A. Background on Chernoff Information

In this section, we provide a brief background on Chernoff bounds and Chernoff information, leading to the derivation of the results under equal priors, i.e., $\pi_0 = \pi_1 = \frac{1}{2}$. We discuss the case of unequal priors in Appendix E.

Consider a detector of the form $T(x) \geq \tau$ for classification between two hypothesis $H_0 : X \sim P_0(x)$ and $H_1 : X \sim P_1(x)$. Recall that the log-generating functions for this detector are defined as follows:

$$\Lambda_0(u) = \log \mathbb{E}[e^{uT(X)}|H_0], \text{ and } \Lambda_1(u) = \log \mathbb{E}[e^{uT(X)}|H_1]. \tag{5}$$

A.1. Proof of Lemma 1

We first state the Chernoff bound (see Chapter 2.2 in (Boucheron et al., 2013)) here, which is a well-known tight bound for approximating error probabilities. For a random variable T,

$$\Pr(T \geq \tau) = \Pr(e^{uT} \geq e^{u\tau}) \leq \frac{\mathbb{E}[e^{uT}]}{e^{u\tau}} \;\; \forall u > 0. \tag{6}$$

Proof of Lemma 1. Using the Chernoff bound, we can bound $P_{\text{FP}}^{(T)}(\tau)$ as follows:

$$P_{\text{FP}}^{(T)}(\tau) = \Pr(T(X) \geq \tau | H_0) \leq \frac{\mathbb{E}[e^{uT(X)}|H_0]}{e^{u\tau}} = \frac{e^{\Lambda_0(u)}}{e^{u\tau}} \;\; \forall u > 0. \tag{7}$$

Thus, $-\log P_{\text{FP}}^{(T)}(\tau) \geq \sup_{u>0} (u\tau - \Lambda_0(u)) = E_{\text{FP}}^{(T)}(\tau)$. Similarly, using the Chernoff bound, we have

$$P_{\text{FN}}^{(T)}(\tau) = \Pr(T(X) < \tau | H_1) \leq \frac{\mathbb{E}[e^{uT(X)}|H_1]}{e^{u\tau}} = \frac{e^{\Lambda_1(u)}}{e^{u\tau}} \;\; \forall u < 0. \tag{8}$$

Thus, $-\log P_{\text{FN}}^{(T)}(\tau) \geq \sup_{u<0} (u\tau - \Lambda_1(u)) = E_{\text{FN}}^{(T)}(\tau)$. □

A.2. Properties of log-generating functions

Here, we state some useful properties of the log-generating functions that are used later in the other proofs/explanations.

Property 1 (Convexity)**.** *The log-generating functions $\Lambda_0(u)$ and $\Lambda_1(u)$ are convex in u.*

Proof of Property 1. The proof follows directly using Hölder's inequality. For any u and v, and $\alpha \in [0,1]$,

$$\mathbb{E}[e^{(\alpha u+(1-\alpha)v)T(X)}|H_0] = \mathbb{E}[e^{\alpha uT(X)}e^{(1-\alpha)vT(X)}|H_0] \leq \left(\mathbb{E}[|e^{\alpha uT(X)}|^{\frac{1}{\alpha}}|H_0]\right)^{\alpha} \left(\mathbb{E}[|e^{(1-\alpha)vT(X)}|^{\frac{1}{1-\alpha}}|H_0]\right)^{1-\alpha}. \tag{9}$$

This leads to,

$$\begin{aligned}\Lambda_0(\alpha u + (1-\alpha)v) = \log \mathbb{E}[e^{(\alpha u+(1-\alpha)v)T(X)}|H_0] &\leq \alpha \log \mathbb{E}[e^{uT(X)}|H_0] + (1-\alpha)\log \mathbb{E}[e^{vT(X)}|H_0] \\ &= \alpha\Lambda_0(u) + (1-\alpha)\Lambda_0(v).\end{aligned} \tag{10}$$

The proof is similar for $\Lambda_1(u)$. □

Property 2 (Zero at origin)**.** *The log-generating functions $\Lambda_0(u)$ and $\Lambda_1(u)$ are both 0 at $u = 0$.*

Proof of Property 2. The proof follows by substituting $u = 0$ in the expressions of $\Lambda_0(u)$ and $\Lambda_1(u)$. □

Next, we prove some properties for the log-generating functions when the detector is *well-behaved*. In general, when using a detector of the form $T(x) \geq \tau$, we would expect $T(X)$ to be high when H_1 is true, and low when H_0 is true. We call a detector *well-behaved* if $\mathbb{E}[T(X)|H_0] < 0$ and $\mathbb{E}[T(X)|H_1] > 0$. The next property provides more intuition on what the log-generating functions look like for *well-behaved* detectors.

Property 3 (Log-generating functions of well-behaved detectors)**.** *Suppose that* $\mathbb{E}[T(X)|H_0] < 0$ *and* $\mathbb{E}[T(X)|H_1] > 0$*, and* $P_0(x)$ *and* $P_1(x)$ *are non-zero for all* x*. Then, the following holds:*

- $\Lambda_0(u)$ *and* $\Lambda_1(u)$ *are strictly convex.*
- $\Lambda_0(u) > 0$ *if* $u < 0$. $\Lambda_1(u) > 0$ *if* $u > 0$.

Proof of Property 3. The convexity of $\Lambda_0(u)$ is proved in Property 1. Now $\Lambda_0(u)$ is strictly convex if, for all distinct reals u and v, and $\alpha \in (0, 1)$, we have,

$$\Lambda_0(\alpha u + (1-\alpha)v) < \alpha\Lambda_0(u) + (1-\alpha)\Lambda_0(v).$$

For the sake of contradiction, let us assume that there exists u and v with $v > u$ such that,

$$\Lambda_0(\alpha u + (1-\alpha)v) = \alpha\Lambda_0(u) + (1-\alpha)\Lambda_0(v).$$

This indicates that Hölder's inequality holds with exact equality in (9), which could happen if and only if $ae^{uT(x)} = be^{vT(x)}$ almost everywhere with respect to the probability measure $P_0(x)$ for constants a and b, i.e., $(v-u)T(x) = \log a/b$. Thus,

$$\mathbb{E}[T(X)|H_0] = \frac{1}{(v-u)}\log a/b = \mathbb{E}[T(X)|H_1], \tag{11}$$

where the last step holds because $P_1(x)$ and $P_0(x)$ are both non-zero everywhere (absolutely continuous with respect to each other). But, this is a contradiction since $\mathbb{E}[T(X)|H_0] < 0 < \mathbb{E}[T(X)|H_1]$. Thus, $\Lambda_0(u)$ is strictly convex. A similar proof can be done for $\Lambda_1(u)$.

For proving the next claim, consider the derivative of $\Lambda_0(u)$.

$$\frac{d\Lambda_0(u)}{du} = \frac{\mathbb{E}[e^{uT(X)}T(X)|H_0]}{e^{\Lambda_0(u)}}. \tag{12}$$

The derivative of $\Lambda_0(u)$ at $u = 0$ is given by $\mathbb{E}[T(X)|H_0]$ which is strictly less than 0. Because $\Lambda_0(u)$ is strictly convex in u and $\Lambda_0(0) = 0$, if $\frac{d\Lambda_0(u)}{du}|_{u=0} < 0$, then $\Lambda_0(u) > 0$ for all $u < 0$.

A similar proof holds for the last claim as well, since the derivative of $\Lambda_1(u)$ at $u = 0$ is given by $\mathbb{E}[T(X)|H_1]$ which is strictly greater than 0, and $\Lambda_1(0) = 0$.

□

Next, we examine the properties of the log-generating functions for likelihood ratio detectors. Consider the likelihood ratio detector $T_0(x) = \log\frac{P_1(x)}{P_0(x)}$. The two conditions $\mathbb{E}[T(X)|H_0] < 0$ and $\mathbb{E}[T(X)|H_1] > 0$ become equivalent to $\mathrm{D}(P_0||P_1) > 0$ and $\mathrm{D}(P_1||P_0) > 0$ where $\mathrm{D}(\cdot||\cdot)$ denotes the Kullback-Leibler (KL) divergence between the two distributions $P_0(x)$ and $P_1(x)$. Thus, a likelihood ratio detector always satisfies these conditions as long as the KL divergences are well-defined and non-zero.

Property 4. *(Log-generating functions of likelihood ratio detectors) Let* $T_0(x) = \log\frac{P_1(x)}{P_0(x)}$*, and* $P_0(x)$ *and* $P_1(x)$ *be non-zero for all* x *with* $\mathrm{D}(P_0||P_1)$ *and* $\mathrm{D}(P_1||P_0)$ *strictly greater than* 0*. Then, the following properties hold:*

- $\Lambda_0(u)$ *is* 0 *at* $u = 0$ *and* 1*, and* $\Lambda_1(u)$ *is* 0 *at* $u = 0$ *and* -1.
- $\Lambda_1(u) = \Lambda_0(u+1)$.
- $\mathrm{C}(P_0, P_1) > 0$.
- $\Lambda_0(u)$ *and* $\Lambda_1(u)$ *are continuous, differentiable and strictly convex.*
- *The derivatives of* $\Lambda_0(u)$ *and* $\Lambda_1(u)$ *are continuous, monotonically increasing and take all values between* $-\infty$ *and* ∞.
- $\Lambda_0(u)$ *attains its global minima for* u *in* $(0, 1)$.
- $\Lambda_1(u)$ *attains its global minima for* u *in* $(-1, 0)$.

We first introduce the arithmetic mean-geometric mean (AM-GM) inequality.

Lemma 6 (AM-GM inequality)**.** *The following inequality is satisfied for $u \in (0,1)$ and $a, b \geq 0$:*

$$a^{1-u}b^u \leq (1-u)a + ub, \tag{13}$$

where the equality holds if and only if $a = b$.

Proof of Property 4. The first claim can be verified by direct substitution.

To show that $\Lambda_1(u) = \Lambda_0(u+1)$, observe that,

$$\Lambda_1(u) = -\log \sum_x P_1(x)^{1+u} P_0(x)^u = -\log \sum_x P_1(x)^{1+u} P_0(x)^{1-(1+u)} = \Lambda_0(u+1).$$

Next, we will show that $\mathrm{C}(P_0, P_1) > 0$. Observe that, $\mathrm{C}(P_0, P_1) = -\log \sum_x P_0(x)^{1-u^*} P_1(x)^{u^*}$ for some $u^* \in (0,1)$. Now, there is at least one x' with $P_0(x') > 0$ and $P_1(x') > 0$ such that $P_0(x') \neq P_1(x')$ because $\mathrm{D}(P_0||P_1) > 0$ and $\mathrm{D}(P_1||P_0) > 0$. This leads to a strict AM-GM inequality (Lemma 6) as follows:

$$P_0(x')^{1-u^*} P_1(x')^{u^*} < (1-u^*)P_0(x') + u^* P_1(x').$$

For all other $x \neq x'$,

$$P_0(x)^{1-u^*} P_1(x)^{u^*} \leq (1-u^*)P_0(x) + u^* P_1(x).$$

Thus,

$$\begin{aligned} &\sum_x P_0(x)^{1-u^*} P_1(x)^{u^*} < \sum_x (1-u^*)P_0(x) + u^* P_1(x) = 1 \\ &\implies -\log \sum_x P_0(x)^{1-u^*} P_1(x)^{u^*} > 0. \end{aligned} \tag{14}$$

Thus, $\mathrm{C}(P_0, P_1) > 0$. A similar proof extends for continuous distributions as well where the strict inequality holds at least over a set of x's that is not measure 0.

We move on to the next claim. Since both $P_0(x)$ and $P_1(x)$ are strictly greater than 0 for all x, we have $P_0(x)^{1-u} P_1(x)^u$ to be well-defined and continuous for all values of u, including $u = 0$ and $u = 1$. Thus, $\Lambda_0(u)$ is continuous over the range $(-\infty, \infty)$.

The derivative of $\Lambda_0(u)$ is given by:

$$\frac{d\Lambda_0(u)}{du} = \frac{\sum_x P_0(x)^{1-u} P_1(x)^u \log \frac{P_1(x)}{P_0(x)}}{e^{\Lambda_0(u)}}, \tag{15}$$

which is well-defined for all values of u.

The strict convexity of $\Lambda_0(u)$ can be proved using Property 3, because the two conditions $\mathbb{E}[T(X)|H_0] < 0$ and $\mathbb{E}[T(X)|H_1] > 0$ become equivalent to $\mathrm{D}(P_0||P_1) > 0$ and $\mathrm{D}(P_1||P_0) > 0$. A similar proof extends to $\Lambda_1(u)$.

Now, we move on to the next claim. Observe from (15) that, the derivative is also continuous for all values of u since both $P_0(x)$ and $P_1(x)$ are strictly greater than 0 for all x. It is monotonically increasing because $\Lambda_0(u)$ is strictly convex. Also note that, as $u \to -\infty$, its derivative tends to $-\infty$. Similarly, as $u \to \infty$, its derivative tends to ∞. A similar proof extends to $\Lambda_1(u)$.

Lastly, because $\Lambda_0(u)$ is 0 at $u = 0$ and $u = 1$, and is a continuous and strictly convex function, it attains its minima for u in $(0,1)$. A similar proof extends to $\Lambda_1(u)$, validating the last claim as well. □

Property 5 (Connection to FL transforms)**.** *For well-behaved detectors, the following properties hold:*

- *If $\tau < \mathbb{E}[T(X)|H_1]$, then* $\sup_{u<0} (u\tau - \Lambda_1(u)) = \sup_{u \in \mathbb{R}} (u\tau - \Lambda_1(u))$.
- *If $\tau > \mathbb{E}[T(X)|H_0]$, then* $\sup_{u>0} (u\tau - \Lambda_0(u)) = \sup_{u \in \mathbb{R}} (u\tau - \Lambda_0(u))$.

Before the proof, we introduce a lemma that will be used in the proof.

Lemma 7 (Supporting line of a strictly convex function)**.** *For a strictly convex and differentiable function $f(u):\mathcal{R}\rightarrow\mathcal{R}$,*

$$u_a \frac{df(u)}{du}|_{u=u_a} - f(u_a) = \sup_{u\in\mathcal{R}} \left(u\frac{df(u)}{du}|_{u=u_a} - f(u)\right).$$

The proof of Lemma 7 holds from the definition of strict convexity.

Proof of Property 5. In general, $\sup_{u\in\mathcal{R}}(u\tau-\Lambda_1(u)) \geq \sup_{u<0}(u\tau-\Lambda_1(u))$. But, here again,

$$\begin{aligned}\sup_{u\in\mathcal{R}}(u\tau-\Lambda_1(u)) &\overset{(a)}{=} \sup_{u\in\mathcal{R}}\left(u\frac{d\Lambda_1(u)}{du}|_{u=u_a} - \Lambda_1(u)\right) \overset{(b)}{=} u_a\frac{d\Lambda_1(u)}{du}|_{u=u_a} - \Lambda_1(u_a) \\ &\overset{(c)}{\leq} \sup_{u<0}\left(u\frac{d\Lambda_1(u)}{du}|_{u=u_a} - \Lambda_1(u)\right) \overset{(d)}{=} \sup_{u<0}(u\tau-\Lambda_1(u)). \end{aligned} \tag{16}$$

Here (a) holds because the derivative of $\Lambda_1(u)$ is continuous, monotonically increasing and takes all values from $(-\infty,\infty)$ (see Property 4). Thus, for any τ, there exists a single u_a such that $\frac{d\Lambda_1(u)}{du}|_{u=u_a}=\tau$. Next, (b) holds from Lemma 7, whereas (c) holds because $\frac{d\Lambda_1(u)}{du}|_{u=u_a}=\tau<\mathbb{E}[T(X)|H_1]=\frac{d\Lambda_1(u)}{du}|_{u=0}$ and the derivative is monotonically increasing (see Property 4) implying $u_a<0$. Lastly (d) holds by again substituting $\tau=\frac{d\Lambda_1(u)}{du}|_{u=u_a}$. This proves the first claim.

Similarly, in general, we have $\sup_{u\in\mathcal{R}}(u\tau-\Lambda_0(u)) \geq \sup_{u>0}(u\tau-\Lambda_0(u))$. But, here again,

$$\begin{aligned}\sup_{u\in\mathcal{R}}(u\tau-\Lambda_0(u)) &\overset{(a)}{=} \sup_{u\in\mathcal{R}}\left(u\frac{d\Lambda_0(u)}{du}|_{u=u_a} - \Lambda_0(u)\right) \overset{(b)}{=} u_a\frac{d\Lambda_0(u)}{du}|_{u=u_a} - \Lambda_0(u_a) \\ &\overset{(c)}{\leq} \sup_{u>0}\left(u\frac{d\Lambda_0(u)}{du}|_{u=u_a} - \Lambda_0(u)\right) \overset{(d)}{=} \sup_{u>0}(u\tau-\Lambda_0(u)). \end{aligned} \tag{17}$$

Here (a) holds because the derivative of $\Lambda_0(u)$ is continuous, monotonically increasing and takes all values from $(-\infty,\infty)$ (see Property 4). Thus, for any τ, there exists a single u_a such that $\frac{d\Lambda_0(u)}{du}|_{u=u_a}=\tau$. Next, (b) holds from Lemma 7, whereas (c) holds because $\frac{d\Lambda_0(u)}{du}|_{u=u_a}=\tau>\mathbb{E}[T(X)|H_0]=\frac{d\Lambda_0(u)}{du}|_{u=0}$ and the derivative is monotonically increasing (see Property 4) implying $u_a>0$. Lastly (d) holds by again substituting $\tau=\frac{d\Lambda_0(u)}{du}|_{u=u_a}$.

□

A.3. Log Generating Functions for Gaussians

Let $P_0(x)\sim\mathcal{N}(\mu_0,\sigma^2\mathbf{I})$ and $P_1(x)\sim\mathcal{N}(\mu_1,\sigma^2\mathbf{I})$, where μ_0 and μ_1 are vectors and $\mathbf{I}$ is an identity matrix. We derive the log-generating functions for likelihood ratio detectors corresponding to these two distributions.

$$\begin{aligned}\Lambda_0(u) = \log\int P_1(x)^u P_0(x)^{1-u}dx &= \log\int e^{\frac{-u}{2\sigma^2}((x-\mu_1)^T(x-\mu_1)-(x-\mu_0)^T(x-\mu_0))}P_0(x)dx \\ &= \log e^{\frac{-u}{2\sigma^2}(\mu_1^T\mu_1-\mu_0^T\mu_0)}\int e^{\frac{-u}{2\sigma^2}(-2x^T(\mu_1-\mu_0)))}P_0(x)dx \\ &\overset{(a)}{=} \log e^{\frac{-u}{2\sigma^2}(\mu_1^T\mu_1-\mu_0^T\mu_0)} e^{\frac{-u}{2\sigma^2}(-2\mu_0^T(\mu_1-\mu_0)))} e^{\frac{u^2}{2\sigma^2}(||\mu_1-\mu_0||_2^2))} \\ &= \log e^{\frac{-u}{2\sigma^2}(||\mu_1-\mu_0||_2^2)} e^{\frac{u^2}{2\sigma^2}(||\mu_1-\mu_0||_2^2))} \\ &= \frac{1}{2\sigma^2}||\mu_1-\mu_0||_2^2 u(u-1), \end{aligned} \tag{18}$$

where (a) is derived using the expression of the moment generating function of a Gaussian distribution.

A.4. Proof of Lemma 2

Proof of Lemma 2. Under equal priors $\pi_0 = \pi_1 = \frac{1}{2}$, the detector that minimizes the Bayesian probability of error, i.e., $P_{e,T}(\tau) = \pi_0 P_{\text{FP},T}(\tau) + \pi_1 P_{\text{FN},T}(\tau)$ is the likelihood ratio detector given by $T(x) = \log \frac{P_1(x)}{P_0(x)} \geq 0$ (for $\pi_0 = \pi_1 = \frac{1}{2}$). The proof is available in Theorem 3.1 of (Gallager, 2012).

Here, we will show that the Chernoff exponent of the probability of error for this detector, i.e., $E_{e,T}(0)$ is equal to $\text{C}(P_0, P_1) = -\min_{u\in(0,1)} \log \sum_x P_0(x)^{(1-u)} P_1(x)^u$.

Note that,

$$E_{\text{FP},T}(0) = \sup_{u>0} -\Lambda_0(u) = -\min_{u\in(0,1)} \log \sum_x P_0(x)^{(1-u)} P_1(x)^u, \tag{19}$$

where the last step follows because $\Lambda_0(u)$ attains its minima in the range $u \in (0,1)$ (see Property 4).

$$\begin{aligned} E_{\text{FN},T}(0) = \sup_{u<0} -\Lambda_1(u) &\overset{(a)}{=} -\min_{u\in(-1,0)} \log \sum_x P_0(x)^{(-u)} P_1(x)^{(1+u)} \\ &= -\min_{u'=u+1\in(0,1)} \log \sum_x P_0(x)^{(1-u')} P_1(x)^{(u')}, \end{aligned} \tag{20}$$

where (a) also holds because $\Lambda_1(u)$ attains its minima in the range $u \in (-1,0)$ (see Property 4). Lastly,

$$E_{e,T}(0) = \min\{E_{\text{FP},T}(0), E_{\text{FN},T}(0)\} = \text{C}(P_0, P_1). \tag{21}$$

□

B. Appendix to Section 3.1

Before the proofs, we introduce a lemma that will be used in the proofs.

Lemma 8. *Let $P_0(x)$ and $P_1(x)$ be non-zero for all x and $\text{D}(P_0||P_1)$ and $\text{D}(P_1||P_0)$ be strictly greater than 0. For likelihood ratio detectors of the form $T_0(x) = \log \frac{P_1(x)}{P_0(x)} \geq \tau_0$, if $\tau_0 \neq 0$, then one of the following statements is true:*

$$E_{\text{FN},T_0}(\tau_0) < \text{C}(P_0,P_1) < E_{\text{FP},T_0}(\tau_0), \textit{or } E_{\text{FP},T_0}(\tau_0) < \text{C}(P_0,P_1) < E_{\text{FN},T_0}(\tau_0).$$

Proof of Lemma 8. Let us analyze the scenario where $\tau_0 > 0$. Observe that,

$$\begin{aligned} E_{\text{FP},T_0}(\tau_0) = \sup_{u>0}(u\tau_0 - \Lambda_0(u)) &\geq u_0^*\tau_0 - \Lambda_0(u_0^*) && [\text{for any } u_0^* > 0] \\ &> -\Lambda_0(u_0^*) && [\text{since } u_0^*\tau_0 > 0] \\ &\overset{(a)}{=} \text{C}(P_0,P_1), \end{aligned} \tag{22}$$

where (a) follows if we choose $u_0^* = \arg\min \Lambda_0(u)$ (from Property 4, $\Lambda_0(u)$ attains its minima for some $u \in (0,1)$) and $\Lambda_0(u_0^*) = -\text{C}(P_0,P_1)$ (by definition).

Now, we will show that $E_{\text{FN},T_0}(\tau_0) < \text{C}(P_0,P_1)$ when $\tau_0 > 0$.

Case 1: $\tau_0 \geq \frac{d\Lambda_1(u)}{du}|_{u=0} = \text{D}(P_1||P_0)$

$$\begin{aligned} E_{\text{FN},T_0}(\tau_0) = \sup_{u<0}(u\tau_0 - \Lambda_1(u)) &\leq \sup_{u<0}(u\text{D}(P_1||P_0) - \Lambda_1(u)) \; [\text{since } \tau_0 \geq \text{D}(P_1||P_0)] \\ &\leq \sup_{u\in\mathcal{R}}(u\text{D}(P_1||P_0) - \Lambda_1(u)) \\ &\overset{(a)}{=} (0\cdot\text{D}(P_1||P_0) - \Lambda_1(0)) \overset{(b)}{=} 0 \overset{(c)}{<} \text{C}(P_0,P_1), \end{aligned} \tag{23}$$

where (a) holds from Lemma 7 because $\frac{d\Lambda_1(u)}{du}|_{u=0} = \mathrm{D}(P_1||P_0)$, and (b) and (c) hold from Property 4 since $\Lambda_1(0) = 0$ and $\mathrm{C}(P_0, P_1) > 0$.

Case 2: $0 < \tau_0 < \frac{d\Lambda_1(u)}{du}|_{u=0} = \mathrm{D}(P_1||P_0)$

$$\begin{aligned} E_{\mathrm{FN},T_0}(\tau_0) = \sup_{u<0}(u\tau_0 - \Lambda_1(u)) &\leq \sup_{u\in\mathcal{R}}(u\tau_0 - \Lambda_1(u)) \\ &\overset{(a)}{=} \sup_{u\in\mathcal{R}}(u\tau_0 - \Lambda_1(u)) \;\; [\text{where } \frac{d\Lambda_1(u)}{du}|_{u=u_a} = \tau_0] \\ &\overset{(b)}{=} u_a\tau_0 - \Lambda_1(u_a) \\ &\overset{(c)}{<} -\Lambda_1(u_a) \quad [\text{since } u_a\tau_0 < 0] \\ &\leq -\min_u \Lambda_1(u) \\ &\overset{(d)}{=} -\min_{u\in(-1,0)} \Lambda_1(u) = \mathrm{C}(P_0, P_1) \end{aligned} \tag{24}$$

Here, (a) holds because the derivative of $\Lambda_1(u)$ is continuous, monotonically increasing and takes all values from $-\infty$ to ∞ (see Property 4). Thus, for any τ_0, there exists a single u_a such that $\frac{d\Lambda_1(u)}{du}|_{u=u_a} = \tau_0$. Next, (b) holds from Lemma 7, (c) holds because $\frac{d\Lambda_1(u)}{du}|_{u=u_a} = \tau_0 < \frac{d\Lambda_1(u)}{du}|_{u=0}$, and the derivative is monotonically increasing, implying $u_a < 0$. Lastly (d) holds because $\Lambda_1(u)$ attains its minima in the range $u \in (-1, 0)$ (see Property 4).

Thus, for $\tau_0 > 0$, we get $E_{\mathrm{FN},T_0}(\tau_0) < \mathrm{C}(P_0, P_1) < E_{\mathrm{FP},T_0}(\tau_0)$.

The proof is similar for the scenario where $\tau_0 < 0$, and leads to $E_{\mathrm{FP},T_0}(\tau_0) < \mathrm{C}(P_0, P_1) < E_{\mathrm{FN},T_0}(\tau_0)$.

□

B.1. Proof of Lemma 3

Proof of Lemma 3. Suppose there exists two likelihood ratio detectors for the two groups such that, $E_{\mathrm{FN},T_0}(\tau_0) = E_{\mathrm{FN},T_1}(\tau_1)$. Since $\mathrm{C}(P_0, P_1) < \mathrm{C}(Q_0, Q_1)$, *at most* one of the two exponents $E_{\mathrm{FN},T_0}(\tau_0)$ and $E_{\mathrm{FN},T_1}(\tau_1)$ can be equal to their corresponding Chernoff information $\mathrm{C}(P_0, P_1)$ or $\mathrm{C}(Q_0, Q_1)$. Without loss of generality, we may assume that $E_{\mathrm{FN},T_0}(\tau_0) \neq \mathrm{C}(P_0, P_1)$. This implies that $\tau_0 \neq 0$ because in the proof of Lemma 2, we already showed that when $\tau_0 = 0$, we always have $E_{\mathrm{FN},T_0}(0) = E_{\mathrm{FP},T_0}(0) = \mathrm{C}(P_0, P_1)$. Since $\tau_0 \neq 0$, using Lemma 8, we either have $E_{\mathrm{FN},T_0}(\tau_0) < \mathrm{C}(P_0, P_1) < E_{\mathrm{FP},T_0}(\tau_0)$ or $E_{\mathrm{FP},T_0}(\tau_0) < \mathrm{C}(P_0, P_1) < E_{\mathrm{FN},T_0}(\tau_0)$. Thus,

$$E_{e,T_0}(\tau_0) = \min\{E_{\mathrm{FP},T_0}(\tau_0), E_{\mathrm{FN},T_0}(\tau_0)\} < \mathrm{C}(P_0, P_1). \tag{25}$$

□

B.2. Proof of Theorem 1

Proof of Theorem 1. The first claim follows directly from Lemma 2 by choosing the likelihood ratio detectors for the two groups with thresholds $\tau_0 = \tau_1 = 0$, i.e., the Bayes optimal detector under equal priors.

Now, we prove the second claim. Suppose that we choose the Bayes optimal classifiers $T_0(x) \geq \tau_0$ and $T_1(x) \geq \tau_1$ for the two groups. Then, we have $E_{\mathrm{FN},T_0}(\tau_0) = C(P_0, P_1)$ and $E_{\mathrm{FN},T_1}(\tau_1) = C(Q_0, Q_1)$ which are not equal. Thus, $|E_{\mathrm{FN},T_0}(\tau_0) - E_{\mathrm{FN},T_1}(\tau_1)| \neq 0$.

Assume (for the sake of contradiction) that there is a likelihood ratio detector such that $E_{e,T_0}(\tau_0) > \mathrm{C}(P_0, P_1)$.

Now, if $\tau_0 = 0$, then we have $E_{e,T_0}(\tau_0) = \mathrm{C}(P_0, P_1)$ (from Lemma 2). Alternately, if $\tau_0 \neq 0$, then we either have $E_{\mathrm{FN},T_0}(\tau_0) < \mathrm{C}(P_0, P_1) < E_{\mathrm{FP},T_0}(\tau_0)$ or $E_{\mathrm{FP},T_0}(\tau_0) < \mathrm{C}(P_0, P_1) < E_{\mathrm{FN},T_0}(\tau_0)$ (from Lemma 8). Thus,

$$E_{e,T_0}(\tau_0) = \min\{E_{\mathrm{FP},T_0}(\tau_0), E_{\mathrm{FN},T_0}(\tau_0)\} < \mathrm{C}(P_0, P_1). \tag{26}$$

For both cases, we have a contradiction, implying that $E_{e,T_0}(\tau_0) \leq \mathrm{C}(P_0, P_1) < \mathrm{C}(Q_0, Q_1)$ for all likelihood ratio detectors. □

B.3. Proofs of Lemma 4 and Lemma 5

Proof of Lemma 4. Let $\tau_0^*=0$. Using Lemma 2, this ensures,

$$E_{\mathrm{FN},T_0}(0)=E_{\mathrm{FP},T_0}(0)=\mathrm{C}(P_0,P_1).$$

Now, we will show that the only value of τ_1^* that will satisfy $E_{\mathrm{FN},T_1}(\tau_1^*)=E_{\mathrm{FN},T_0}(0)$ is a $\tau_1^*{>}0$ such that $E_{\mathrm{FN},T_1}(\tau_1^*)=\mathrm{C}(P_0,P_1)$. To prove that such a τ_1^* exists, consider the function:

$$g(u)=u\frac{d\Lambda_1(u)}{d(u)}-\Lambda_1(u),$$

where $\Lambda_1(u)$ is the log-generating transform for $z=1$. The function $g(u)$ is continuous. At $u=0$, $g(u)=0$ and at $u=u_1^*$ (where $u_1^*=\arg\min\Lambda_1(u)$ and lies in $(-1,0)$ from Property 4) we have $g(u)=\mathrm{C}(Q_0,Q_1)$. Because $g(u)$ is continuous, there exists a $u_a\in(u_1^*,0)$ such that $g(u_a)=\mathrm{C}(P_0,P_1)$ which lies between 0 and $\mathrm{C}(Q_0,Q_1)$. If we set $\tau_1^*=\frac{d\Lambda_1(u)}{d(u)}|_{u=u_a}$, we have,

$$\mathrm{C}(P_0,P_1)=g(u_a)\overset{\text{Lemma 7}}{=}\sup_{u\in\mathcal{R}}(u\tau_1^*-\Lambda_1(u)).$$

Now, in general, $\sup_{u<0}(u\tau_1^*-\Lambda_1(u))\leq\sup_{u\in\mathcal{R}}(u\tau_1^*-\Lambda_1(u))=g(u_a)$. But again, $\sup_{u<0}(u\tau_1^*-\Lambda_1(u))\geq u_a\tau_1^*-\Lambda_1(u_a)=g(u_a)$ since $u_a\in(u_1^*,0)$. Thus,

$$E_{\mathrm{FN},T_0}(\tau_1^*)=\sup_{u<0}(u\tau_1^*-\Lambda_1(u))=g(u_a)=\mathrm{C}(P_0,P_1).$$

Also note that $\tau_1^*>0$ because the derivative of $\Lambda_1(u)$ is monotonically increasing and $u_a>u_1^*$, leading to $\tau_1^*=\frac{d\Lambda_1(u)}{d(u)}|_{u=u_a}>\frac{d\Lambda_1(u)}{d(u)}|_{u=u_1^*}=0$.

Now that we have a τ_1^* such that $E_{\mathrm{FN},T_1}(\tau_1^*)=\mathrm{C}(P_0,P_1)$ which is strictly less that $\mathrm{C}(Q_0,Q_1)$, we must have $E_{\mathrm{FP},T_1}(\tau_1^*)>\mathrm{C}(Q_0,Q_1)$ (from Lemma 8).

This leads to,

$$\min\{E_{\mathrm{FP},T_0}(0),E_{\mathrm{FN},T_0}(0),E_{\mathrm{FP},T_1}(\tau_1^*),E_{\mathrm{FN},T_1}(\tau_1^*)\}{=}\mathrm{C}(P_0,P_1).$$

For any other choice of $\tau_0^*\neq0$, we either have $E_{\mathrm{FP},T_0}(\tau_0^*)<\mathrm{C}(P_0,P_1)<E_{\mathrm{FN},T_0}(\tau_0^*)$, or $E_{\mathrm{FN},T_0}(\tau_0^*)<\mathrm{C}(P_0,P_1)<E_{\mathrm{FP},T_0}(\tau_0^*)$, implying

$$\min\{E_{\mathrm{FP},T_0}(\tau_0^*),E_{\mathrm{FN},T_0}(\tau_0^*),E_{\mathrm{FP},T_1}(\tau_1^*),E_{\mathrm{FN},T_1}(\tau_1^*)\}{<}\mathrm{C}(P_0,P_1).$$

□

Proof of Lemma 5. We are given that,

$$E_{\mathrm{FN},T_1}(\tau_1)=E_{\mathrm{FP},T_1}(\tau_1)=\mathrm{C}(Q_0,Q_1).$$

Now, we will show that the only value of τ_0^* that will satisfy $E_{\mathrm{FN},T_0}(\tau_0^*)=\mathrm{C}(Q_0,Q_1)$ is a $\tau_0^*<0$. To prove that such a τ_0^* exists, consider the function

$$g(u)=u\frac{d\Lambda_1(u)}{d(u)}-\Lambda_1(u),$$

where $\Lambda_1(u)$ is the log-generating transform for $z=0$. The function $g(u)$ is continuous. At $u=u_1^*$ (where $u_1^*=\arg\min\Lambda_1(u)$ and lies in $(-1,0)$ from Property 4), we have $g(u_1^*)=\mathrm{C}(P_0,P_1)$ and as $u\to-\infty$, we have $g(u)\to\infty$. Because $g(u)$ is continuous, there exists a $u_a\in(-\infty,u_1^*)$ such that $g(u_a)=\mathrm{C}(Q_0,Q_1)$ which lies between $\mathrm{C}(P_0,P_1)$ and ∞. If we set $\tau_0^*=\frac{d\Lambda_1(u)}{d(u)}|_{u=u_a}$, we have,

$$\mathrm{C}(Q_0,Q_1)=g(u_a)\overset{\text{Lemma 7}}{=}\sup_{u\in\mathcal{R}}(u\tau_0^*-\Lambda_1(u)).$$

Now, in general, $\sup_{u<0}(u\tau_0^* - \Lambda_1(u)) \leq \sup_{u\in\mathcal{R}}(u\tau_0^* - \Lambda_1(u)) = g(u_a)$. But again, $\sup_{u<0}(u\tau_0^* - \Lambda_1(u)) \geq u_a\tau_0^* - \Lambda_1(u_a) = g(u_a)$ since $u_a < u_1^* < 0$. Thus,

$$E_{\text{FN},T_0}(\tau_0^*) = \sup_{u<0}(u\tau_0^* - \Lambda_1(u)) = g(u_a) = \mathrm{C}(Q_0, Q_1).$$

This τ_0^* is less than 0 because the derivative of $\Lambda_1(u)$ is monotonically increasing and $u_a < u_1^*$, leading to $\tau_0^* = \frac{\Lambda_1(u)}{d(u)}|_{u=u_a} < \frac{\Lambda_1(u)}{d(u)}|_{u=u_1^*} = 0$.

Now that we have a τ_0^* such that $E_{\text{FN},T_0}(\tau_0^*) = \mathrm{C}(Q_0, Q_1)$ which is strictly greater that $\mathrm{C}(P_0, P_1)$, we must have $E_{\text{FP},T_0}(\tau_0^*) < \mathrm{C}(P_0, P_1)$ (from Lemma 8).

This leads to,

$$\min\{E_{\text{FP},T_0}(\tau_0^*), E_{\text{FN},T_0}(\tau_0^*)\} < \mathrm{C}(P_0, P_1).$$

□

C. Appendix to Section 3.2

Proof of Theorem 2. From Lemma 5, there exists a likelihood ratio detector of the form $T_0(x) = \log \frac{P_1(x)}{P_0(x)} \geq \tau_0^*$ such that

$$E_{\text{FN},T_0}(\tau_0^*) = \mathrm{C}(Q_0, Q_1). \tag{27}$$

In the proof of Lemma 5, we showed that this $\tau_0^* < 0$.

Now, we will show that there exists $\widetilde{P}_0(x)$ and $\widetilde{P}_1(x)$ such that their optimal detector $\widetilde{T_0}(x) = \log \frac{\widetilde{P}_1(x)}{\widetilde{P}_0(x)} \geq 0$ is equivalent to the detector $T_0(x) \geq \tau_0^*$.

Let $\widetilde{P}_0(x) = \frac{P_0(x)^{(1-w)}P_1(x)^w}{\sum_x P_0(x)^{(1-w)}P_1(x)^w}$ and $\widetilde{P}_1(x) = \frac{P_0(x)^{(1-v)}P_1(x)^v}{\sum_x P_0(x)^{(1-v)}P_1(x)^v}$ for some $w, v \in \mathcal{R}$ with $w \neq v$. Observe that,

$$\begin{aligned}
\widetilde{T_0}(x) = \log \frac{\widetilde{P}_1(x)}{\widetilde{P}_0(x)} &= (v-w)\log\frac{P_1(x)}{P_0(x)} + \log\frac{\sum_x P_0(x)^{(1-w)}P_1(x)^w}{\sum_x P_0(x)^{(1-v)}P_1(x)^v} \\
&= (v-w)\log\frac{P_1(x)}{P_0(x)} + \Lambda_0(w) - \Lambda_0(v) \\
&= (v-w)\left(\log\frac{P_1(x)}{P_0(x)} - \frac{\Lambda_0(v)-\Lambda_0(w)}{v-w}\right).
\end{aligned} \tag{28}$$

Because $\Lambda_0(u)$ is strictly convex with its derivative taking all values from $-\infty$ to ∞, one can always find a tangent to $\Lambda_0(u)$ that has a slope τ_0^* at (say) $u = u_a$. Thus, one can always find pairs of points (w, v) on either sides of $u = u_a$ such that $\tau_0^* = \frac{\Lambda_0(v)-\Lambda_0(w)}{v-w}$, which are essentially pairs of points (w, v) at which a straight line with slope τ_0^* cuts $\Lambda_0(u)$. In particular, we can fix $v = 1$ and always find a $w < 0$ such that

$$\tau_0^* = \frac{\Lambda_0(v)-\Lambda_0(w)}{v-w} = \frac{-\Lambda_0(w)}{1-w}, \tag{29}$$

because $\Lambda_0(u)$ is continuous taking values 0 at $u = 0$ and $u = 1$, and takes all values from $(0, \infty)$ in the range $(-\infty, 0)$. Thus, the first claim is proved.

Now, we calculate $\mathrm{C}(\widetilde{P}_0, \widetilde{P}_1)$.

$$\begin{aligned}
\mathrm{C}(\widetilde{P}_0, \widetilde{P}_1) = \max_{u\in(0,1)} -\log\sum_x \widetilde{P}_0(x)^{1-u}\widetilde{P}_1(x)^u &\overset{(a)}{=} \max_{u\in\mathcal{R}} -\log\sum_x \widetilde{P}_0(x)^{1-u}\widetilde{P}_1(x)^u \\
&\overset{(b)}{=} \max_{u\in\mathcal{R}} -\log\sum_x P_0(x)^{(1-w)(1-u)}P_1(x)^{w(1-u)+u} + (1-u)\Lambda_0(w)
\end{aligned}$$

$$\begin{aligned}
&\overset{(c)}{=} \max_{u\in\mathcal{R}} -\log \sum_x P_0(x)^{(1-w)(1-u)} P_1(x)^{w(1-u)+u} + (1-u)(w-1)\tau_0^* \\
&\overset{(d)}{=} \max_{u\in\mathcal{R}} (1-u)(w-1)\tau_0^* - \Lambda_1((1-u)(w-1)) \\
&\overset{(e)}{=} \sup_{u'\in\mathcal{R}} (u'\tau_0^* - \Lambda_1(u')) \;\; [u' = (1-u)(w-1)] \\
&\overset{(f)}{=} \sup_{u'<0} (u'\tau_0^* - \Lambda_1(u')) \;\; [u' = (1-u)(w-1)] \\
&\overset{(g)}{=} \mathrm{C}(Q_0, Q_1).
\end{aligned} \tag{30}$$

Here (a) holds because the log-generating function $-\log \sum_x \widetilde{P}_0(x)^{1-u}\widetilde{P}_1(x)^u$ of a likelihood ratio detector attains its global minima at $(0,1)$ (see Property 4) and (b) holds by substituting $\widetilde{P}_0(x) = \frac{P_0(x)^{(1-w)}P_1(x)^w}{\sum_x P_0(x)^{(1-w)}P_1(x)^w}$ and $\widetilde{P}_1(x) = \frac{P_0(x)^{(1-v)}P_1(x)^v}{\sum_x P_0(x)^{(1-v)}P_1(x)^v}$ with $v = 1$. Next, (c) holds by using $\tau_0^* = \frac{\Lambda_0(v)-\Lambda_0(w)}{v-w} = \frac{-\Lambda_0(w)}{1-w}$ (see (29)), (d) holds from the definition of $\Lambda_1((1-u)(w-1))$, (e) holds by a change of variable $u' = (1-u)(w-1)$, (f) holds because $\tau_0^* < 0 \leq \mathrm{D}(\widetilde{P}_1||\widetilde{P}_0) = \mathbb{E}[\widetilde{T_0}(X)|\widetilde{H_1}]$ and the detector is well-behaved (see Property 5), and lastly (g) holds because $E_{\mathrm{FN},T_0}(\tau_0^*) = \mathrm{C}(Q_0, Q_1)$ (see (27)). □

D. Appendix to Section 3.3

D.1. Proof of Theorem 3

Proof of Theorem 3. We remind the readers that,

$$\frac{W_0(x,x')}{P_0(x)} = \Pr(X' = x'|X = x, Z = 0, Y = 0), \text{ and } \frac{W_1(x,x')}{P_1(x)} = \Pr(X' = x'|X = x, Z = 0, Y = 1). \tag{31}$$

First, we would like to prove: $I(X';Y|X, Z = 0) > 0 \implies \mathrm{C}(W_0, W_1) > \mathrm{C}(P_0, P_1)$.

Suppose that X' is not independent of Y given X and $Z = 0$, i.e., $I(X';Y|X, Z = 0) > 0$. This implies that there exists at least one $X = x_a$ such that the distributions of $X'|_{X=x_a,Z=0,Y=0}$ and $X'|_{X=x_a,Z=0,Y=1}$ are different. Therefore, there exists at least one pair $(x', x) = (x'_a, x_a)$ for which the following AM-GM inequality (Lemma 6) holds with strict inequality for all $u \in (0,1)$, i.e,

$$\left(\frac{W_0(x_a,x'_a)}{P_0(x_a)}\right)^{1-u}\left(\frac{W_1(x_a,x'_a)}{P_1(x_a)}\right)^{u} < (1-u)\frac{W_0(x_a,x'_a)}{P_0(x_a)} + u\frac{W_1(x_a,x'_a)}{P_1(x_a)}. \tag{32}$$

For all other $(x', x) \neq (x'_a, x_a)$, we have (from the AM-GM inequality in Lemma 6):

$$\left(\frac{W_0(x,x')}{P_0(x)}\right)^{1-u}\left(\frac{W_1(x,x')}{P_1(x)}\right)^{u} \leq (1-u)\frac{W_0(x,x')}{P_0(x)} + u\frac{W_1(x,x')}{P_1(x)}. \tag{33}$$

Using (32) and (33),

$$\sum_{x'}\left(\frac{W_0(x_a,x')}{P_0(x_a)}\right)^{1-u}\left(\frac{W_1(x_a,x')}{P_1(x_a)}\right)^{u} < \sum_{x'}\left((1-u)\frac{W_0(x_a,x')}{P_0(x_a)} + u\frac{W_1(x_a,x')}{P_1(x_a)}\right) = 1. \tag{34}$$

This leads to,

$$\sum_{x'} W_0(x_a,x')^{1-u}W_1(x_a,x')^u < P_0(x_a)^{1-u}P_1(x_a)^u. \tag{35}$$

For all other $x \neq x_a$, we have (using (33) alone),

$$\sum_{x'}\left(\frac{W_0(x,x')}{P_0(x)}\right)^{1-u}\left(\frac{W_1(x,x')}{P_1(x)}\right)^{u} \leq \sum_{x'}\left((1-u)\frac{W_0(x,x')}{P_0(x)} + u\frac{W_1(x,x')}{P_1(x)}\right) = 1, \tag{36}$$

leading to

$$\sum_{x'} W_0(x,x')^{1-u} W_1(x,x')^u \leq P_0(x)^{1-u} P_1(x)^u. \tag{37}$$

Lastly, using (35) and (37),

$$\sum_{x}\sum_{x'} W_0(x,x')^{1-u} W_1(x,x')^u < \sum_{x} P_0(x)^{1-u} P_1(x)^u, \tag{38}$$

leading to the claim:

$$\mathrm{C}(W_0,W_1) = -\min_{u\in(0,1)} \log \sum_{x}\sum_{x'} W_0(x,x')^{1-u} W_1(x,x')^u > -\min_{u\in(0,1)} \log \sum_{x} P_0(x)^{1-u} P_1(x)^u = \mathrm{C}(P_0,P_1). \tag{39}$$

We would now like to prove:
$\mathrm{C}(W_0,W_1) > \mathrm{C}(P_0,P_1) \implies I(X';Y|X,Z=0) > 0$, or, $I(X';Y|X,Z=0)=0 \Longrightarrow \mathrm{C}(W_0,W_1) \not> \mathrm{C}(P_0,P_1)$.

First note that, from the previous proof, $\mathrm{C}(W_0,W_1) \geq \mathrm{C}(P_0,P_1)$ always holds using the AM-GM inequality. Thus, $\mathrm{C}(W_0,W_1) \not> \mathrm{C}(P_0,P_1)$ is same as $\mathrm{C}(W_0,W_1) = \mathrm{C}(P_0,P_1)$.

Suppose that X' is independent of Y given X and $Z=0$, i.e., $I(X';Y|X,Z=0)=0$. This implies that,

$$\begin{aligned}
&\Pr(X'=x'|X,Z=0,Y=0) = \Pr(X'=x'|X,Z=0,Y=1)\ \forall x' \\
&\Rightarrow \frac{W_0(x,x')}{P_0(x)} = \frac{W_1(x,x')}{P_1(x)} \quad \forall x', x \\
&\Rightarrow \sum_{x'} \Big(\frac{W_0(x,x')}{P_0(x)}\Big)^{1-u} \Big(\frac{W_1(x,x')}{P_1(x)}\Big)^u = 1\ \forall x \\
&\Rightarrow \sum_{x}\sum_{x'} W_0(x,x')^{1-u} W_1(x,x')^u = \sum_{x} P_0(x)^{1-u} P_1(x)^u. && (40)
\end{aligned}$$

This leads to

$$\mathrm{C}(W_0,W_1) = -\min_{u\in(0,1)} \log \sum_{x}\sum_{x'} W_0(x,x')^{1-u} W_1(x,x')^u = -\min_{u\in(0,1)} \log \sum_{x} P_0(x)^{1-u} P_1(x)^u = \mathrm{C}(P_0,P_1). \tag{41}$$

□

E. Unequal Priors

E.1. Unequal Priors on Y but Equal Priors on Z

When the prior probabilities are unequal, we can write $P_{e,T_z}(\tau_z)$ as:

$$P_{e,T_z}(\tau_z) = \frac{1}{2}(2\pi_0 P_{\mathrm{FP},T_z}(\tau_z)) + \frac{1}{2}(2\pi_1 P_{\mathrm{FN},T_z}(\tau_z)),$$

and define the Chernoff exponent of $P_{e,T_z}(\tau_z)$, i.e., $E_{e,T_z}(\tau_z)$ more generally as follows:

$$\min\{E_{\mathrm{FP},T_z}(\tau_z) - \log 2\pi_0, E_{\mathrm{FN},T_z}(\tau_z) - \log 2\pi_1\}.$$

Lemma 9. *Let the absolute continuity and distinct hypotheses assumptions of Section 2 hold, and $T_z(x)$ be the likelihood ratio detector for the group $Z=z$. Then, the value of τ_z that maximizes $E_{e,T_z}(\tau_z)$, i.e.,*

$$\max_{\tau_z} \min\{E_{\mathrm{FP},T_z}(\tau_z) - \log 2\pi_0, E_{\mathrm{FN},T_z}(\tau_z) - \log 2\pi_1\},$$

is given by $\tau_z^* = \log\frac{\pi_0}{\pi_1}$, which is the same as the value of τ_z that minimizes $P_{e,T_z}(\tau_z)$, i.e.,

$$\min_{\tau_z} \pi_0 P_{\mathrm{FP},T_z}(\tau_z) + \pi_1 P_{\mathrm{FN},T_z}(\tau_z).$$

This likelihood ratio detector $T_z(x) \geq \log \frac{\pi_0}{\pi_1}$ is the Bayes optimal detector for the group.

Before we proceed to the proof, we discuss another result. Observe that,

$$u\tau_0 - \Lambda_0(u) - \log 2\pi_0 = u(\tau_0 - \log \frac{\pi_0}{\pi_1}) + u \log \frac{\pi_0}{\pi_1} - \Lambda_0(u) - \log 2\pi_0 = u\tau' - \widetilde{\Lambda}_0(u) - \log 2, \tag{42}$$

where $\tau' = \tau_0 - \log \frac{\pi_0}{\pi_1}$, and $\widetilde{\Lambda}_0(u) = \Lambda_0(u) - u \log \frac{\pi_0}{\pi_1} + \log \pi_0$. Similarly,

$$u\tau_0 - \Lambda_1(u) - \log 2\pi_1 = u(\tau_0 - \log \frac{\pi_0}{\pi_1}) + u \log \frac{\pi_0}{\pi_1} - \Lambda_1(u) - \log 2\pi_1 = u\tau' - \widetilde{\Lambda}_1(u) - \log 2, \tag{43}$$

where $\tau' = \tau_0 - \log \frac{\pi_0}{\pi_1}$, and $\widetilde{\Lambda}_1(u) = \Lambda_1(u) - u \log \frac{\pi_0}{\pi_1} + \log \pi_1$.

We first derive some properties of $\widetilde{\Lambda}_0(u)$ and $\widetilde{\Lambda}_1(u)$.

Lemma 10. *Let $P_0(x)$ and $P_1(x)$ be strictly greater than 0 everywhere and $\mathrm{D}(P_0||P_1)$ and $\mathrm{D}(P_1||P_0)$ be strictly greater than 0 and π_0 and π_1 lie in $(0,1)$. Then, the following properties hold:*

- *$\widetilde{\Lambda}_0(u)$ and $\widetilde{\Lambda}_1(u)$ are continuous, differentiable and strictly convex.*
- *The derivatives of $\widetilde{\Lambda}_0(u)$ and $\widetilde{\Lambda}_1(u)$ are continuous, monotonically increasing, and take all values from $-\infty$ to ∞.*
- *$\widetilde{\Lambda}_1(u) = \widetilde{\Lambda}_0(u+1)$.*

Proof of Lemma 10. Note that, $\widetilde{\Lambda}_0(u)$ is the sum of $\Lambda_0(u)$ and an affine function $-u \log \frac{\pi_0}{\pi_1} + \log \pi_0$. Because $\Lambda_0(u)$ is continuous, differentiable and strictly convex (from Property 4), $\widetilde{\Lambda}_0(u)$ also satisfies those properties. The second claim also holds for the same reason because the derivative of $\Lambda_0(u)$ satisfies all these properties (from Property 4).

Lastly,

$$\begin{aligned}\widetilde{\Lambda}_0(u+1) = \Lambda_0(u+1) - (u+1)\log \frac{\pi_0}{\pi_1} + \log \pi_0 &= \Lambda_0(u+1) - u \log \frac{\pi_0}{\pi_1} + \log \pi_1 \\ &\overset{(a)}{=} \Lambda_1(u) - u \log \frac{\pi_0}{\pi_1} + \log \pi_1 = \widetilde{\Lambda}_1(u),\end{aligned} \tag{44}$$

where (a) holds because $\Lambda_1(u) = \Lambda_0(u+1)$ from Property 4.

□

Proof of Lemma 9. We specifically consider the case where $\pi_0 \neq \pi_1$ in this proof because the case of equal priors $\pi_0 = \pi_1$ can be proved using Lemma 2 and Lemma 8.

Without loss of generality, we assume $\pi_0 > \pi_1$. Thus, $\log \frac{\pi_0}{\pi_1} > 0$.

Case 1: $\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} = \mathrm{D}(P_1||P_0) - \log \frac{\pi_0}{\pi_1} > 0$.

Observe that, $\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=-1} = -\mathrm{D}(P_0||P_1) - \log \frac{\pi_0}{\pi_1} < 0$ and $\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} = \mathrm{D}(P_1||P_0) - \log \frac{\pi_0}{\pi_1} > 0$. Thus, the strictly convex function $\widetilde{\Lambda}_1(u)$ attains its minima in $(-1,0)$ (using Lemma 10). Next, using $\widetilde{\Lambda}_0(u+1) = \widetilde{\Lambda}_1(u)$ (also from Lemma 10), we have $\widetilde{\Lambda}_0(u)$ attaining its minima in $(0,1)$.

For $\tau' = 0$ (equivalently $\tau_0 = \log \frac{\pi_0}{\pi_1}$), we have

$$\begin{aligned}E_{\mathrm{FP},T_0}(\log \frac{\pi_0}{\pi_1}) - \log 2\pi_0 \overset{(a)}{=} \sup_{u>0}(u \cdot 0 - \widetilde{\Lambda}_0(u) - \log 2) &\overset{(b)}{=} -\min_u \widetilde{\Lambda}_0(u) - \log 2 \\ &\overset{(c)}{=} -\min_u \widetilde{\Lambda}_1(u) - \log 2 \\ &\overset{(d)}{=} \sup_{u<0}(u \cdot 0 - \widetilde{\Lambda}_1(u) - \log 2) \\ &\overset{(e)}{=} E_{\mathrm{FN},T_0}(\log \frac{\pi_0}{\pi_1}) - \log 2\pi_1.\end{aligned} \tag{45}$$

Here, (a) holds from (42), (b) holds because $\widetilde{\Lambda}_0(u)$ attains its minima in $(0,1)$, (c) holds from $\widetilde{\Lambda}_0(u+1)=\widetilde{\Lambda}_1(u)$ (see Lemma 10), (d) holds because $\widetilde{\Lambda}_1(u)$ attains its minima in $(-1,0)$, and (e) holds from (43).

Next, we will show that, for any other value of $\tau' \neq 0$ ($\tau_0 \neq \log\frac{\pi_0}{\pi_1}$), we either have

$$E_{\text{FP},T_0}(\tau_0)-\log 2\pi_0 < E_{\text{FP},T_0}(\log\frac{\pi_0}{\pi_1})-\log 2\pi_0 < E_{\text{FN},T_0}(\tau_0)-\log 2\pi_1, \text{ or}$$
$$E_{\text{FN},T_0}(\tau_0)-\log 2\pi_1 < E_{\text{FP},T_0}(\log\frac{\pi_0}{\pi_1})-\log 2\pi_0 < E_{\text{FP},T_0}(\tau_0)-\log 2\pi_0. \tag{46}$$

Let $\tau'>0$. Then,

$$E_{\text{FP},T_0}(\tau_0)-\log 2\pi_0 \overset{(a)}{=} \sup_{u>0}(u\tau'-\widetilde{\Lambda}_0(u)-\log 2) \overset{(b)}{\geq} (u_0^*\tau'-\widetilde{\Lambda}_0(u_0^*)-\log 2) \overset{(c)}{>} -\widetilde{\Lambda}_0(u_0^*)-\log 2$$
$$\overset{(d)}{=} E_{\text{FP},T_0}(\log\frac{\pi_0}{\pi_1})-\log 2\pi_0. \tag{47}$$

Here (a) holds from (42), (b) holds for any $u_0^*>0$, (c) holds because $u_0\tau'>0$, and (d) holds if we set $u_0^*=\arg\min\widetilde{\Lambda}_0(u)$ since $\widetilde{\Lambda}_0(u)$ attains its minima in $(0,1)$.

Sub-case 1a: $\tau' \geq \frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} = \mathrm{D}(P_1||P_0)-\log\frac{\pi_0}{\pi_1}$

$$\begin{aligned}
E_{\text{FN},T_0}(\tau_0)-\log 2\pi_1 = \sup_{u<0}(u\tau'-\widetilde{\Lambda}_1(u)-\log 2) &\overset{(a)}{\leq} \sup_{u<0}(u\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0}-\widetilde{\Lambda}_1(u)-\log 2)\\
&\leq \sup_{u\in\mathcal{R}}(u\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0}-\widetilde{\Lambda}_1(u)-\log 2)\\
&\overset{(b)}{=} (0\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0}-\widetilde{\Lambda}_1(0)-\log 2)\\
&= (-\widetilde{\Lambda}_1(0)-\log 2)\\
&\overset{(c)}{<} -\min_u\widetilde{\Lambda}_1(u)-\log 2\\
&\overset{(d)}{=} E_{\text{FP},T_0}(\log\frac{\pi_0}{\pi_1})-\log 2\pi_0,
\end{aligned} \tag{48}$$

where (a) holds because $\tau' \geq \frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0}$, (b) holds from Lemma 7, (c) holds from the strict convexity of $\widetilde{\Lambda}_1(u)$ because it attains its minima in $(-1,0)$, and (d) holds from (45).

Sub-case 1b: $0<\tau'<\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0}$

$$\begin{aligned}
E_{\text{FN},T_0}(\tau_0)-\log 2\pi_0 = \sup_{u<0}(u\tau'-\widetilde{\Lambda}_1(u)-\log 2) &\leq \sup_{u\in\mathcal{R}}(u\tau'-\widetilde{\Lambda}_1(u)-\log 2)\\
&\overset{(a)}{=} u_a\tau'-\widetilde{\Lambda}_1(u_a)-\log 2\\
&\overset{(b)}{<} -\widetilde{\Lambda}_1(u_a)-\log 2 \quad [\text{since } u_a\tau'<0]\\
&\leq -\min_u\Lambda_1(u)-\log 2\\
&\overset{(c)}{=} E_{\text{FP},T_0}(\log\frac{\pi_0}{\pi_1})-\log 2\pi_0
\end{aligned} \tag{49}$$

Here, (a) holds from Lemma 7 because $\widetilde{\Lambda}_1(u)$ is a strictly convex and differentiable function, and its derivative is also continuous, monotonically increasing and takes all values from $-\infty$ to ∞ (see Lemma 10). Thus, there exists a single u_a such that $\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=u_a}=\tau'$. Next, (b) holds because $\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=u_a}=\tau'<\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0}$, and the derivative is monotonically increasing, implying $u_a<0$. Lastly (c) holds from (45).

Thus,

$$E_{\mathrm{FN},T_0}(\tau_0) - \log 2\pi_1 < E_{\mathrm{FP},T_0}(\log\frac{\pi_0}{\pi_1}) - \log 2\pi_0 < E_{\mathrm{FP},T_0}(\tau_0) - \log 2\pi_0. \tag{50}$$

For $\tau' < 0$, a similar proof holds, leading to

$$E_{\mathrm{FP},T_0}(\tau_0) - \log 2\pi_0 < E_{\mathrm{FP},T_0}(\log\frac{\pi_0}{\pi_1}) - \log 2\pi_0 < E_{\mathrm{FN},T_0}(\tau_0) - \log 2\pi_1, \tag{51}$$

Then, the value of τ_0 that maximizes the Chernoff exponent $E_{e,T_0}(\tau_0)$, i.e.,

$$\max_{\tau_0} \min\{E_{\mathrm{FP},T_0}(\tau_0) - \log 2\pi_0, E_{\mathrm{FN},T_0}(\tau_0) - \log 2\pi_1\},$$

is given by $\tau_0^* = \log\frac{\pi_0}{\pi_1}$ ($\tau' = 0$).

This matches with the detector that minimizes the Bayesian probability of error under unequal priors (see Theorem 3.1 in (Gallager, 2012)).

Case 2: $\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} = \mathrm{D}(P_1||P_0) - \log\frac{\pi_0}{\pi_1} \leq 0$.

For this case, note that, both $\widetilde{\Lambda}_1(u)$ and $\widetilde{\Lambda}_0(u)$ attain their minima in $u \in [0,\infty)$.

For $\tau' = 0$ (equivalently $\tau_0 = \log\frac{\pi_0}{\pi_1}$), we have

$$E_{\mathrm{FN},T_0}(\log\frac{\pi_0}{\pi_1}) - \log 2\pi_1 = \sup_{u<0}(u\cdot 0 - \widetilde{\Lambda}_1(u) - \log 2) = -\widetilde{\Lambda}_1(0) - \log 2. \tag{52}$$

And,

$$\begin{aligned} E_{\mathrm{FP},T_0}(\log\frac{\pi_0}{\pi_1}) - \log 2\pi_0 = \sup_{u>0}(u\cdot 0 - \widetilde{\Lambda}_0(u) - \log 2) &= -\min_u \widetilde{\Lambda}_0(u) - \log 2 \\ &= -\min_u \widetilde{\Lambda}_1(u) - \log 2 \\ &\geq -\widetilde{\Lambda}_1(0) - \log 2. \end{aligned} \tag{53}$$

Thus,

$$\min\{E_{\mathrm{FP},T_0}(\log\frac{\pi_0}{\pi_1}) - \log 2\pi_0, E_{\mathrm{FN},T_0}(\log\frac{\pi_0}{\pi_1}) - \log 2\pi_1\} = -\widetilde{\Lambda}_1(0) - \log 2. \tag{54}$$

Now, we will show that any other value of $\tau' \neq 0$ (equivalently $\tau_0 \neq \log\frac{\pi_0}{\pi_1}$) cannot increase the Chernoff exponent of the probability of error beyond $-\widetilde{\Lambda}_1(0) - \log 2$.

Sub-case 2a: $\tau' \geq \frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} = \mathrm{D}(P_1||P_0) - \log\frac{\pi_0}{\pi_1}$

$$\begin{aligned} E_{\mathrm{FN},T_0}(\tau_0) - \log 2\pi_1 = \sup_{u<0}(u\tau' - \widetilde{\Lambda}_1(u) - \log 2) &\overset{(a)}{\leq} \sup_{u<0}(u\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} - \widetilde{\Lambda}_1(u) - \log 2) \\ &\leq \sup_{u\in\mathcal{R}}(u\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} - \widetilde{\Lambda}_1(u) - \log 2) \\ &\overset{(b)}{=} (0\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} - \widetilde{\Lambda}_1(0) - \log 2) \\ &= (-\widetilde{\Lambda}_1(0) - \log 2), \end{aligned} \tag{55}$$

where (a) holds because $\tau' \geq \frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0}$ and (b) holds from Lemma 7. Thus,

$$\min\{E_{\mathrm{FP},T_0}(\tau_0) - \log 2\pi_0, E_{\mathrm{FN},T_0}(\tau_0) - \log 2\pi_1\} \leq -\widetilde{\Lambda}_1(0) - \log 2. \tag{56}$$

Sub-case 2b: $\tau' < \frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} = \mathrm{D}(P_1||P_0) - \log\frac{\pi_0}{\pi_1}$

$$\begin{aligned}
E_{\mathrm{FP},T_0}(\tau_0) - \log 2\pi_0 = \sup_{u>0}(u\tau' - \widetilde{\Lambda}_0(u) - \log 2) &\overset{(a)}{\leq} \sup_{u>0}(u\frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} - \widetilde{\Lambda}_0(u) - \log 2) \\
&\overset{(b)}{\leq} \sup_{u>0}(u\frac{d\widetilde{\Lambda}_0(u)}{du}|_{u=1} - \widetilde{\Lambda}_0(u) - \log 2) \\
&\overset{(c)}{\leq} \sup_{u\in\mathcal{R}}(u\frac{d\widetilde{\Lambda}_0(u)}{du}|_{u=1} - \widetilde{\Lambda}_0(u) - \log 2) \\
&\overset{(d)}{=} \frac{d\widetilde{\Lambda}_0(u)}{du}|_{u=1} - \widetilde{\Lambda}_0(1) - \log 2 \\
&\overset{(e)}{\leq} -\widetilde{\Lambda}_0(1) - \log 2 \\
&\overset{(f)}{=} -\widetilde{\Lambda}_1(0) - \log 2. \qquad (57)
\end{aligned}$$

Here (a) holds because $\tau' < \frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0}$, (b) holds from Lemma 10 since $\widetilde{\Lambda}_1(u) = \widetilde{\Lambda}_0(u+1)$, (c) holds because the supremum is taken over a larger superset, (d) holds from Lemma 7, (e) holds because $\frac{d\widetilde{\Lambda}_0(u)}{du}|_{u=1} = \frac{d\widetilde{\Lambda}_1(u)}{du}|_{u=0} = \mathrm{D}(P_1||P_0) - \log\frac{\pi_0}{\pi_1} \leq 0$, and (f) holds again from from Lemma 10 since $\widetilde{\Lambda}_1(u) = \widetilde{\Lambda}_0(u+1)$. Thus,

$$\max_{\tau_0} \min\{E_{\mathrm{FP},T_0}(\tau_0) - \log 2\pi_0, E_{\mathrm{FN},T_0}(\tau_0) - \log 2\pi_1\} = -\widetilde{\Lambda}_1(0) - \log 2, \qquad (58)$$

which is attained at $\tau_0 = \log\frac{\pi_0}{\pi_1}$.

□

E.2. Unequal priors on both Z and Y

Here we discuss a modification of optimization (2) proposed in Section 3.1 to account for the case of unequal priors on both Z and Y.

Let $\Pr(Z=0) = \lambda_0$ and $\Pr(Z=1) = \lambda_1$. Also let, $\Pr(Y=0|Z=0) = \pi_{00}$, $\Pr(Y=1|Z=0) = \pi_{10}$, $\Pr(Y=0|Z=1) = \pi_{01}$ and $\Pr(Y=1|Z=1) = \pi_{11}$.

Then, the overall probability of error considering both groups together is given by:

$$\begin{aligned}
&\lambda_0 P_e^{T_0}(\tau_0) + \lambda_1 P_e^{T_1}(\tau_1) \\
&= \frac{1}{2}(2\lambda_0)P_e^{T_0}(\tau_0) + \frac{1}{2}(2\lambda_1)P_e^{T_1}(\tau_1) \\
&= \frac{1}{4}(4\lambda_0\pi_{00})P_{\mathrm{FP},T_0}(\tau_0) + \frac{1}{4}(4\lambda_0\pi_{10})P_{\mathrm{FN},T_0}(\tau_0) + \frac{1}{4}(4\lambda_1\pi_{01})P_{\mathrm{FP},T_1}(\tau_1) + \frac{1}{4}(4\lambda_1\pi_{11})P_{\mathrm{FN},T_1}(\tau_1). \qquad (59)
\end{aligned}$$

Then, the error exponent of the overall probability of error considering both groups is defined as:

$$\min\{E_{\mathrm{FP},T_0}(\tau_0) - 4\pi_{00}\lambda_0, E_{\mathrm{FN},T_0}(\tau_0) - 4\pi_{10}\lambda_0, E_{\mathrm{FP},T_1}(\tau_1) - 4\pi_{01}\lambda_1, E_{\mathrm{FN},T_1}(\tau_1) - 4\pi_{11}\lambda_1\}. \qquad (60)$$

These log-generating functions can be plotted, and the intercepts made by their tangents can be examined again to obtain the error exponents, leading to the optimal detector.