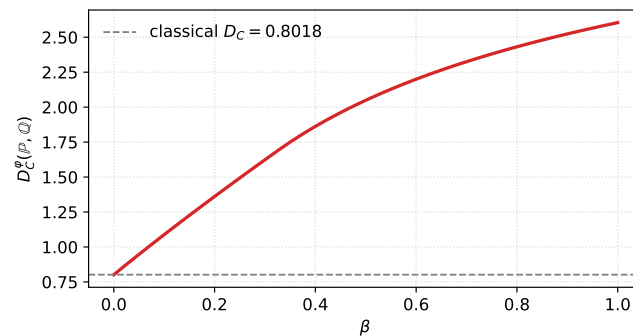


Figure 3. Weighted Chernoff information $\beta \mapsto D_C^w(\mathbb{P}, \mathbb{Q})$. The dashed line marks the classical value D_C recovered at $\beta = 0$.

4.2. Gaussian Models

Throughout this subsection, the reference measure is the Lebesgue measure on \mathbb{R}^d . We compute the weighted Bhattacharyya coefficient

$$\rho_\alpha^w(\mathbb{P}, \mathbb{Q}) = \int_{\mathbb{R}^d} \varphi(x) p(x)^\alpha q(x)^{1-\alpha} dx, \quad \alpha \in [0, 1],$$

together with the weighted Bhattacharyya distance $D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) := -\ln \rho_\alpha^w(\mathbb{P}, \mathbb{Q})$ and the weighted Chernoff information $D_C^w(\mathbb{P}, \mathbb{Q}) = \max_{\alpha \in [0, 1]} D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q})$ (Definition 2). Note that, unlike the unweighted case, ρ_α^w is not restricted to $[0, 1]$ and $D_{B,\alpha}^w$ (hence also D_C^w) may take negative values.

Example 1 (Gaussian weighted Bhattacharyya coefficient with exponential weight). Let $\mathbb{P} = \mathcal{N}(\mu_1, \Sigma_1)$ and $\mathbb{Q} = \mathcal{N}(\mu_2, \Sigma_2)$ on \mathbb{R}^d , where $\Sigma_1 \succ 0$ and $\Sigma_2 \succ 0$, and let $\varphi(x) = e^{\gamma^\top x}$ for some $\gamma \in \mathbb{R}^d$. Denote by p, q the corresponding densities. For $\alpha \in [0, 1]$ define

$$\begin{aligned} \Sigma_\alpha &:= \left(\alpha \Sigma_1^{-1} + (1 - \alpha) \Sigma_2^{-1}\right)^{-1}, \\ \tilde{\mu}_\alpha &:= \Sigma_\alpha \left(\alpha \Sigma_1^{-1} \mu_1 + (1 - \alpha) \Sigma_2^{-1} \mu_2 + \gamma\right). \end{aligned}$$

Then

$$\begin{aligned} \rho_\alpha^w(\mathbb{P}, \mathbb{Q}) &= \int_{\mathbb{R}^d} e^{\gamma^\top x} p(x)^\alpha q(x)^{1-\alpha} dx \\ &= \frac{|\Sigma_\alpha|^{1/2}}{|\Sigma_1|^{\alpha/2} |\Sigma_2|^{(1-\alpha)/2}} \exp\left\{-\frac{1}{2} \left(\alpha \mu_1^\top \Sigma_1^{-1} \mu_1 + (1 - \alpha) \mu_2^\top \Sigma_2^{-1} \mu_2 - \tilde{\mu}_\alpha^\top \Sigma_\alpha^{-1} \tilde{\mu}_\alpha\right)\right\}. \end{aligned} \tag{56}$$

Consequently,

$$\begin{aligned} D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) &= \\ &= \frac{1}{2} \left(\alpha \mu_1^\top \Sigma_1^{-1} \mu_1 + (1 - \alpha) \mu_2^\top \Sigma_2^{-1} \mu_2 - \tilde{\mu}_\alpha^\top \Sigma_\alpha^{-1} \tilde{\mu}_\alpha + \ln \frac{|\Sigma_1|^\alpha |\Sigma_2|^{1-\alpha}}{|\Sigma_\alpha|}\right). \end{aligned} \tag{57}$$

In particular, setting $\gamma = 0$ (i.e., $\varphi \equiv 1$) reduces (57) to the classical (unweighted) Gaussian Bhattacharyya distance; see, e.g., [12].

Corollary 3 (Common covariance). In Example 1, assume $\Sigma_1 = \Sigma_2 = \Sigma \succ 0$ and keep the exponential weight $\varphi(x) = e^{\gamma^\top x}$. Set $\delta := \mu_1 - \mu_2$, $\|v\|_{\Sigma^{-1}}^2 := v^\top \Sigma^{-1} v$, and $\mu_\alpha := \alpha \mu_1 + (1 - \alpha) \mu_2$. Then, for any $\alpha \in [0, 1]$,

$$\begin{aligned} \rho_\alpha^w(\mathbb{P}, \mathbb{Q}) &= \int_{\mathbb{R}^d} e^{\gamma^\top x} p(x)^\alpha q(x)^{1-\alpha} dx \\ &= \exp\left\{-\frac{\alpha(1 - \alpha)}{2} \|\delta\|_{\Sigma^{-1}}^2 + \gamma^\top \mu_\alpha + \frac{1}{2} \gamma^\top \Sigma \gamma\right\}, \end{aligned} \tag{58}$$

and therefore

$$D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) = -\ln \rho_\alpha^w(\mathbb{P}, \mathbb{Q}) = \frac{\alpha(1 - \alpha)}{2} \|\delta\|_{\Sigma^{-1}}^2 - \gamma^\top \mu_\alpha - \frac{1}{2} \gamma^\top \Sigma \gamma. \tag{59}$$

If $\mu_1 \neq \mu_2$ and the unconstrained maximiser

$$\tilde{\alpha} = \frac{1}{2} - \frac{\gamma^\top \delta}{\|\delta\|_{\Sigma^{-1}}^2}$$

belongs to $(0, 1)$, then $\alpha^* = \tilde{\alpha}$; otherwise the maximum over $\alpha \in [0, 1]$ is attained at the nearest boundary point $\alpha^* \in \{0, 1\}$. In all cases,

$$D_C^w(\mathbb{P}, \mathbb{Q}) = \max_{\alpha \in [0,1]} D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) = D_{B,\alpha^*}^w(\mathbb{P}, \mathbb{Q}).$$

In particular, for $\gamma = 0$ (i.e., $\varphi \equiv 1$) we recover $\alpha^* = 1/2$ and $D_C(\mathbb{P}, \mathbb{Q}) = \|\delta\|_{\Sigma^{-1}}^2 / 8$.

Proof. The expression for ρ_α^w follows by simplifying Example 1 under $\Sigma_1 = \Sigma_2 = \Sigma$, which makes the determinant prefactor equal to 1 and yields a Gaussian MGF term $\exp(\gamma^\top \mu_\alpha + \frac{1}{2} \gamma^\top \Sigma \gamma)$. The maximiser follows by differentiating (59) in α . \square

Choosing an exponential weight $\varphi(x) = e^{\gamma^T x}$ corresponds to exponential tilting: for a Gaussian $X \sim \mathcal{N}(\mu, \Sigma)$, this tilting keeps the covariance and shifts the mean to $\mu + \Sigma\gamma$ (with normalisation factor $\exp(\gamma^T \mu + \frac{1}{2} \gamma^T \Sigma \gamma)$), which is why the weighted affinities remain available in closed form. In particular, the optimal Chernoff parameter is no longer forced to be $\alpha^* = 1/2$ and, as (59) shows, sufficiently strong tilting can push the maximiser to the boundary $\alpha^* \in \{0, 1\}$.

4.3. Poisson Models

Example 2 (Poisson model with exponential weight). Let $\mathcal{X} = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ and let μ be the counting measure on \mathcal{X} . Fix two hypotheses $\mathbb{P} = \text{Poi}(\lambda_1)$ and $\mathbb{Q} = \text{Poi}(\lambda_2)$ with $\lambda_1, \lambda_2 > 0$, and write $p = p_{\lambda_1}$ and $q = q_{\lambda_2}$. Throughout this subsection we work under the standing assumption of Section 2, namely, that the observations X_1, \dots, X_n are i.i.d. (distributed as \mathbb{P} under H_0 and as \mathbb{Q} under H_1), and that the weight φ factorises across observations (Assumption 1). We still consider the exponential weight $\varphi_\gamma(k) = e^{\gamma k}$, $\gamma \in \mathbb{R}$. (For $\gamma = 0$ we recover the unweighted case $\varphi \equiv 1$.) Equivalently, setting $\varepsilon := e^\gamma > 0$, the weight takes the form $\varphi(k) = \varepsilon^k$; this reparameterisation is convenient in applications where $\varepsilon \in (0, 1)$ models a per-event discount factor, while $\gamma \in \mathbb{R}$ is the natural parameter for the exponential-family calculations below. The two parameterisations are equivalent.

For $\alpha \in [0, 1]$, set

$$\lambda_\alpha := \lambda_1^\alpha \lambda_2^{1-\alpha}.$$

(a) **Weighted Bhattacharyya coefficient and Chernoff arc.** A direct summation gives

$$\rho_\alpha^w(\mathbb{P}, \mathbb{Q}) = \sum_{k=0}^\infty \varphi_\gamma(k) p(k)^\alpha q(k)^{1-\alpha} = \exp\{-\alpha\lambda_1 - (1-\alpha)\lambda_2 + e^\gamma \lambda_\alpha\}. \tag{60}$$

Hence, by Definition 1,

$$D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) = -\ln \rho_\alpha^w(\mathbb{P}, \mathbb{Q}) = \alpha\lambda_1 + (1-\alpha)\lambda_2 - e^\gamma \lambda_\alpha. \tag{61}$$

Moreover, the normalised weighted geometric mixture (Chernoff-tilted density) from (18) takes the form

$$(pq)_\alpha(k) = \frac{\varphi_\gamma(k) p(k)^\alpha q(k)^{1-\alpha}}{\rho_\alpha^w(\mathbb{P}, \mathbb{Q})} = \exp\{-e^\gamma \lambda_\alpha\} \frac{(e^\gamma \lambda_\alpha)^k}{k!} = \text{Poi}(e^\gamma \lambda_\alpha).$$

(b) **Optimal Chernoff parameter.** If $\lambda_1 \neq \lambda_2$, then $\alpha \mapsto D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q})$ is strictly concave on $[0, 1]$ since

$$\frac{d^2}{d\alpha^2} D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) = -e^\gamma \lambda_\alpha \left(\ln \frac{\lambda_1}{\lambda_2}\right)^2 < 0;$$

hence, the maximiser α^* in Definition 2 is unique. Differentiating (61) yields the critical point condition

$$0 = \frac{d}{d\alpha} D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) = \lambda_1 - \lambda_2 - e^\gamma \lambda_\alpha \ln\left(\frac{\lambda_1}{\lambda_2}\right).$$

Equivalently, the (unconstrained) critical point $\alpha = \tilde{\alpha}$ satisfies

$$\lambda_{\tilde{\alpha}} = e^{-\gamma} L(\lambda_1, \lambda_2), \quad L(\lambda_1, \lambda_2) := \frac{\lambda_1 - \lambda_2}{\ln \lambda_1 - \ln \lambda_2}.$$

In contrast to the unweighted case ($\gamma = 0$), the context tilt γ may push the optimal Chernoff parameter to the boundary $\alpha^* \in \{0, 1\}$.

Thus, the unconstrained maximiser is

$$\tilde{\alpha} = \frac{\ln L(\lambda_1, \lambda_2) - \gamma - \ln \lambda_2}{\ln \lambda_1 - \ln \lambda_2},$$

and the maximiser on $[0, 1]$ is $\alpha^* = \Pi_{[0,1]}(\tilde{\alpha})$, where

$$\Pi_{[0,1]}(a) := \min\{1, \max\{0, a\}\}.$$

Finally,

$$D_C^w(\mathbb{P}, \mathbb{Q}) = \max_{\alpha \in [0,1]} D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) = D_{B,\alpha^*}^w(\mathbb{P}, \mathbb{Q}).$$

If $\lambda_1 = \lambda_2$, then $\rho_\alpha^w(\mathbb{P}, \mathbb{Q})$ does not depend on α and $D_C^w(\mathbb{P}, \mathbb{Q}) = D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q})$ for any $\alpha \in [0, 1]$.

Derivation of (60). For $k \in \mathbb{N}_0$,

$$p(k)^\alpha q(k)^{1-\alpha} = \exp\{-\alpha\lambda_1 - (1-\alpha)\lambda_2\} \frac{\lambda_\alpha^k}{k!},$$

so multiplying by $e^{\gamma k}$ and summing over k gives

$$\begin{aligned} \rho_\alpha^w(\mathbb{P}, \mathbb{Q}) &= \exp\{-\alpha\lambda_1 - (1-\alpha)\lambda_2\} \sum_{k \geq 0} (e^\gamma \lambda_\alpha)^k / k! \\ &= \exp\{-\alpha\lambda_1 - (1-\alpha)\lambda_2 + e^\gamma \lambda_\alpha\}. \end{aligned}$$

□

4.4. Exponential Models

Example 3 (Exponential model with exponential weight). Let $\mathcal{X} = \mathbb{R}_+ = [0, \infty)$ with Lebesgue measure. Fix two hypotheses $\mathbb{P} = \text{Exp}(\lambda_1)$ and $\mathbb{Q} = \text{Exp}(\lambda_2)$ with rates $\lambda_1, \lambda_2 > 0$, and write

$$p(x) = \lambda_1 e^{-\lambda_1 x} \mathbf{1}\{x \geq 0\}, \quad q(x) = \lambda_2 e^{-\lambda_2 x} \mathbf{1}\{x \geq 0\}.$$

Consider the exponential weight $\varphi_\gamma(x) = e^{\gamma x}$ with

$$\gamma < \min\{\lambda_1, \lambda_2\},$$

so that $\rho_\alpha^w(\mathbb{P}, \mathbb{Q}) \in (0, \infty)$ for all $\alpha \in [0, 1]$. For $\alpha \in [0, 1]$ set

$$\lambda_\alpha := \alpha\lambda_1 + (1-\alpha)\lambda_2.$$

(a) **Weighted Bhattacharyya coefficient and Chernoff arc.** A direct computation gives

$$\rho_\alpha^w(\mathbb{P}, \mathbb{Q}) = \int_0^\infty e^{\gamma x} p(x)^\alpha q(x)^{1-\alpha} dx = \frac{\lambda_1^\alpha \lambda_2^{1-\alpha}}{\lambda_\alpha - \gamma}.$$

Hence

$$D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) = -\ln \rho_\alpha^w(\mathbb{P}, \mathbb{Q}) = \ln(\lambda_\alpha - \gamma) - \alpha \ln \lambda_1 - (1-\alpha) \ln \lambda_2.$$

Moreover, the Chernoff-tilted density $(pq)_\alpha$ from (18) is again exponential:

$$(pq)_\alpha(x) = \frac{e^{\gamma x} p(x)^\alpha q(x)^{1-\alpha}}{\rho_\alpha^w(\mathbb{P}, \mathbb{Q})} = (\lambda_\alpha - \gamma) e^{-(\lambda_\alpha - \gamma)x} \mathbf{1}\{x \geq 0\} = \text{Exp}(\lambda_\alpha - \gamma).$$

(b) Optimal Chernoff parameter. If $\lambda_1 \neq \lambda_2$, then $\alpha \mapsto D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q})$ is strictly concave on $[0, 1]$; hence, the maximiser α^* in Definition 2 is unique. Differentiating yields the critical point condition

$$\frac{\lambda_1 - \lambda_2}{\lambda_\alpha - \gamma} = \ln\left(\frac{\lambda_1}{\lambda_2}\right).$$

Equivalently, the (unconstrained) critical point $\alpha = \tilde{\alpha}$ satisfies

$$\lambda_{\tilde{\alpha}} - \gamma = L(\lambda_1, \lambda_2), \quad L(\lambda_1, \lambda_2) := \frac{\lambda_1 - \lambda_2}{\ln \lambda_1 - \ln \lambda_2},$$

so that

$$\tilde{\alpha} = \frac{\gamma + L(\lambda_1, \lambda_2) - \lambda_2}{\lambda_1 - \lambda_2}.$$

The maximiser on $[0, 1]$ is $\alpha^* = \Pi_{[0,1]}(\tilde{\alpha})$ (projection onto $[0, 1]$), and

$$D_C^w(\mathbb{P}, \mathbb{Q}) = \max_{\alpha \in [0,1]} D_{B,\alpha}^w(\mathbb{P}, \mathbb{Q}) = D_{B,\alpha^*}^w(\mathbb{P}, \mathbb{Q}).$$

If $\lambda_1 = \lambda_2 = \lambda$, then $\rho_\alpha^w(\mathbb{P}, \mathbb{Q}) = \lambda / (\lambda - \gamma)$ does not depend on α , so any $\alpha \in [0, 1]$ is optimal and $D_C^w(\mathbb{P}, \mathbb{Q}) = \ln(\lambda - \gamma) - \ln \lambda$. Setting $\gamma = 0$ (i.e., $\varphi \equiv 1$) recovers the classical unweighted expressions.

Additional Example (Baseline, Non-Exponential Family)

Appendix A contains a closed-form illustration for the Cauchy location–scale family. Since the Cauchy family is not an exponential family, this example complements the main text by showing that, even in the unweighted baseline case $\varphi \equiv 1$, the Bhattacharyya coefficient (in particular $\rho_{1/2}$) and the Chernoff information may involve special functions (complete elliptic integrals). For nontrivial weights φ , the symmetry $\rho_\alpha = \rho_{1-\alpha}$ (hence $\alpha^* = 1/2$) typically fails and a comparable closed form is not available, so we keep the baseline Cauchy computation in the appendix.

4.5. Extension to M-ary Hypothesis Testing

We now record the finite- M analogue of Theorem 1. The key observation is that the optimal M -ary pointwise loss is squeezed between pairwise minima (Lemma 3), and each pairwise term has logarithmic rate given by the corresponding weighted Chernoff information. Hence the overall M -ary rate is determined by the closest pair in terms of D_C^w .

Fix an integer $M \geq 2$ and let $\mathbb{P}_1, \dots, \mathbb{P}_M$ be probability measures on \mathcal{X} dominated by μ , with strictly positive densities p_1, \dots, p_M . Let $X_1^n = (X_1, \dots, X_n)$ be i.i.d. under each hypothesis $H_i : X_1^n \sim \mathbb{P}_i^{\otimes n}$. Assume that the weight function factorises as in Assumption 1.

Assume moreover that for every $1 \leq i < j \leq M$ and every $\alpha \in [0, 1]$,

$$\rho_\alpha^w(p_i, p_j) = \int_{\mathcal{X}} \varphi(x) p_i(x)^\alpha p_j(x)^{1-\alpha} d\mu(x) \in (0, \infty),$$

so that all pairwise weighted Chernoff information values are well-defined and inequality

$$\max_{i \neq j} \sup_{\alpha \in [0,1]} \int \varphi(x) \left| \ln \frac{p_i(x)}{p_j(x)} \right| p_i(x)^\alpha p_j(x)^{1-\alpha} d\mu(x) < \infty \tag{62}$$

holds true.

A (deterministic) M -ary decision rule is a measurable map $\delta_n : \mathcal{X}^n \rightarrow \{1, \dots, M\}$. Define the context-sensitive loss under H_i by

$$L_{i,n}(\delta_n) := \mathbb{E}_{\mathbb{P}_i^{\otimes n}}[\varphi(X_1^n) \mathbf{1}\{\delta_n(X_1^n) \neq i\}],$$

and the total loss

$$L_{n,M}(\delta_n) := \sum_{i=1}^M L_{i,n}(\delta_n), \quad L_{n,M}^* := \inf_{\delta_n} L_{n,M}(\delta_n).$$

Proposition 6 (Pointwise form of the optimal M -ary total loss). *For each $n \geq 1$,*

$$L_{n,M}^* = \int_{\mathcal{X}^n} \varphi(x_1^n) \left(\sum_{i=1}^M p_i(x_1^n) - \max_{1 \leq j \leq M} p_j(x_1^n) \right) d\mu^{\otimes n}(x_1^n), \tag{63}$$

where $p_i(x_1^n) = \prod_{k=1}^n p_i(x_k)$. Moreover, an optimal rule is given by the maximum-likelihood classifier

$$\delta_n^*(x_1^n) \in \arg \max_{1 \leq j \leq M} p_j(x_1^n)$$

(with any measurable tie-breaking).

Proof. Fix δ_n . Using $\sum_{i=1}^M \mathbf{1}\{\delta_n \neq i\} p_i = \sum_{i=1}^M p_i - p_{\delta_n}$ pointwise, we obtain

$$L_{n,M}(\delta_n) = \int_{\mathcal{X}^n} \varphi(x_1^n) \left(\sum_{i=1}^M p_i(x_1^n) - p_{\delta_n(x_1^n)}(x_1^n) \right) d\mu^{\otimes n}(x_1^n).$$

Minimisation over δ_n is therefore pointwise in x_1^n and is achieved by selecting an index maximising $p_j(x_1^n)$, yielding (63). \square

Lemma 3 (Pairwise minima). *For any non-negative numbers a_1, \dots, a_M ,*

$$\max_{1 \leq i < j \leq M} \min(a_i, a_j) \leq \sum_{k=1}^M a_k - \max_{1 \leq k \leq M} a_k \leq \sum_{1 \leq i < j \leq M} \min(a_i, a_j). \tag{64}$$

Consequently, defining for $i < j$

$$I_n^{i,j} := \int_{\mathcal{X}^n} \varphi(x_1^n) \min\{p_i(x_1^n), p_j(x_1^n)\} d\mu^{\otimes n}(x_1^n),$$

we have the sandwich inequality

$$\max_{i < j} I_n^{i,j} \leq L_{n,M}^* \leq \sum_{i < j} I_n^{i,j}. \tag{65}$$

Proof. Let $a_{(1)} \geq a_{(2)} \geq \dots \geq a_{(M)}$ be the decreasing rearrangement of (a_1, \dots, a_M) . Then $\sum_k a_k - \max_k a_k = \sum_{r=2}^M a_{(r)} \geq a_{(2)}$. Moreover, $\max_{i < j} \min(a_i, a_j) = a_{(2)}$, proving the left inequality in (64).

For the right inequality, let $k^* \in \arg \max_k a_k$. Then

$$\sum_{1 \leq i < j \leq M} \min(a_i, a_j) \geq \sum_{k \neq k^*} \min(a_k, a_{k^*}) = \sum_{k \neq k^*} a_k = \sum_{k=1}^M a_k - \max_k a_k.$$

Applying (64) pointwise to $a_i = p_i(x_1^n)$, multiplying by $\varphi(x_1^n)$ and integrating yields (65). \square

Theorem 3 (*M*-ary exponent equals the minimum pairwise weighted Chernoff information). For $1 \leq i < j \leq M$, let $D_C^w(\mathbb{P}_i, \mathbb{P}_j)$ be the weighted Chernoff information as in Definition 2, and (62) holds true. Set

$$C_M^w := \min_{1 \leq i < j \leq M} D_C^w(\mathbb{P}_i, \mathbb{P}_j).$$

Then the optimal *M*-ary total loss satisfies

$$L_{n,M}^* = \exp\{-nC_M^w + o(n)\}, \quad n \rightarrow \infty, \tag{66}$$

or equivalently,

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \ln L_{n,M}^* = C_M^w.$$

Proof. Fix $1 \leq i < j \leq M$ and consider the binary testing problem between \mathbb{P}_i and \mathbb{P}_j with the same factorised weight φ . The optimal binary total loss equals

$$I_n^{i,j} = \int_{\mathcal{X}^n} \varphi(x_1^n) \min\{p_i(x_1^n), p_j(x_1^n)\} d\mu^{\otimes n}(x_1^n),$$

and by Theorem 1 applied to the pair $(\mathbb{P}_i, \mathbb{P}_j)$,

$$I_n^{i,j} = \exp\{-nD_C^w(\mathbb{P}_i, \mathbb{P}_j) + o_{i,j}(n)\}.$$

Since the number of pairs is finite, letting $r_n := \max_{i < j} |o_{i,j}(n)|$ yields $r_n = o(n)$ and

$$I_n^{i,j} = \exp\{-nD_C^w(\mathbb{P}_i, \mathbb{P}_j) + O(r_n)\} \quad \text{uniformly over } i < j.$$

Now use the sandwich inequality(65). Let (i^*, j^*) attain the minimum $C_M^w = D_C^w(\mathbb{P}_{i^*}, \mathbb{P}_{j^*})$. From the lower bound,

$$L_{n,M}^* \geq I_n^{i^*,j^*} = \exp\{-nC_M^w + O(r_n)\}.$$

From the upper bound,

$$L_{n,M}^* \leq \sum_{i < j} I_n^{i,j} \leq \binom{M}{2} \exp\{-nC_M^w + O(r_n)\}.$$

Taking $-\frac{1}{n} \ln(\cdot)$ and letting $n \rightarrow \infty$ yields (66). \square

Remark 8 (Nonzero priors do not change the exponent). Let $w_1, \dots, w_M > 0, \sum_i w_i = 1$, and consider the Bayesian weighted total loss $L_{n,M}^{(w)}(\delta_n) = \sum_{i=1}^M w_i L_{i,n}(\delta_n)$ with optimum $L_{n,M}^{(w)*} := \inf_{\delta_n} L_{n,M}^{(w)}(\delta_n)$. Then, the exponent remains C_M^w . Indeed, writing $w_{\min} := \min_i w_i$ and $w_{\max} := \max_i w_i$, for any δ_n ,

$$w_{\min} L_{n,M}(\delta_n) \leq L_{n,M}^{(w)}(\delta_n) \leq w_{\max} L_{n,M}(\delta_n),$$

and taking infimum over δ_n gives $w_{\min} L_{n,M}^* \leq L_{n,M}^{(w)*} \leq w_{\max} L_{n,M}^*$. Hence, $-\frac{1}{n} \ln L_{n,M}^{(w)*}$ and $-\frac{1}{n} \ln L_{n,M}^*$ have the same limit C_M^w .

5. Conclusions

We studied context-sensitive simple hypothesis testing under a multiplicative weight and proved that the optimal total loss admits the single-letter logarithmic asymptotic

$$L_n^* = \exp\{-nD_C^w(\mathbb{P}, \mathbb{Q}) + o(n)\}.$$

The rate is the weighted Chernoff information. The main structural ingredient is an exponential-family embedding of the weighted geometric mixtures $\varphi p^\alpha q^{1-\alpha}$, which yields the characterisation of the optimal Chernoff parameter through the log-normaliser and leads to weighted information-geometric identities. We also derived finite- n concentration bounds for the tilted weighted log-likelihood, obtained explicit formulas in Gaussian, Poisson, and exponential models, and extended the logarithmic asymptotic to finitely many hypotheses through the minimum pairwise weighted Chernoff information.

Open Problems

The single-letter representation (4) rests on Assumption 1 (factorised weights), on the integrability of the tilted log-normaliser \hat{F} , and on i.i.d. sampling of simple hypotheses. Relaxing these assumptions defines several natural open directions.

- (a) *Non-factorised weights.* Replacing Assumption 1 by a weight $\varphi(x_1^n) = \psi(\frac{1}{n} \sum_{i=1}^n h(x_i))$ or a pairwise-interaction weight; the single-letter rate is then expected to be replaced by a variational formula over the space of probability measures, in the spirit of Sanov/Gibbs conditioning.
- (b) *Integrability.* Weights with heavy tails in the sufficient statistic $t(x)$ may violate the finiteness of \hat{F} near $\alpha^* \theta_1 + (1 - \alpha^*) \theta_2$; the boundary cases $\alpha^* \in \{0, 1\}$ (cf. Proposition 5 and [3]) call for a systematic treatment via truncation or a change in base measure.
- (c) *Dependent observations.* A weighted counterpart of the Gärtner–Ellis theorem for stationary ergodic sequences, along the lines of the weighted extensions in [4].
- (d) *Composite hypotheses.* A weighted analogue of the generalised likelihood-ratio test and its exponent, with sup/inf characterisations in terms of D_C^w over the composite parameter sets.
- (e) *Information geometry of weighted manifolds.* Extending the Chentsov–Amari framework [8–12] to weighted statistical manifolds; the Fisher metric and the dually flat structure depend on symmetries that φ breaks in a controlled way.

Author Contributions: Conceptualization, methodology, and supervision, M.K.; formal analysis and validation, M.K. and E.Y.K.; software, visualization, data curation, and project administration, E.Y.K.; writing—original draft preparation, M.K. and E.Y.K.; writing—review and editing, E.Y.K.; funding acquisition, M.K. and E.Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: The work by MK was carried out in the framework of a research project HSE-BR-2025-039 implemented as part of the Basic Research Program at HSE University. The second author (E.Yu. Kalimulina) was supported by the Ministry of Science and Higher Education of the Russian Federation under the state assignment (project FFNU-2025-0029).

Data Availability Statement: The Python/Jupyter notebook reproducing Table 1 and Figures 1–3, together with the direct numerical-integration verification of the Gaussian formula in Section 4.1, is openly available on Zenodo. No new experimental datasets were generated.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Cauchy Location–Scale Family

We consider the univariate Cauchy location–scale family

$$p_{l,s}(x) = \frac{s}{\pi(s^2 + (x - l)^2)}, \quad x \in \mathbb{R}, \quad l \in \mathbb{R}, \quad s > 0.$$

In this appendix, we set $\varphi \equiv 1$, so the weighted quantities from Section 2 coincide with their classical counterparts (e.g., $\rho_\alpha^w = \rho_\alpha$, $D_{B,\alpha}^w = D_{B,\alpha}$, $D_C^w = D_C$). This example provides a closed-form illustration outside the exponential-family setting.

Although the main body of the paper develops *weighted* Chernoff/Bhattacharyya quantities, we include the Cauchy location–scale family as an *unweighted baseline* closed-form benchmark outside the exponential-family setting (note the appearance of complete elliptic integrals even in the classical case). This appendix is not used in the proofs of the weighted results; it serves as a sanity check and illustrates the analytic complexity of ρ_α beyond exponential families. Nontrivial weights must satisfy the finiteness conditions from Section 2; for heavy-tailed Cauchy laws, common exponential tilts $\varphi(x) = e^{\gamma x}$ violate these conditions for any $\gamma \neq 0$, and in general, a nonconstant weight also breaks the symmetry leading to $\alpha^* = 1/2$.

Appendix A.1. KL Divergence (Closed Form)

Proposition A1 (KL divergence between two Cauchy laws). *Let $\mathbb{P} = \text{Cauchy}(l_1, s_1)$ and $\mathbb{Q} = \text{Cauchy}(l_2, s_2)$ with $s_1, s_2 > 0$. Then*

$$D_{\text{KL}}(\mathbb{P}||\mathbb{Q}) = \int_{\mathbb{R}} p_{l_1, s_1}(x) \ln \frac{p_{l_1, s_1}(x)}{p_{l_2, s_2}(x)} dx = \ln \frac{(s_1 + s_2)^2 + (l_1 - l_2)^2}{4s_1 s_2}. \tag{A1}$$

Appendix A.2. Chernoff Parameter and Bhattacharyya Coefficient

Recall the (unweighted) α -Bhattacharyya coefficient and distance

$$\rho_\alpha(\mathbb{P}, \mathbb{Q}) = \int_{\mathbb{R}} p_{l_1, s_1}(x)^\alpha p_{l_2, s_2}(x)^{1-\alpha} dx, \quad D_{B,\alpha}(\mathbb{P}, \mathbb{Q}) = -\ln \rho_\alpha(\mathbb{P}, \mathbb{Q}), \quad \alpha \in [0, 1].$$

Proposition A2 (Chernoff parameter for Cauchy). *For the Cauchy location–scale family (with $\varphi \equiv 1$), one has the symmetry*

$$\rho_\alpha(\mathbb{P}, \mathbb{Q}) = \rho_{1-\alpha}(\mathbb{P}, \mathbb{Q}), \quad \alpha \in [0, 1]; \tag{A2}$$

equivalently, $D_{B,\alpha}(\mathbb{P}, \mathbb{Q}) = D_{B,1-\alpha}(\mathbb{P}, \mathbb{Q})$. Consequently, if $\mathbb{P} \neq \mathbb{Q}$, then the Chernoff maximiser is unique and equals

$$\alpha^* = \frac{1}{2}, \quad \text{and hence} \quad D_C(\mathbb{P}, \mathbb{Q}) = \max_{\alpha \in [0,1]} D_{B,\alpha}(\mathbb{P}, \mathbb{Q}) = D_{B,1/2}(\mathbb{P}, \mathbb{Q}).$$

If $\mathbb{P} = \mathbb{Q}$, then $D_{B,\alpha}(\mathbb{P}, \mathbb{Q}) \equiv 0$ and every $\alpha \in [0, 1]$ is a maximiser.

Proof. Set $F(\alpha) := \ln \rho_\alpha(\mathbb{P}, \mathbb{Q})$. By Hölder’s inequality, F is convex on $[0, 1]$, and it is strictly convex unless $p_{l_1, s_1} / p_{l_2, s_2}$ is μ -a.e. constant (equivalently, unless $\mathbb{P} = \mathbb{Q}$). Hence $D_{B,\alpha}(\mathbb{P}, \mathbb{Q}) = -F(\alpha)$ is concave (strictly concave when $\mathbb{P} \neq \mathbb{Q}$). For Cauchy laws, the symmetry (A2) holds; see, e.g., [17] for an invariance-based proof. Therefore, $D_{B,\alpha}$ is symmetric about $1/2$, and concavity implies that it attains its maximum at $\alpha = 1/2$. Strict concavity yields uniqueness when $\mathbb{P} \neq \mathbb{Q}$. \square

Appendix A.3. Closed Form for $\rho_{1/2}$ and D_C

Lemma A1 (A standard elliptic-integral identity). *Let $a, b > 0$ and $d \in \mathbb{R}$. Define the complete elliptic integral of the first kind*

$$K(m) := \int_0^{\pi/2} \frac{du}{\sqrt{1 - m \sin^2 u}}, \quad m \in [0, 1].$$

Then

$$\int_{-\infty}^{\infty} \frac{dx}{\sqrt{(x^2 + a^2)((x - d)^2 + b^2)}} = \frac{4}{\sqrt{(a + b)^2 + d^2}} K\left(\frac{(a - b)^2 + d^2}{(a + b)^2 + d^2}\right). \tag{A3}$$

Proposition A3 (Bhattacharyya coefficient for Cauchy, closed form). Let $\mathbb{P} = \text{Cauchy}(l_1, s_1)$ and $\mathbb{Q} = \text{Cauchy}(l_2, s_2)$ with $s_1, s_2 > 0$, and set $\delta := l_1 - l_2$. Then

$$\rho_{1/2}(\mathbb{P}, \mathbb{Q}) = \int_{\mathbb{R}} \sqrt{p_{l_1, s_1}(x) p_{l_2, s_2}(x)} dx = \frac{4\sqrt{s_1 s_2}}{\pi\sqrt{(s_1 + s_2)^2 + \delta^2}} K\left(\frac{(s_1 - s_2)^2 + \delta^2}{(s_1 + s_2)^2 + \delta^2}\right). \tag{A4}$$

Consequently,

$$D_C(\mathbb{P}, \mathbb{Q}) = D_{B,1/2}(\mathbb{P}, \mathbb{Q}) = -\ln \rho_{1/2}(\mathbb{P}, \mathbb{Q}), \tag{A5}$$

with $\rho_{1/2}(\mathbb{P}, \mathbb{Q})$ given by (A4).

Proof. By definition,

$$\sqrt{p_{l_1, s_1}(x) p_{l_2, s_2}(x)} = \frac{\sqrt{s_1 s_2}}{\pi\sqrt{((x - l_1)^2 + s_1^2)((x - l_2)^2 + s_2^2)}}.$$

Shift $x \mapsto x + l_2$ to obtain

$$\rho_{1/2}(\mathbb{P}, \mathbb{Q}) = \frac{\sqrt{s_1 s_2}}{\pi} \int_{-\infty}^{\infty} \frac{dx}{\sqrt{(x^2 + s_2^2)((x - \delta)^2 + s_1^2)}}.$$

Apply Lemma A1 with $a = s_2, b = s_1, d = \delta$ to get (A4). Finally, Proposition A2 yields $D_C(\mathbb{P}, \mathbb{Q}) = D_{B,1/2}(\mathbb{P}, \mathbb{Q})$, hence (A5). \square

Remark A1 (Useful special cases). If $l_1 = l_2$ (common location), then $\delta = 0$ and (A4) reduces to

$$\rho_{1/2}(\mathbb{P}, \mathbb{Q}) = \frac{4\sqrt{s_1 s_2}}{\pi(s_1 + s_2)} K\left(\left(\frac{s_1 - s_2}{s_1 + s_2}\right)^2\right).$$

If $s_1 = s_2 = s$ (common scale), then

$$\rho_{1/2}(\mathbb{P}, \mathbb{Q}) = \frac{4s}{\pi\sqrt{4s^2 + \delta^2}} K\left(\frac{\delta^2}{4s^2 + \delta^2}\right).$$

Remark A2 (Weighted case). For a general weight φ , one has

$$\rho_{1/2}^w(\mathbb{P}, \mathbb{Q}) = \int_{\mathbb{R}} \varphi(x) \sqrt{p_{l_1, s_1}(x) p_{l_2, s_2}(x)} dx.$$

The symmetry (A2) (and hence $\alpha^* = 1/2$) typically fails unless φ is compatible with the invariance argument used in [17]. We therefore use the Cauchy model mainly as a closed-form baseline example at $\varphi \equiv 1$.

References

1. Chernoff, H. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **1952**, *23*, 493–507. <https://doi.org/10.1214/aoms/1177729330>.
2. Hoeffding, W. Asymptotically optimal tests for multinomial distributions. *Ann. Math. Stat.* **1965**, *36*, 369–401. <https://doi.org/10.1214/aoms/1177700150>.
3. Kelbert, M.; Suhov Y. Context-sensitive hypothesis-testing and exponential families, *Statistics* **2025**, *59*, 845–878.
4. Kelbert, M.; Suhov Y. On basic context-dependent concepts of Information Theory and Statistics, *Theory Probab. Appl.* **2026**, *70*, 563–583.

5. Kalimulina, E.Y. Application of multi-valued logic models in traffic aggregation problems in mobile networks. In Proceedings of the 2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, 13–15 October 2021; pp. 1–6.
6. Esin, A.A.; Kalimulina, E.Y. Markov-modulated queueing network for mobile traffic aggregation with threshold-controlled buffers. *Math. Model. Numer. Simul. Appl.* **2026**, *6*, 4. <https://doi.org/10.53391/2791-8564.1019>.
7. Nielsen, F. Hypothesis testing, information divergence and computational geometry. In *Geometric Science of Information*; Nielsen, F., Barbaresco, F., Eds.; GSI 2013, Lecture Notes in Computer Science. Springer: Berlin/Heidelberg, Germany, 2013; Volume 8085, pp. 241–248. https://doi.org/10.1007/978-3-642-40020-9_25.
8. Amari, S.-I.; Nagaoka, H. *Methods of Information Geometry*; Translations of Mathematical Monographs; American Mathematical Society: Providence, RI, USA, 2000; Volume 191. <https://doi.org/10.1090/mmono/191>.
9. Chentsov, N.N. *Statistical Decision Rules and Optimal Inference*; Translations of Mathematical Monographs; American Mathematical Society: Providence, RI, USA, 1982, Volume 53.
10. Amari, S.-I. *Information Geometry and Its Applications*; Applied Mathematical Sciences; Springer: Japan, Tokyo, 2016; Volume 194. <https://doi.org/10.1007/978-4-431-55978-8>.
11. Nielsen, F. An information-geometric characterization of Chernoff information. *IEEE Signal Process. Lett.* **2013**, *20*, 269–272. <https://doi.org/10.1109/LSP.2013.2243726>.
12. Nielsen, F. Revisiting Chernoff information with likelihood ratio exponential families. *Entropy* **2022**, *24*, 1400. <https://doi.org/10.3390/e24101400>.
13. Azuma, K. Weighted sums of certain dependent random variables, *Tôhoku Math. J.* **1967**, *19*, 357–367.
14. McDiarmid, C. On the method of bounded differences. In *Surveys in Combinatorics*; Cambridge University Press: Cambridge, UK, 1989; Volume 141, pp. 148–188.
15. Ren, Y.; Zhang, J.; Xia, Y.; Wang, R.; Xie, F.; Guan, J.; Zhang, H.; Zhou, S. Regression-based conditional independence test with adaptive kernels. *Artif. Intell.* **2025**, *347*, 104391. <https://doi.org/10.1016/j.artint.2025.104391>.
16. Kalimulina, E.Y. Weighted Chernoff Information–Numerical Illustration. Software, companion to the present paper, version 1.0.0. *Zenodo* **2026**. <https://doi.org/10.5281/zenodo.19736237>.
17. Nielsen, F.; Okamura K. On f -divergences between Cauchy distributions, *IEEE Trans. Inf. Theory* **2023**, *69*, 3150–3170.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.