Calibrated Surrogate Losses for Adversarially Robust Classification

Han Bao*
*The University of Tokyo
RIKEN AIP

Clayton Scott
University of Michigan

Masashi Sugiyama
RIKEN AIP
The University of Tokyo

Editors: Jacob Abernethy and Shivani Agarwal

Abstract

Adversarially robust classification seeks a classifier that is insensitive to adversarial perturbations of test patterns. This problem is often formulated via a minimax objective, where the target loss is the worst-case value of the 0-1 loss subject to a bound on the size of perturbation. Recent work has proposed convex surrogates for the adversarial 0-1 loss, in an effort to make optimization more tractable. In this work, we consider the question of which surrogate losses are calibrated with respect to the adversarial 0-1 loss, meaning that minimization of the former implies minimization of the latter. We show that no convex surrogate loss is calibrated with respect to the adversarial 0-1 loss when restricted to the class of linear models. We further introduce a class of nonconvex losses and offer necessary and sufficient conditions for losses in this class to be calibrated.

Keywords: surrogate loss, classification calibration, adversarial robustness

1. Introduction

In conventional machine learning, training and testing instances are assumed to follow the same probability distribution. In adversarially robust machine learning, test instances may be perturbed by an adversary before being presented to the predictor. Recent work has shown that seemingly insignificant adversarial perturbations can lead to significant performance degradations of otherwise highly accurate classifiers (Goodfellow et al., 2015). This has led to the development of a number of methods for learning predictors with decreased sensitivity to adversarial perturbations (Xu et al., 2009; Xu and Mannor, 2012; Goodfellow et al., 2015; Cisse et al., 2017; Wong and Kolter, 2018; Raghunathan et al., 2018a; Tsuzuku et al., 2018).

Adversarially robust classification is typically formulated as empirical risk minimization with an adversarial 0-1 loss, which is the maximum of the usual 0-1 loss over a set of possible perturbations of the test instance. This minimax optimization problem is nonconvex, and recent work, reviewed in Section 4, has proposed several convex surrogate losses. However, it is still unknown whether minimizing these convex surrogates leads to minimization of the adversarial 0-1 loss.

In this work, we examine the question of which surrogate losses are calibrated with respect to (wrt) the adversarial 0-1 loss. A surrogate loss is said to be calibrated wrt a target loss if minimiza-

* This work was performed while the first author was a visitor at University of Michigan.

Adversarially Robust Classification Calibration

Figure 1: The best linear classifier under each loss. The shift parameter \( \beta \) for a surrogate loss is defined in Section 8. The \( \ell_2 \)-balls associated to each instance indicate adversarial perturbations with radii 0.1. The yellow balls indicate instances vulnerable to perturbations, in that they are within 0.1 of the decision boundary. In this example, 1.2% of instances are vulnerable under the ramp loss, while 24.8% of instances are vulnerable under the hinge loss.

Our results demonstrate that adversarial robustness requires different surrogates than other notions of robustness. For example, symmetric losses such as the sigmoid and ramp losses are robust to label noise (Ghosh et al., 2015), but not calibrated wrt the adversarial 0-1 loss. Figure 1 illustrates the results of learning a linear classifier with respect a shifted ramp loss, which is calibrated wrt the adversarial 0-1 loss, and a shifted hinge loss, which is not (these losses are discussed in detail later). While the hinge loss yields a classifier with smaller misclassification rate wrt the conventional 0-1 loss, this classifier is quite sensitive to small perturbations of the test instances. The classifier learned by the ramp loss, on the other hand, makes fewer errors when subjected to adversarial perturbations.

The rest of this paper is organized as follows. Section 3 formalizes notation and the problem. Related work on robust learning and calibration analysis is reviewed in Section 4. Technical details of calibration analysis are reviewed in Section 5. Section 6 describes the nonexistence of convex calibrated surrogate losses, while Section 7 presents general calibration conditions for a certain
class of nonconvex losses. Section 8 applies our theory to several convex and nonconvex losses, and presents excess risk bounds for the calibrated nonconvex losses. Section 9 shows simulation results to verify that calibrated losses achieve target excess risk close to zero under the robust 0-1 loss. Conclusions are stated in Section 10.

2. Notation

Let \( \|x\|_p \) for a vector \( x \in \mathbb{R}^d \) be the \( \ell_p \)-norm, namely, \( \|x\|_p = \sqrt[d]{\sum_{i=1}^{d} |x_i|^p} \). Let \( B_p^d(r) = \{ v \in \mathbb{R}^d \mid \|v\|_p \leq r \} \) be the \( d \)-dimensional \( \ell_p \)-ball with radius \( r \). Let \( \text{dom}(h) \) be the domain of a function \( h \). We define the epigraph of the function \( h \): \( \text{epi}(h) = \{ (x, t) \mid x \in S, h(x) \leq t \} \). A function \( h : S \to \mathbb{R} \) is said to be quasiconcave if for all \( x_1, x_2 \in S \) and \( \lambda \in [0, 1] \), \( h(\lambda x_1 + (1 - \lambda)x_2) \geq \min\{h(x_1), h(x_2)\} \).

Let \( X, Y \subseteq \mathbb{R} \) be a function class. We consider symmetric \( F \), that is, \(-f \in F \) for all \( f \in F \). We write \( F_{\text{all}} \subseteq \mathcal{X} \times \mathcal{Y} \times F \) for the space of all measurable functions. Let \( \ell : \mathcal{Y} \times X \times F \to \mathbb{R}_{\geq 0} \) be a loss function. Then, we write \( \mathcal{R}_{\ell}(f) = \mathbb{E}_{(Y,X,f)}[\ell(Y,X,f)] \) for the \( \ell \)-risk of \( f \in F \). If \( \ell \) can be represented by \( \ell(y,x,f) = \phi(yf(x)) \) with some \( \phi : \mathbb{R} \to \mathbb{R}_{\geq 0} \) for any \( y \in \mathcal{Y}, x \in \mathcal{X}, f \in F \), \( \phi \) is called a margin-based loss function. We define the \( \phi \)-risk of \( f \in F \) for a margin-based loss \( \phi \) by

\[
\mathcal{R}_{\phi}(f) = \mathbb{E}_{(X,Y)}[\phi(Yf(X))] = \mathbb{E}_X \mathbb{E}_{Y|X}[\phi(Yf(X))],
\]

where \( \mathbb{E}_X \) and \( \mathbb{E}_{Y|X} \) mean the expectation over \( \mathbb{P}(X) \) and \( \mathbb{P}(Y|X) \), respectively. We can rewrite (1) as \( \mathcal{R}_{\phi}(f) = \mathbb{E}_X[\mathcal{C}_{\phi}(f(X), \mathbb{P}(Y = +1|X))] \) with \( \mathcal{C}_{\phi}(\alpha, \eta) \) defined as \( \eta \phi(\alpha) + (1 - \eta)\phi(-\alpha) \). We call \( \mathcal{C}_{\phi}(\alpha, \eta) \) the class-conditional \( \phi \)-risk (\( \phi \)-CCR). The minimal \( \phi \)-risk \( \mathcal{R}_{\phi,F}^* = \inf_{f \in F} \mathcal{R}_{\phi}(f) \) is called the Bayes \( (\phi, F) \)-risk, and the minimal \( \phi \)-CCR on \( F \) is denoted by \( \mathcal{C}_{\phi,F}^*(\eta) = \inf_{f \in F} \mathcal{C}_{\phi,F}(\alpha, \eta) \), where \( \mathcal{C}_{\phi,F}(\alpha, \eta) = \{ \alpha = f(x) \mid f \in F, x \in \mathcal{X} \} \). We refer to \( \mathcal{R}_{\phi}(f) - \mathcal{R}_{\phi,F}^* \) as the \( (\phi, F) \)-excess risk.

We occasionally use the abbreviation \( \Delta \mathcal{C}_{\phi,F}(\alpha, \eta) = \mathcal{C}_{\phi}(\alpha, \eta) - \mathcal{C}_{\phi,F}^*(\eta) \).

3. Surrogate losses for Adversarially Robust Classification

In supervised binary classification, a learner is asked to output a predictor \( f : \mathcal{X} \to \mathbb{R} \) that minimizes the classification error \( \mathbb{P}\{Yf(\mathcal{X}) \leq 0\} \), where \( \mathbb{P} \) is the unknown underlying distribution. This can be equivalently interpreted as the minimization of the risk \( \mathbb{E}_{(X,Y)}[\ell_{01}(Y, X, f)] \) wrt \( f \), where

\[
\ell_{01}(y, x, f) = \begin{cases} 
1 & \text{if } yf(x) \leq 0, \\
0 & \text{otherwise}
\end{cases}
\]

is the 0-1 loss. Letting \( \phi_{01}(\alpha) = \mathbb{1}_{\{\alpha \leq 0\}} \), then \( \ell_{01}(y, x, f) = \phi_{01}(yf(x)) \). On the other hand, an adversarially robust learner is asked to output a predictor \( f \) that minimizes the 0-1 loss while being tolerant to small perturbations to input data points. Following existing literature (Xu et al., 2009;
Tsuzuku et al., 2018; Bubeck et al., 2019), we consider $\ell_2$-ball perturbations and define the goal as the minimization of $P\{\exists \Delta_x \in B^d_2(\gamma) \text{ s.t. } X + \Delta_x \in \mathcal{X} \text{ and } Yf(X + \Delta_x) \leq 0\}$, where $\Delta_x$ is a perturbation vector and $\gamma \in (0, 1)$ is a pre-defined perturbation budget. Equivalently, the goal of adversarially robust classification is to minimize $\mathbb{E}_{(X,Y)}[\ell_\gamma(Y, X, f)]$ wrt $f$, where

$$\ell_\gamma(y, x, f) \triangleq \begin{cases} 1 & \text{if } \exists \Delta_x \in B^d_2(\gamma) \text{ s.t. } x + \Delta_x \in \mathcal{X} \text{ and } yf(x + \Delta_x) \leq 0, \\ 0 & \text{otherwise.} \end{cases}$$

We call this loss function $\ell_\gamma$ the **adversarially robust 0-1 loss**, or the **robust 0-1 loss** for short.

The robust 0-1 loss is also a margin-based loss when restricted to the class of linear models $\mathcal{F}_{lin} \triangleq \{x \mapsto \theta^\top x \mid \theta \in \mathbb{R}^d, \|\theta\|_2 = 1\} \subseteq \mathbb{R}^\mathcal{X}$. Note that $\mathcal{F}_{lin}$ is symmetric.

**Proposition 1** For any $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $f \in \mathcal{F}_{lin}$, we have $\ell_\gamma(y, x, f) = 1_{\{yf(x) \leq \gamma\}}$.

We include the proof in Appendix B for completeness though it is mentioned as a fact by Diakonikolas et al. (2019). Subsequently, when considering $\mathcal{F}_{lin}$, we work with the loss function $\phi_\gamma(\alpha) \triangleq 1_{\{\alpha \leq \gamma\}}$ and call $\phi_\gamma$ the **$\gamma$-robust 0-1 loss**. We will study calibrated surrogates wrt $\phi_\gamma$ instead of $\ell_\gamma$, and both are equivalent under the restricted function class $\mathcal{F}_{lin}$. We can view $\phi_\gamma$ as a shifted version of $\phi_{01}$.

In many machine learning problems, there are often dichotomies between optimization (learning) and evaluation. For instance, binary classification is evaluated by the 0-1 loss, while common learning methods such as SVM and logistic regression minimize surrogates to the 0-1 loss. This dichotomy arises because minimizing the 0-1 loss directly is known to be NP-hard (Feldman et al., 2012). Much research has investigated surrogates $\phi$ satisfying

$$\mathcal{R}_\phi(f_i) - \mathcal{R}_{\phi,\mathcal{F}}^* \rightarrow 0 \implies \mathcal{R}_\ell(f_i) - \mathcal{R}_{\ell,\mathcal{F}}^* \rightarrow 0,$$

for all probability distributions and sequence of $(f_i)_{i \in \mathbb{N}} \subseteq \mathcal{F}$.

Our learning goal is to minimize the expected $\gamma$-robust 0-1 loss on a given function class $\mathcal{F}$:

$$\min_{f \in \mathcal{F}} \mathcal{R}_{\ell_\gamma}(f).$$

(3)

In order to solve (3), we aim to characterize surrogate losses $\phi$ satisfying (2) with $\ell = \ell_\gamma$ and $\mathcal{F} = \mathcal{F}_{lin}$. By Proposition 1, we have $\mathcal{R}_{\ell_\gamma}(f) = \mathcal{R}_{\phi_\gamma}(f)$ when $\mathcal{F} = \mathcal{F}_{lin}$. 

### 4. Related Work

From the viewpoint of robust optimization (Ben-Tal et al., 2009; Bertsimas et al., 2011), adversarially robust binary classification can be formulated as

$$\min_{f \in \mathcal{F}} \mathbb{E}_{(X,Y)} \left[ \max_{\tilde{X} \in \mathcal{U}(X)} \ell(Y, \tilde{X}, f) \right],$$

(4)

where $\ell$ is a loss function and $\mathcal{U}(x)$ is a user-specified uncertainty set. Our formulation of adversarially robust classification (3) can be regarded as the special case $\ell = \ell_{01}$ and $\mathcal{U}(x) = x + B^d_2(\gamma)$.

Since the minimax problem (4) is generally nonconvex, it is traditionally tackled by minimizing a convex upper bound. Lanckriet et al. (2002) and Shivaswamy et al. (2006) pick $\mathcal{U}(x) = \{x \sim$
as an uncertainty set, where $x \sim (\bar{x}, \Sigma_x)$ means that $x$ is drawn from a distribution that has prespecified mean $\bar{x}$, covariance $\Sigma_x$, and arbitrary higher moments. Lanckriet et al. (2002) and Shivaswamy et al. (2006) convexified (4) and obtained a second-order cone program. Xu et al. (2009) studied the relationship between robustness and regularization, and showed that (4) with the hinge loss and $U(x) = x + B^{d_2}_2(\gamma)$ is equivalent to $\ell_2$-regularized SVM. Recently, Wong and Kolter (2018), Madry et al. (2018), Raghunathan et al. (2018a), Raghunathan et al. (2018b), and Khim and Loh (2019) examined (4) with the softmax cross entropy loss and $U(x) = x + B^{d_\infty}_\infty(\gamma)$ when $F$ is a set of deep nets, and provided convex upper bounds of the worst-case loss in (4). However, no work except Cranko et al. (2019) studied whether the surrogate objectives minimize the robust 0-1 excess risk. Cranko et al. (2019) showed that no canonical proper loss (Reid and Williamson, 2010) can minimize the robust 0-1 loss. Since canonical proper losses are convex, this result aligns with ours. We show more general results via calibration analysis for $U(x) = x + B^{d_2}_2(\gamma)$.

There are several other approaches to the robust classification such as minimizing the Taylor approximation of the worst-case loss in (4) (Goodfellow et al., 2015; Gu and Rigazio, 2015; Shaham et al., 2018), regularization on the Lipschitz norm of models (Cisse et al., 2017; Hein and Andriushchenko, 2017; Tsuzuku et al., 2018), and injection of random noises to model parameters (Lecuyer et al., 2019; Cohen et al., 2019; Pinot et al., 2019; Salman et al., 2019). It is not known whether these methods imply the minimization of the robust 0-1 excess risk.

Other forms of robustness have also been considered in the literature. A number of existing works considered the worst-case test distribution. This line includes divergence-based methods (Namkoong and Duchi, 2016, 2017; Hu et al., 2018; Sinha et al., 2018), domain adaptation (Mansour et al., 2009; Ben-David et al., 2010; Germain et al., 2013; Kuroki et al., 2019; Zhang et al., 2019b), and methods based on constraints on feature moments (Farnia and Tse, 2016; Fathony et al., 2016).

In addition to adversarial robustness, it is worthwhile to mention outlier and label-noise robustness. It is known that convex losses are vulnerable to outliers, thus truncation making losses nonconvex is useful (Huber, 2011). In the machine learning context, Masnadi-Shirazi and Vasconcelos (2009) and Holland (2019) designed nonconvex losses robust to outliers. On the other hand, label-noise robustness, especially the random classification noise model, has been studied extensively (Angluin and Laird, 1988), where training labels are flipped with a fixed probability. Long and Servedio (2010) showed that there is no convex loss that is robust to label noises. Later, Ghosh et al. (2015), van Rooyen et al. (2015), and Charoenphakdee et al. (2019) discovered a certain class of nonconvex losses is a good alternative for label-noise robustness. In both outlier and label-noise robustness, nonconvex loss functions play an important role as we see in adversarial robustness.

Calibration analysis has been formalized in Lin (2004), Zhang et al. (2004), Bartlett et al. (2006), and Steinwart (2007), and employed to analyze not only binary classification, but also complicated problems such as multi-class classification (Zhang, 2004; Tewari and Bartlett, 2007; Long and Servedio, 2013; Ávila Pires and Szepesvári, 2016; Ramaswamy and Agarwal, 2016), multi-label classification (Gao and Zhou, 2011; Dembczynski et al., 2012), cost-sensitive learning (Scott, 2011, 2012; Ávila Pires et al., 2013), ranking (Duchi et al., 2010; Ravikumar et al., 2011; Ramaswamy et al., 2013), structured prediction (Hazan et al., 2010; Ramaswamy and Agarwal, 2012; Osokin et al., 2017; Blondel, 2019), AUC optimization (Gao and Zhou, 2015), and optimization of non-decomposable metrics (Bao and Sugiyama, 2020). Zhang et al. (2004), Ravikumar et al. (2011), and Gao and Zhou (2015) figured out ad hoc derivations of excess risk bounds, while Bartlett et al. (2006), Steinwart (2007), Scott (2012), Ávila Pires et al. (2013), Ávila Pires and Szepesvári (2016),
Osokin et al. (2017), and Blondel (2019) used more systematic approaches. As for adversarially robust classification, Zhang et al. (2019a, Theorem 3.1) applied the classical result of calibration analysis on convex losses to upper bound the robust classification risk, resulting in a term requiring numerical approximation in practice.

5. Calibration Analysis

Calibration analysis is a tool to study the relationship between surrogate losses and target losses. This section is devoted to explaining the calibration function introduced in Steinwart (2007) and specializing it to the current paper.

**Definition 2** For a loss \( \psi : \mathbb{R} \to \mathbb{R}_{\geq 0} \) and a function class \( \mathcal{F} \), we say a loss \( \phi : \mathbb{R} \to \mathbb{R}_{\geq 0} \) is calibrated wrt \( (\psi, \mathcal{F}) \), or \( (\psi, \mathcal{F}) \)-calibrated, if for any \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that for all \( \eta \in [0, 1] \) and \( \alpha \in \mathcal{A}_\mathcal{F} \), we have

\[
C_\phi(\alpha, \eta) < C^*_\psi(\eta) + \delta \implies C_\psi(\alpha, \eta) < C^*_\psi(\eta) + \varepsilon.
\] (5)

If \( \phi \) is \( (\psi, \mathcal{F}) \)-calibrated, the condition (2) holds for any probability distribution on \( \mathcal{X} \times \mathcal{Y} \) (Steinwart, 2007, Theorem 2.8). Thus, consistency of a learner wrt the \( \phi \)-risk implies consistency wrt the \( \psi \)-risk.

Next, we introduce the calibration function (Steinwart, 2007, Lemma 2.16).

**Definition 3** For a margin-based loss \( \psi \) and \( \phi \), and a function class \( \mathcal{F} \), the calibration function of \( \phi \) wrt \( (\psi, \mathcal{F}) \), or simply calibration function if the context is clear, is defined as

\[
\delta(\varepsilon) = \inf_{\eta \in [0, 1]} \inf_{\alpha \in \mathcal{A}_\mathcal{F}} C_\phi(\alpha, \eta) - C^*_\psi(\eta) \quad \text{s.t.} \quad C_\psi(\alpha, \eta) - C^*_\psi(\eta) \geq \varepsilon.
\] (6)

Note that \( \delta(\varepsilon) \) is nondecreasing for \( \varepsilon > 0 \). The calibration function \( \delta(\varepsilon) \) is the maximal \( \delta \) satisfying the CCR condition (5). Steinwart (2007) established the following two important results.

**Proposition 4 (Lemma 2.9 in Steinwart (2007))** A surrogate loss \( \phi \) is \( (\psi, \mathcal{F}) \)-calibrated if and only if its calibration function \( \delta \) satisfies \( \delta(\varepsilon) > 0 \) for all \( \varepsilon > 0 \).

**Proposition 5 (Theorem 2.13 in Steinwart (2007))** Let \( \delta : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \) be the calibration function of \( \phi \) wrt \( (\psi, \mathcal{F}) \). Define \( \tilde{\delta} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \) as \( \tilde{\delta}(\varepsilon) = \delta(\varepsilon) \) if \( \varepsilon > 0 \) and \( \tilde{\delta}(0) = 0 \). Then, for all \( f \in \mathcal{F} \), we have

\[
\tilde{\delta}^{**}(\mathcal{R}_\psi(f) - \mathcal{R}^*_\psi) \leq \mathcal{R}_\phi(f) - \mathcal{R}^*_\phi,
\] (7)

where \( \tilde{\delta}^{**} \) denotes the Fenchel-Legendre biconjugate of \( \tilde{\delta} \).

The relationship in (7) is called an excess risk transform. The excess risk transform is invertible iff \( \phi \) is \( (\psi, \mathcal{F}) \)-calibrated (Steinwart, 2007, Remark 2.14). In this case, we obtain the excess risk bound \( \mathcal{R}_\psi(f) - \mathcal{R}^*_\psi \leq (\tilde{\delta}^{**})^{-1}(\mathcal{R}_\phi(f) - \mathcal{R}^*_\phi) \). In the end, the calibration function can be used in two

---

1. We import toolsets from Steinwart (2007) because of two reasons: (i) Steinwart (2007) formalized calibration analysis that is dependent on user-specified function classes, which is useful for our analysis on \( \mathcal{F}_{\text{lin}} \). (ii) Steinwart (2007) gave a general form of the calibration function (6), while most of literature focuses on specific target losses.
ways: Proposition 4 enables us to check if a surrogate loss is calibrated, and Proposition 5 gives us a quantitative relationship between the surrogate excess risk and the target excess risk. Such an analysis has been carried out in a number of learning problems as we mention in Section 4.

Next, we review an important result regarding convex surrogates for the non-robust 0-1 loss $\phi_{01}$.

**Proposition 6 (Theorem 6 in Bartlett et al. (2006))** Let $\phi$ be a convex surrogate loss. Then, $\phi$ is calibrated wrt ($\phi_{01}, F_{\text{all}}$) if and only if it is differentiable at 0 and $\phi'(0) < 0$.

As a result of Proposition 6, we know that many surrogate losses used in practice such as the hinge loss, logistic loss, and squared loss are calibrated wrt ($\phi_{01}, F_{\text{all}}$).

Finally, we characterize the calibration function of an arbitrary surrogate loss $\phi$ wrt $\phi_\gamma$. Its proof is deferred in Appendix B.

**Lemma 7** Let $\mathcal{F} \subseteq \mathbb{R}^X$ be a function class such that $A_{\mathcal{F}} \supseteq [-1, 1]$. For a surrogate loss $\phi$, the ($\phi_\gamma, \mathcal{F}$)-calibration function is $\tilde{\delta}(\varepsilon) = \inf_{\eta \in [0, 1]} \tilde{\delta}(\varepsilon, \eta)$, where

$$
\tilde{\delta}(\varepsilon, \eta) = \begin{cases} 
\infty & \text{if } \varepsilon > \max\{\eta, 1 - \eta\}, \\
\inf_{|\alpha| \leq \gamma} \Delta C_{\phi, \mathcal{F}}(\alpha, \eta) & \text{if } |2\eta - 1| < \varepsilon \leq \max\{\eta, 1 - \eta\}, \\
\inf_{\alpha \in A_{\mathcal{F}}(2\eta - 1) \alpha \leq 0 \text{ or } |\alpha| \leq \gamma} \Delta C_{\phi, \mathcal{F}}(\alpha, \eta) & \text{if } \varepsilon \leq |2\eta - 1|.
\end{cases}
$$

(8)

Lemma 7 is used in the proofs and examples below. Note that $A_{\mathcal{F}_{\text{lin}}} = [-1, 1]$ and $A_{\mathcal{F}_{\text{all}}} = \mathbb{R}$.

### 6. Convex Surrogates are Not ($\phi_\gamma, F_{\text{all}}$)-calibrated

Our first result concerns calibration of convex surrogate losses wrt the $\gamma$-robust 0-1 loss.

**Theorem 8** For any margin-based surrogate loss $\phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ and function class $\mathcal{F} \subseteq \mathbb{R}^X$ such that $A_{\mathcal{F}} \supseteq [-1, 1]$, if $\phi$ is convex, then $\phi$ is not calibrated wrt ($\phi_\gamma, \mathcal{F}$).

**Corollary 9** For any margin-based surrogate loss $\phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$, if $\phi$ is convex, then $\phi$ is not calibrated wrt ($\phi_\gamma, F_{\text{lin}}$), nor is it calibrated wrt ($\phi_\gamma, F_{\text{all}}$).

**Proof** (Sketch) Here we focus on function class $F_{\text{all}}$. In the non-robust setup, Bartlett et al. (2006) showed that a surrogate loss is calibrated wrt ($\phi_{01}, F_{\text{all}}$) iff $\inf_{(2\eta - 1) \alpha \leq 0} C_\phi(\alpha, \eta)$ (the minimum $\phi$-risk over ‘wrong’ predictions) is larger than $\inf_{\alpha \in \mathbb{R}} C_\phi(\alpha, \eta)$ (the minimum $\phi$-risk over all predictions) for $\eta \neq \frac{1}{2}$. This means wrong predictions must be penalized more. In our robust setup, we must penalize not only wrong predictions but also predictions that fall in the $\gamma$-margin, i.e.,

$$
\inf_{|\alpha| \leq \gamma} C_\phi(\alpha, \eta) > \inf_{\alpha \in \mathbb{R}} C_\phi(\alpha, \eta),
$$

(9)

which is an immediate corollary of Proposition 4 and Lemma 7 and stated in part 3 of Lemma 12 in Appendix B. Condition (9) becomes harder to satisfy as a data point gets more uncertain ($\eta \to \frac{1}{2}$). In the limit, we have $\inf_{|\alpha| \leq \gamma} \phi(\alpha) + \phi(-\alpha) > \inf_{\alpha \in \mathbb{R}} \phi(\alpha) + \phi(-\alpha)$, meaning that the even part of $\phi$ “should take larger values in $|\alpha| \leq \gamma$ than in the rest of $\alpha$.” However, $\phi(\alpha) + \phi(-\alpha)$ attains the
Infinum at $\alpha = 0$ because $\phi(\alpha) + \phi(-\alpha)$ is convex and even as long as $\phi$ is convex. Therefore, the condition (9) would never be satisfied by convex surrogate $\phi$. This idea is illustrated in Figure 2.

Hence, many popular surrogate losses such as the hinge, logistic, and squared losses are not calibrated wrt $(\phi, F_{\text{all}})$. We defer all proofs to Appendix B.

Note that convex losses can be calibrated wrt $(\phi, F_{\text{all}})$ under restricted distributions while we are primarily interested in calibrated losses under all distributions (see Definition 2). Indeed, $C_\phi(\alpha, \eta)$ would not be minimized in $|\alpha| \leq \gamma$ unless $\eta$ is close enough to $\frac{1}{2}$. In other words, convex losses may be calibrated wrt $(\phi, F_{\text{all}})$ under low-noise conditions (Mammen and Tsybakov, 1999).

### 7. Calibration Conditions for Nonconvex Surrogates

As seen in Section 6, convex surrogate losses that are calibrated wrt $(\phi, F_{\text{all}})$ do not exist. This motivates a search for nonconvex surrogate losses. Nonconvex surrogates are used for outlier robustness (Collobert et al., 2006; Masnadi-Shirazi and Vasconcelos, 2009; Holland, 2019) or label-noise robustness (Ghosh et al., 2015; van Rooyen et al., 2015; Charoenphakdee et al., 2019). Bounded monotone surrogates such as the ramp loss and the sigmoid loss are simple and common choices for those purposes. In this section, we also look for good surrogates from bounded monotone losses.

First, we introduce an important notion that constrains our search space of loss functions.

**Definition 10** We say a margin-based loss function $\phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is quasiconcave even if $\phi(\alpha) + \phi(-\alpha)$ is quasiconcave. Such $\phi$ is called a quasiconcave even loss.

The name comes from the fact that any function $h(x)$ may be uniquely expressed as the sum of its even part $\frac{h(x) + h(-x)}{2}$ and odd part $\frac{h(x) - h(-x)}{2}$. This fact is also utilized to study the relationship between loss functions and sufficiency (Patrini et al., 2016).

Next, we state our main positive result. Its proof is included in Appendix B.

**Theorem 11** Let $\phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a surrogate loss. Assume that $\phi$ is bounded, nonincreasing, and quasiconcave even. Let $B \overset{\text{def}}{=} \phi(1) + \phi(-1)$ and assume $\phi(-1) > \phi(1)$. Let $F \subseteq \mathbb{R}^Y$ be a function class such that $A_F \supseteq [-1, 1]$. Then,

1. $\phi$ is $(\phi_0, F)$-calibrated.
2. φ is \((\phi_{\gamma}, F)\)-calibrated if and only if \(\phi(\gamma) + \phi(-\gamma) > B\).

Proof (Sketch of 2) As in the proof sketch of Theorem 8, (9) is needed for \((\phi_{\gamma}, F)\)-calibration, and \(\phi(\alpha) + \phi(-\alpha)\) “should take larger values in \(|\alpha| \leq \gamma\) than in the rest of \(\alpha\).” Quasiconcavity of \(\phi(\alpha) + \phi(-\alpha)\) naturally implies this property with a non-strict inequality, and the condition \(\phi(\gamma) + \phi(-\gamma) > B\) ensures the strict inequality. Figure 3 illustrates it with the ramp loss.

To the best our knowledge, this is the first characterization of losses calibrated to \(\phi_{\gamma}\). This result is especially interesting when \(F = F_{\text{lin}}\), ensuring that a quasiconcave even surrogate \(\phi\) such that \(\phi(\gamma) + \phi(-\gamma) > B\) is \((\phi_{\gamma}, F_{\text{lin}})\)-calibrated.

We remark that \(\phi(\gamma) + \phi(-\gamma) \geq B\) always holds when \(\phi\) is bounded, nonincreasing, and quasiconcave even (see part 4 of Lemma 13 in Appendix B). The strict inequality \(\phi(\gamma) + \phi(-\gamma) > B\) is necessary and sufficient for \((\phi_{\gamma}, F_{\text{lin}})\)-calibration.

We additionally remark that the ramp loss and the sigmoid loss are \((\phi_{01}, F_{\text{all}})\)-calibrated (Bartlett et al., 2006; Charoenphakdee et al., 2019). Note that these two losses are bounded, nonincreasing, and quasiconcave even, hence \((\phi_{01}, F_{\text{lin}})\)-calibrated.

8. Examples

Several examples of loss functions are shown in Figure 4. For each base surrogate \(\phi\), we consider the shifted surrogate \(\phi_{\beta}(\alpha) \overset{\text{def}}{=} \phi(\alpha - \beta)\) with the horizontal shift parameter \(\beta\). The ramp, sigmoid, modified squared losses are examples of nonconvex and quasiconcave even losses when \(\beta \geq 0\), while the hinge, logistic, and squared losses are examples of convex losses. We show \((\phi_{\gamma}, F_{\text{lin}})\)-calibration functions in this subsection. As a result, we will see that the ramp, sigmoid, and modified squared losses are calibrated with appropriate shift parameters. Detailed derivations of the calibration functions and the proofs of quasiconcavity are deferred to Appendix C.

8.1. Ramp Loss

The ramp loss is \(\phi(\alpha) = \min \{1, \max \{0, 1/2\alpha\}\}\). We consider the shifted ramp loss: \(\phi_{\beta}(\alpha) = \phi(\alpha - \beta) = \min \{1, \max \{0, 1/2(1-\alpha+\beta)\}\}\). The \((\phi_{\gamma}, F_{\text{lin}})\)-calibration function and its Fenchel-Legendre biconjugate of the ramp loss are plotted in Figure 5. We can see that the ramp loss

---

2. We only rely on the fact that \(F_{\text{lin}} \supseteq [-1, 1]\). The results can be extrapolated to \(F\) such that \(F \supseteq [-1, 1]\).
is calibrated wrt $(\phi_\gamma, F_{\text{lin}})$ when $0 < \beta < 2$. Since the ramp loss is quasi-concave even when $\beta \geq 0$, we also observe that the ramp loss is not calibrated when $\beta = 0$ because it is symmetric loss (Charoenphakdee et al., 2019), that is, $\phi_0(\alpha) + \phi_0(-\alpha) = 1$ for all $\alpha \in \mathbb{R}$, which does not satisfy the condition $\phi_0(\gamma) + \phi_0(-\gamma) > B = 1$ in Theorem 2.

8.2. Sigmoid Loss

The sigmoid loss is $\phi(\alpha) = \frac{1}{1 + e^{-\alpha}}$. We consider the shifted sigmoid loss: $\phi_\beta(\alpha) = \frac{1}{1 + e^{-(\alpha - \beta)}}$ for $\beta > 0$. The $(\phi_\gamma, F_{\text{lin}})$-calibration function is plotted in Figure 6. Thus, the sigmoid loss is $(\phi_\gamma, F_{\text{lin}})$-calibrated when $A_1 > 0$, which is equivalent to $\beta > 0$. Again, we observe that the sigmoid loss with $\beta = 0$ is not calibrated in the same way as the ramp loss because it is symmetric.

8.3. Modified Squared Loss

We make a bounded monotone surrogate $\phi(\alpha) = \text{clip}_{[0,1]}(\max\{0, 1 - \alpha\}^2)$ by modifying the squared loss, where $\text{clip}_{[a,b]}(\cdot)$ clips values outside the interval $[a, b]$, and consider the shifted version $\phi_\beta(\alpha) \overset{\text{def}}{=} \phi(\alpha - \beta)$. The $(\phi_\gamma, F_{\text{lin}})$-calibration function and its Fenchel-Legendre biconjugate are plotted in Figure 7. We can deduce that the modified squared loss is calibrated wrt $(\phi_\gamma, F_{\text{lin}})$ for all $0 \leq \beta < 1$. In contrast to the proceeding examples, the modified squared loss is not symmetric.

Moreover, the modified squared loss is $(\phi_\gamma, F_{\text{lin}})$-calibrated even if $\phi_\beta$ for $\beta < 0$ is not a quasi-concave even loss. We plot two examples in Figure 8. As seen in the proof sketch of Theorem 11, it is crucial that $\phi_\beta(\alpha) + \phi_\beta(-\alpha)$ takes higher values in $|\alpha| \leq \gamma$ than in $|\alpha| > \gamma$. The modified squared loss with $-1 + \frac{1}{\sqrt{2}} < \beta < 0$ satisfies this property (see Figure 9).
8.4. Hinge Loss and Squared Loss

Here we consider the shifted hinge loss \( \phi_\beta(\alpha) = \max\{0, 1 - \alpha + \beta\} \), and the shifted squared loss \( \phi_\beta(\alpha) = (1 - \alpha + \beta)^2 \) as examples of convex losses. Their \( (\phi_\gamma, F_{\text{lin}}) \)-calibration functions are plotted in Figures 10 and 11, respectively, which tell us that the hinge and squared losses are not \( (\phi_\gamma, F_{\text{lin}}) \)-calibrated. This result aligns with Theorem 8.

9. Simulation

Learning Curve on Synthetic Data. We generate positive and negative data from \( N([2 2]^{\top}, I_2) \) and \( N([-2 2]^{\top}, I_2) \), respectively, and normalize with the maximum \( \ell_2 \)-norm among all data points. This ensures that data points lie in the \( \ell_2 \) unit ball. We generate 800 training and 200 test points.

Linear models \( f(x) = \theta^{\top} x + \theta_0 \) are used, where \( \theta \) and \( \theta_0 \) are learnable parameters. As surrogate losses, we use the ramp, sigmoid, logistic, and hinge losses, with shift parameter \( \beta = 0.2 \). Batch gradient descent with the fixed step size 0.1 is used in optimization, and 1,000 steps are run for each trial. After every parameter update, the parameters are normalized to ensure \( \|\theta \theta_0^{\top}\|_2 = 1 \).

The robust 0-1 loss with \( \gamma = 0.2 \) is used as the target loss. To compute the excess risk, the Bayes risk for each surrogate loss and the robust 0-1 loss is numerically computed. The detail of numerical approximation of the Bayes risks is explained in Appendix D. The surrogate and target excess risks are shown in Figure 12. 20 trials are run for each data realization.

As you can see from Figure 12, optimization trajectories of calibrated surrogates (ramp and sigmoid) have target excess risks close to zero, while those of convex surrogates (logistic and hinge)
Figure 10: The calibration function of the hinge loss.

Figure 11: The calibration function of the squared loss. \( \eta_0 \overset{\text{def}}{=} \frac{1 + \gamma + \beta}{2(1 + \beta)}, \eta_2 \overset{\text{def}}{=} \frac{2 + \beta}{2(1 + \beta)}, \) and \( \eta_1 \overset{\text{def}}{=} \frac{\eta_0 + \eta_2}{2} \).

Table 1: The simulation results of the \( \gamma \)-adversarially robust 0-1 loss with \( \gamma = 0.1 \) and \( \beta = 0.5 \). 50 trials are conducted for each pair of a method and dataset. Standard errors (multiplied by \( 10^4 \)) are shown in parentheses. Bold-faces indicate outperforming methods, chosen by one-sided t-test with the significant level 5%.

<table>
<thead>
<tr>
<th></th>
<th>Ramp</th>
<th>Sigmoid</th>
<th>Hinge</th>
<th>Logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 vs 1</td>
<td>0.034 (3)</td>
<td>0.017 (2)</td>
<td>0.087 (12)</td>
<td>0.321 (19)</td>
</tr>
<tr>
<td>0 vs 2</td>
<td>0.111 (7)</td>
<td>0.133 (10)</td>
<td>0.109 (8)</td>
<td>0.281 (19)</td>
</tr>
<tr>
<td>0 vs 3</td>
<td>0.107 (7)</td>
<td>0.126 (8)</td>
<td>0.120 (9)</td>
<td>0.307 (18)</td>
</tr>
<tr>
<td>0 vs 4</td>
<td>0.069 (6)</td>
<td>0.093 (12)</td>
<td>0.072 (7)</td>
<td>0.269 (21)</td>
</tr>
<tr>
<td>0 vs 5</td>
<td>0.233 (21)</td>
<td>0.340 (25)</td>
<td>0.233 (21)</td>
<td>0.269 (16)</td>
</tr>
<tr>
<td>0 vs 6</td>
<td>0.129 (8)</td>
<td>0.167 (13)</td>
<td>0.127 (8)</td>
<td>0.287 (22)</td>
</tr>
<tr>
<td>0 vs 7</td>
<td>0.067 (6)</td>
<td>0.073 (6)</td>
<td>0.090 (9)</td>
<td>0.302 (18)</td>
</tr>
<tr>
<td>0 vs 8</td>
<td>0.096 (7)</td>
<td>0.123 (12)</td>
<td>0.100 (9)</td>
<td>0.263 (20)</td>
</tr>
<tr>
<td>0 vs 9</td>
<td>0.082 (6)</td>
<td>0.101 (8)</td>
<td>0.092 (8)</td>
<td>0.279 (22)</td>
</tr>
</tbody>
</table>

Figure 12: 20 trials of optimization trajectories are shown with standard errors. The horizontal (vertical, resp.) axis shows surrogate excess risk (excess risk of the robust 0-1 loss, resp.) on test data.

fail. This observation supports our theoretical findings in Theorems 8 and 11. Different values of \( \beta \) were tried for the hinge and logistic losses, but the conclusions are not affected.

**Benchmark Data.** We compare the ramp, sigmoid, hinge, and logistic losses on MNIST. The results are shown in Table 1, where we see that nonconvex losses, especially the ramp loss, outperform convex losses in terms of the robust 0-1 loss. Details and full results appear in Appendix D.

**10. Conclusion**

Calibration analysis was leveraged to analyze the adversarially robust 0-1 loss. We found that no convex surrogate loss is calibrated wrt the adversarially robust 0-1 loss. We also established necessary and sufficient conditions for a certain class of nonconvex surrogate losses to be calibrated wrt the adversarially robust 0-1 loss, which includes shifted versions of the ramp and sigmoid losses. An important open problem is to extend our calibration results to nonlinear classifier models.

**Acknowledgments**

HB was supported by JST ACT-I Grant Number JPMJPR18UI. CS was supported in part by NSF Grant Number 1838179. MS was supported by JST CREST Grant Number JPMJCR18A2.
References


Appendix A. Convex and Quasiconvex Analysis

This section summarizes basic tools for convex and quasiconvex analysis.

**Quasiconvex function:** A function $h : S \to \mathbb{R}$ on a (finite-dimensional) vector space $S$ is said to be quasiconvex if for all $x, y \in S$ and $\lambda \in [0, 1]$, $h(\lambda x + (1 - \lambda) y) \leq \max\{h(x), h(y)\}$. A function $h$ is said quasiconcave if $-h$ is quasiconvex: For all $x, y \in S$ and $\lambda \in [0, 1]$, $h(\lambda x + (1 - \lambda) y) \geq \min\{h(x), h(y)\}$. Intuitively, quasiconvexity relaxes convexity in that a function still preserves ‘unimodality’ though it loses definite curvature. There is an equivalent definition (here we only show for quasiconcavity): $h$ is quasiconcave if every superlevel set $\{x \mid h(x) \geq t\}$ for $t \in \mathbb{R}$ is a convex set (Boyd and Vandenberghe, 2004).

**Subderivative:** In order to analyze convexity and quasiconvexity, subderivative is a useful tool. We adopt the Clarke definition of subderivative (Clarke, 1990; Aussel et al., 1994). Let $S^*$ be the dual space of $S$ and $\langle \cdot, \cdot \rangle$ be the dual pairing. The (Clarke) subderivative of a lower semicontinuous function $h$ is the operator $\partial h : S \to S^*$ defined for each $x \in S$ such that

$$\partial h(x) \overset{\text{def}}{=} \{x^* \in S^* \mid \langle x^*, x \rangle \leq h^\circ(x; v) \quad \forall v \in S\},$$

where $h^\circ(x; v)$ is the Rockafellar directional derivative (see Clarke (1990) and Aussel et al. (1994) for the formal definition). When $h$ is locally Lipschitz at $x \in S$, Clarke (1990) states that this is equivalent to $\partial h(x) = \text{co}\{\lim \nabla f(x_i) \mid x_i \to x, x_i \notin \mathcal{Y} \cup \Omega_h\}$, where $\text{co}$ is the convex hull, $\mathcal{Y}$ is any set of measure zero, and $\Omega_h$ is the set of points where $h$ is non-differentiable. In the case $S = \mathbb{R}$, this simply reduces to $\partial h(x) = [\partial_+ h(x), \partial_- h(x)]$, where $\partial_+ h$ and $\partial_- h$ are the right-/left-derivatives of $h$, respectively.

**Properties of subderivative:** Several basic properties of subderivatives are shown in Clarke (1990, Section 2.3) such as $\partial(\lambda h)(x) = \lambda \partial h(x)$ (scalar multiples), $\partial(\sum h_i)(x) \subseteq \sum \partial h_i(x)$ (finite sums), and $0 \nsubseteq \partial h(x)$ if $h$ attains a local extrema at $x$. When $h$ is locally Lipschitz, it clearly holds that $\partial h(x) = \{h'(x)\}$ if $h$ is differentiable at $x$.

**Operator monotonicity:** Convex smooth functions have monotonically nondecreasing derivatives. This can be extended to non-smooth functions via subderivatives. Let $h : S \to \mathbb{R}$ be a lower semicontinuous function. Then $h$ is convex if and only if $\partial h : S \to S^*$ is a monotone operator (Aussel et al., 1994), that is, $\langle y_s - x_s, y - x \rangle \geq 0$ for all $x, y \in \text{dom}(h)$ and $x_s \in \partial h(x), y_s \in \partial h(y)$. In addition, $h$ is quasiconvex if and only if $\partial h$ is a quasimonotone operator (Aussel et al., 1994), that is, $\langle x_s, y - x \rangle > 0 \implies \langle y_s, y - x \rangle \geq 0$ for all $x, y \in \text{dom}(h)$ and $x_s \in \partial h(x), y_s \in \partial h(y)$.

---

3. For two vector spaces $U$ and $V$ over the same field $F$ and a bilinear map $\langle \cdot, \cdot \rangle : U \times V \to F$, we say a triple $(U, V, \langle \cdot, \cdot \rangle)$ is a dual pair if there exists $v \in V$ such that $\langle u, v \rangle \neq 0$ for all $u \in U$ and there exists $u \in U$ such that $\langle u, v \rangle \neq 0$ for all $v \in V$. Here, $V$ is called a dual space of $V$, and $\langle \cdot, \cdot \rangle$ is called a dual pairing.
Appendix B. Deferred Proofs

B.1. Proof of Proposition 1

Proof. Fix \((x, y) \in \mathcal{X} \times \mathcal{Y}\) and \(f \in \mathcal{F}_{\text{lin}}\) associated with parameter \(\theta \in \mathbb{R}^d\).

When \(y = +1\), we divide into three cases depending on the value of \(y f(x) = \theta^T x\). If \(\theta^T x \leq 0\), then \(\Delta_x = 0\) simply gives \(\theta^T (x + \Delta_x) \leq 0\). If \(0 < \theta^T x \leq \gamma\), fix \(\Delta_x = -\gamma \theta \in B^d_{\|\theta\|_2}(\gamma)\). Then, \(\theta^T (x + \Delta_x) = \theta^T x - \gamma \leq 0\). If \(\theta^T x > \gamma\), we observe \(\theta^T \Delta_x\) is minimized by \(\Delta_x = -\frac{\gamma}{\|\theta\|_2} \theta \in B^d_{\|\theta\|_2}(\gamma)\). Then, \(\theta^T (x + \Delta_x) > \gamma + \theta^T \Delta_x \geq \gamma - \gamma = 0\). In all cases, \(\ell_\gamma(1, x, f) = 1 \{f(x) \leq \gamma\}\).

When \(y = -1\), we divide the cases as well. If \(\theta^T x > 0\), then \(\Delta_x = 0\) simply gives \(\theta^T (x + \Delta_x) > 0\). If \(-\gamma \leq \theta^T x \leq 0\), fix \(\Delta_x = \gamma \theta \in B^d_{\|\theta\|_2}(\gamma)\). Then, \(\theta^T (x + \Delta_x) = \theta^T x + \gamma \geq 0\). If \(\theta^T x < -\gamma\), we observe \(\theta^T \Delta_x\) is maxmized by \(\Delta_x = \frac{-\gamma}{\|\theta\|_2} \theta \in B^d_{\|\theta\|_2}(\gamma)\). Then, \(\theta^T (x + \Delta_x) < -\gamma + \theta^T \Delta_x \leq -\gamma + \gamma = 0\). In all cases, \(\ell_\gamma(-1, x, f) = 1 \{f(x) \geq -\gamma\} = 1 \{-f(x) \leq \gamma\}\).

B.2. Proof of Lemma 7

Proof. We first simplify the constraint in the calibration function (6). The \(\phi_\gamma\)-CCR for \(\alpha \in \mathcal{A}_{\mathcal{F}}\) is

\[
C_{\phi_\gamma}(\alpha, \eta) = \eta \mathbb{1}_{\{\alpha \leq \gamma\}} + (1 - \eta) \mathbb{1}_{\{\alpha \geq -\gamma\}} = \begin{cases} 1 & \text{if } |\alpha| \leq \gamma, \\ 1 - \eta & \text{if } \gamma < |\alpha|, \\ \eta & \text{if } \alpha < -\gamma. \end{cases}
\]

and the minimal \((\phi_\gamma, \mathcal{F})\)-CCR is \(C^*_{\phi_\gamma, \mathcal{F}}(\eta) = \min\{\eta, 1 - \eta\}\). If \(\gamma < |\alpha|\), a well-known algebra in the binary classification case shows that \(C_{\phi_\gamma}(\alpha, \eta) - C^*_{\phi_\gamma, \mathcal{F}}(\eta) = 2|\eta| - 1 \cdot \mathbb{1}_{\{2|\eta| - 1 \alpha \leq 0\}}\) (see, e.g., Bartlett et al. (2006, Proof of Theorem 3)). If \(|\alpha| \leq \gamma\), it follows that \(C_{\phi_\gamma}(\alpha, \eta) - C^*_{\phi_\gamma, \mathcal{F}}(\eta) = 1 - \min\{\eta, 1 - \eta\} = \max\{\eta, 1 - \eta\}\). Hence,

\[
\Delta C_{\phi_\gamma, \mathcal{F}}(\alpha, \eta) = \begin{cases} \max\{\eta, 1 - \eta\} & \text{if } |\alpha| \leq \gamma, \\ |2\eta - 1| \cdot \mathbb{1}_{\{2\eta - 1 \alpha \leq 0\}} & \text{if } \gamma < |\alpha|. \end{cases}
\]

Next, we simplify the inner infimum on \(\alpha\), \(\inf_{\alpha \in \mathbb{R}} \{\Delta C_{\phi_\gamma, \mathcal{F}}(\alpha, \eta) \mid \Delta C_{\phi_\gamma, \mathcal{F}}(\alpha, \eta) \geq \varepsilon\} = \bar{\delta}(\varepsilon, \eta)\) in (6), for a fixed \(\eta \in [0, 1]\). If \(\varepsilon > \max\{\eta, 1 - \eta\}\), no \(\alpha \in \mathcal{A}_{\mathcal{F}}\) achieves \(\Delta C_{\phi_\gamma, \mathcal{F}}(\alpha, \eta) \geq \varepsilon\), meaning that \(\bar{\delta}(\varepsilon, \eta) = \infty\). If \(|2\eta - 1| < \varepsilon \leq \max\{\eta, 1 - \eta\}\), \(\Delta C_{\phi_\gamma, \mathcal{F}}(\alpha, \eta) \geq \varepsilon\) is achieved when \(|\alpha| \leq \gamma\). Hence, \(\bar{\delta}(\varepsilon, \eta) = \inf_{\alpha} \{\Delta C_{\phi_\gamma, \mathcal{F}}(\alpha, \eta) \mid |\alpha| \leq \gamma\}\). Note that \(|2\eta - 1| \leq \max\{\eta, 1 - \eta\} = \frac{1+|2\eta-1|}{2}\) for all \(\eta \in [0, 1]\). If \(\varepsilon \leq |2\eta - 1|\), \(\Delta C_{\phi_\gamma, \mathcal{F}}(\alpha, \eta) \geq \varepsilon\) is achieved if either \(|\alpha| \leq \gamma\) or \((2\eta - 1)\alpha \leq 0\) holds. Hence, \(\bar{\delta}(\varepsilon, \eta) = \inf_{\alpha} \{\Delta C_{\phi_\gamma, \mathcal{F}}(\alpha, \eta) \mid |\alpha| \leq \gamma\) or \((2\eta - 1)\alpha \leq 0\}\). These verify the statement of this lemma.

B.3. Useful Lemmas

The following lemmas are useful in the remaining proofs. Their proofs appear in Sections B.6 and B.7.

Lemma 12. Let \(\phi : \mathbb{R} \to \mathbb{R}_{\geq 0}\) be a margin-based loss function and \(\mathcal{F} \subseteq \mathbb{R}^X\) be a symmetric function class such that \(\mathcal{A}_{\mathcal{F}} \supseteq [-1, 1]\).
1. For all $\alpha \in \mathbb{R}$, $C_{\phi}(\alpha, \eta)$ and $\Delta C_{\phi,F}(\alpha, \eta)$ are symmetric about $\eta = \frac{1}{2}$, i.e., $C_{\phi}(\alpha, \eta) = C_{\phi}(-\alpha, 1 - \eta)$ and $\Delta C_{\phi,F}(\alpha, \eta) = \Delta C_{\phi,F}(-\alpha, 1 - \eta)$ for all $\eta \in [0, 1]$.

2. When $\eta = \frac{1}{2}$, we have

$$\inf_{|\alpha| \leq \gamma} \Delta C_{\phi,F}(\alpha, \frac{1}{2}) = \inf_{0 \leq \alpha \leq \gamma} \Delta C_{\phi,F}(\alpha, \frac{1}{2}) = \inf_{0 \leq \alpha \leq \gamma} C_{\phi}(\alpha, \frac{1}{2}) - \inf_{\alpha \in A_F : \alpha \geq 0} C_{\phi}(\alpha, \frac{1}{2})$$

3. A surrogate loss $\phi$ is calibrated wrt $(\phi, F)$ if and only if

$$\inf_{|\alpha| \leq \gamma} C_{\phi}(\alpha, \frac{1}{2}) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \frac{1}{2}) \quad \text{and} \quad \inf_{\alpha \in A_F : \alpha \leq \gamma} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta)$$

for all $\eta \in \left(\frac{1}{2}, 1\right]$.

4. A surrogate loss $\phi$ is calibrated wrt $(\phi_{01}, F)$ if and only if

$$\inf_{\alpha \in A_F : \alpha \leq 0} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta)$$

for all $\eta \in \left(\frac{1}{2}, 1\right]$.

Note that part 4 of Lemma 12 can be regarded as a generalization of classification calibration (Bartlett et al., 2006, Definition 1) when $F \neq F_{all}$.

**Lemma 13** Let $\phi : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ be a margin-based loss function. If $\phi$ is nonincreasing, bounded, $\phi \not\equiv 0$, and quasiconcave even, then

1. the class-conditional $\phi$-risk $C_{\phi}(\alpha, \eta)$ is quasiconcave in $\alpha \in \mathbb{R}$ for all $\eta \in [0, 1]$.

2. for all $\eta \in \left(\frac{1}{2}, 1\right]$, $C_{\phi}(\alpha, \eta)$ is nonincreasing in $\alpha$ when $\alpha \geq 0$.

3. for all $\eta \in \left(\frac{1}{2}, 1\right]$, $C_{\phi}(-1, \eta) > C_{\phi}(1, \eta)$.

4. $\phi(\alpha) + \phi(-\alpha)$ is nonincreasing in $\alpha$ when $\alpha \geq 0$.

5. for $l, u \in [-1, 1]$ ($l \leq u$), $\inf_{\alpha \in [l, u]} C_{\phi}(\alpha, \eta) = \min\{C_{\phi}(l, \eta), C_{\phi}(u, \eta)\}$ for all $\eta \in [0, 1]$.

**B.4. Proof of Theorem 8**

**Proof** Part 3 of Lemma 12 states that $\phi$ is calibrated wrt $(\phi, F)$ if and only if

$$\inf_{0 \leq \alpha \leq \gamma} C_{\phi}(\alpha, \frac{1}{2}) > \inf_{\alpha \in A_F : \alpha \geq 0} C_{\phi}(\alpha, \frac{1}{2}) \quad \text{and} \quad \inf_{\alpha \in A_F : \alpha \leq \gamma} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F : \alpha \geq 0} C_{\phi}(\alpha, \eta) \quad \text{for any } \eta \in \left(\frac{1}{2}, 1\right].$$
In order to show $\phi$ is not calibrated wrt $(\phi, F)$, it is sufficient to show that
\[
\inf_{0 \leq \alpha \leq \gamma} C_\phi(\alpha, \frac{1}{2}) = \inf_{\alpha \in A_F : \alpha \geq 0} C_\phi(\alpha, \frac{1}{2}),
\]
which is equivalent to
\[
\inf_{0 \leq \alpha \leq \gamma} \phi(\alpha) + \phi(-\alpha) = \inf_{\alpha \in A_F : \alpha \geq 0} \phi(\alpha) + \phi(-\alpha).
\]
Since $\bar{\phi}(\alpha) \overset{\text{def}}{=} \phi(\alpha) + \phi(-\alpha)$ is a convex even function, we have $\bar{\phi}(0) \leq \bar{\phi}(\alpha)$ for all $\alpha \in A_F$. Too see this, assume that there exists $\alpha_* \in A_F$ such that $\alpha_* \neq 0$ and $\bar{\phi}(0) > \bar{\phi}(\alpha_*)$. Then, we also have $\bar{\phi}(-\alpha_*) < \bar{\phi}(0)$ since $\bar{\phi}$ is even. It follows that $\frac{1}{2}\{\bar{\phi}(-\alpha_*) + \bar{\phi}(\alpha_*)\} < \bar{\phi}(0)$. However, we have $\frac{1}{2}\{\bar{\phi}(-\alpha_*) + \bar{\phi}(\alpha_*)\} \geq \bar{\phi}\left(\frac{-\alpha_* + \alpha_*}{2}\right) = \bar{\phi}(0)$ because of convexity of $\bar{\phi}$. Hence, we see $\bar{\phi}(0) \leq \bar{\phi}(\alpha)$ for all $\alpha \in A_F$. This means that $\inf_{0 \leq \alpha \leq \gamma} \phi(\alpha) = \inf_{\alpha \in A_F : 0 \leq \alpha} \phi(\alpha) = \phi(0)$.  

B.5. Proof of Theorem 11  

Proof of part 1 By part 4 of Lemma 12, $(\phi_{01}, F)$-calibration is equivalent to
\[
\inf_{-1 \leq \alpha \leq 0} C_\phi(\alpha, \eta) > \inf_{\alpha \in [-1, 1]} C_\phi(\alpha, \eta) \quad \text{for all } \eta \in \left(\frac{1}{2}, 1\right].
\]

First, we observe
\[
2\phi(0) = \phi(0) + \phi(0) \geq \inf_{0 \leq \alpha \leq 1} \phi(\alpha) + \phi(-\alpha)
\]
\[
= \phi(1) + \phi(-1) \quad \text{(quasiconcavity of } \phi(\alpha) + \phi(-\alpha))
\]
\[
= B.
\]

Next, fix $\eta$ such that $\frac{1}{2} < \eta \leq 1$. We observe with part 5 of Lemma 13 that
\[
\inf_{-1 \leq \alpha \leq 0} C_\phi(\alpha, \eta) = \min\{C_\phi(-1, \eta), C_\phi(0, \eta)\} \quad \text{(part 5 of Lemma 13)}
\]
\[
= \min\{\eta B + (1 - \eta) B, \phi(0)\},
\]
and
\[
\inf_{-1 \leq \alpha \leq 1} C_\phi(\alpha, \eta) = \min\{C_\phi(-1, \eta), C_\phi(1, \eta)\} \quad \text{(part 5 of Lemma 13)}
\]
\[
= C_\phi(1, \eta) = \eta B + (1 - \eta) B,
\]
where $B \overset{\text{def}}{=} \phi(1)$ and $\overline{B} = \phi(-1)$. Here,
\[
C_\phi(-1, \eta) - C_\phi(1, \eta) = (\overline{B} - B)(2\eta - 1) > 0,
\]
\[
C_\phi(0, \eta) - C_\phi(1, \eta) = \phi(0) - \overline{B} + \eta(\overline{B} - B)
\]
\[
\geq \frac{\overline{B} + B}{2} - \overline{B} + \eta(\overline{B} - B) \quad (2\phi(0) \geq B)
\]
\[
> \frac{\overline{B} + B}{2} - \overline{B} + \frac{\overline{B} - B}{2} \quad (\phi(-1) > \phi(1) \text{ and } \eta > \frac{1}{2})
\]
\[
= 0.
\]
Then, we have for all \( \eta \in \left( \frac{1}{2}, 1 \right] \),
\[
\inf_{-1 \leq \alpha \leq 0} C_\phi(\alpha, \eta) - \inf_{-1 \leq \alpha \leq 1} C_\phi(\alpha, \eta) = \min\{C_\phi(-1, \eta) - C_\phi(1, \eta), C_\phi(0, \eta) - C_\phi(1, \eta)\} > 0.
\]

This verifies the condition (10).

**Proof of part 2** \( \phi \) is calibrated wrt \((\phi, \mathcal{F})\) if and only if
\[
\begin{align*}
(\text{i}) \quad & \inf_{|\alpha| \leq \gamma} C_\phi\left(\alpha, \frac{1}{2}\right) > \inf_{-1 \leq \alpha \leq 1} C_\phi\left(\alpha, \frac{1}{2}\right), \\
(\text{ii}) \quad & \inf_{-1 \leq \alpha \leq \gamma} C_\phi(\alpha, \eta) > \inf_{-1 \leq \alpha \leq 1} C_\phi(\alpha, \eta) \quad \text{for all } \eta \in \left( \frac{1}{2}, 1 \right]
\end{align*}
\]  

(11)

by part 3 of Lemma 12. Now we show \( \phi(\gamma) + \phi(-\gamma) > B \) assuming (i) and (ii).

\[
\phi(\gamma) + \phi(-\gamma) = \inf_{0 \leq \alpha \leq \gamma} \phi(\alpha) + \phi(-\alpha) \quad \text{(part 4 of Lemma 13)}
\]
\[
> \inf_{-1 \leq \alpha \leq 1} \phi(\alpha) + \phi(-\alpha) \quad \text{(i) is used}
\]
\[
= \inf_{0 \leq \alpha \leq 1} \phi(\alpha) + \phi(-\alpha) \quad (\phi(\alpha) + \phi(-\alpha) \text{ is even})
\]
\[
= \phi(1) + \phi(-1) \quad \text{(part 4 of Lemma 13)}
\]
\[
= B.
\]

Conversely, assume \( \phi(\gamma) + \phi(-\gamma) > B \). We will show (i) and (ii) in (11). Since \( \phi(\alpha) + \phi(-\alpha) \) is nonincreasing in \( \alpha \geq 0 \) (part 4 of Lemma 13), we have
\[
\inf_{|\alpha| \leq \gamma} \phi(\alpha) + \phi(-\alpha) = \inf_{0 \leq \alpha \leq \gamma} \phi(\alpha) + \phi(-\alpha) \quad (\phi(\alpha) + \phi(-\alpha) \text{ is even})
\]
\[
= \phi(\gamma) + \phi(-\gamma) \quad \text{(part 4 of Lemma 13)}
\]
\[
> B
\]
\[
= \phi(1) + \phi(-1)
\]
\[
= \inf_{0 \leq \alpha \leq 1} \phi(\alpha) + \phi(-\alpha), \quad \text{(part 4 of Lemma 13)}
\]
\[
= \inf_{-1 \leq \alpha \leq 1} \phi(\alpha) + \phi(-\alpha), \quad (\phi(\alpha) + \phi(-\alpha) \text{ is even})
\]

which is equivalent to (i). For (ii), fix \( \eta \) such that \( \frac{1}{2} < \eta \leq 1 \). We first observe with parts 3 and 5 of Lemma 13 that
\[
\inf_{-1 \leq \alpha \leq \gamma} C_\phi(\alpha, \eta) = \min\{C_\phi(-1, \eta), C_\phi(\gamma, \eta)\},
\]
\[
\inf_{-1 \leq \alpha \leq 1} C_\phi(\alpha, \eta) = \min\{C_\phi(-1, \eta), C_\phi(1, \eta)\} = C_\phi(1, \eta).
\]

Here, we have
\[
C_\phi(1, \eta) = (B - \overline{B})\eta + \overline{B},
\]
\[
C_\phi(\gamma, \eta) = (\phi(\gamma) - \phi(-\gamma))\eta + \phi(-\gamma),
\]
where \( \overline{B} \equiv \phi(1) \) and \( \underline{B} \equiv \phi(-1) \). Observing that
\[
\overline{B} - B + \phi(\gamma) - \phi(-\gamma) \geq \overline{B} - B + \phi(1) - \phi(-1) = 0, \quad (\phi \text{ is nonincreasing})
\]
we have for all \( \eta \in \left( \frac{1}{2}, 1 \right] \),
\[
C_\phi(\gamma, \eta) - C_\phi(1, \eta) = (\phi(\gamma) - \phi(-\gamma) + \overline{B} - B)\eta + (\phi(-\gamma) - \overline{B}) \geq (\phi(\gamma) - \phi(-\gamma) + \overline{B} - B)\frac{1}{2} + \phi(-\gamma) - \overline{B} = \frac{\phi(\gamma) + \phi(-\gamma) - B}{2} > 0,
\]
where the first inequality holds since \( (\phi(\gamma) - \phi(-\gamma) + \overline{B} - B) > 0 \) and \( \eta > \frac{1}{2} \), and the second inequality holds because of the assumption \( \phi(\gamma) + \phi(-\gamma) > B \). In addition, we have \( C_\phi(-1, \eta) > C_\phi(1, \eta) \) for \( \eta > \frac{1}{2} \) by part 3 of Lemma 13. Therefore,
\[
\inf_{-1 \leq \alpha \leq \gamma} C_\phi(\alpha, \eta) - \inf_{-1 \leq \alpha \leq 1} C_\phi(\alpha, \eta) = \min\{C_\phi(-1, \eta) - C_\phi(1, \eta), C_\phi(\gamma, \eta) - C_\phi(1, \eta)\} > 0
\]
holds for all \( \eta \) such that \( \frac{1}{2} < \eta \leq 1 \), and this verifies (ii).

\[\square\]

**B.6. Proof of Lemma 12**

Proof Parts 1 and 2 are obvious from the definition of the class-conditional \( \phi \)-risk.

Part 3: Let \( \delta : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0} \) be the \((\phi_\gamma, \mathcal{F})\)-calibration function of \( \phi \), and \( \delta : \mathbb{R}_{\geq 0} \times [0, 1] \to \mathbb{R}_{\geq 0} \) be the inner infimum of \( \delta \) in (8):
\[
\delta(\varepsilon, \eta) = \begin{cases} 
\inf_{|\alpha| \leq \gamma} \Delta C_{\phi, \mathcal{F}}(\alpha, \eta) & \text{if } |2\eta - 1| < \varepsilon \leq \max\{\eta, 1 - \eta\}, \\
\inf_{\alpha \in \mathcal{A}_F : |\alpha| \leq \gamma \text{ or } (2\eta - 1)\alpha \leq 0} \Delta C_{\phi, \mathcal{F}}(\alpha, \eta) & \text{if } \varepsilon \leq |2\eta - 1|,
\end{cases}
\]
and \( \delta(\varepsilon) = \inf_{\eta \in [0, 1]} \delta(\varepsilon, \eta) \). Then, by Proposition 4, \( \phi \) is \((\phi_\gamma, \mathcal{F})\)-calibrated if and only if \( \delta(\varepsilon) > 0 \) for all \( \varepsilon > 0 \). If \( \delta(\varepsilon, \eta) \) is lower semicontinuous in \( \eta \), this is equivalent to \( \delta(\varepsilon, \eta) > 0 \) for all \( \varepsilon > 0 \) and \( \eta \in [0, 1] \). Using part 1 of Lemma 12 and symmetry of \( \mathcal{F} \), since we have for \( \eta \leq \frac{1}{2} \),
\[
\inf_{\alpha \in \mathcal{A}_F : |\alpha| \leq \gamma \text{ or } (2\eta - 1)\alpha \leq 0} C_\phi(\alpha, \eta) = \inf_{\alpha \in \mathcal{A}_F : \alpha \geq -\gamma} C_\phi(\alpha, \eta) = \inf_{\alpha \in \mathcal{A}_F : \alpha \geq -\gamma} C_\phi(-\alpha, 1 - \eta) \quad \text{(part 1 of Lemma 12)}
\]
and for \( \eta \geq \frac{1}{2} \),
\[
\inf_{\alpha \in \mathcal{A}_F : |\alpha| \leq \gamma \text{ or } (2\eta - 1)\alpha \leq 0} C_\phi(\alpha, \eta) = \inf_{\alpha \in \mathcal{A}_F : \alpha \leq \gamma} C_\phi(\alpha, \eta),
\]
\[
\inf_{\alpha \in A_F : \alpha \leq \gamma} \Delta C_{\phi,F}(\alpha, \eta) > 0 \text{ for all } \eta \geq \frac{1}{2} \implies \inf_{\alpha \in A_F : \alpha \leq \gamma} \Delta C_{\phi,F}(\alpha, \eta) > 0 \text{ for all } \eta \in [0, 1]. \text{ Hence,}
\]

\[
\inf_{\alpha \in A_F : \alpha \leq \gamma} \Delta C_{\phi,F}(\alpha, \eta) > 0
\]

for \(\varepsilon > 0\) and \(\eta \in [0, 1]\) such that \(\varepsilon \leq |2\eta - 1|\) if and only if

\[
\inf_{\alpha \in A_F : \alpha \leq \gamma} \Delta C_{\phi,F}(\alpha, \eta) > 0
\]

for \(\varepsilon > 0\) and \(\eta \in [\frac{1}{2}, 1]\) such that \(\varepsilon \leq 2\eta - 1\).

Therefore, \(\delta(\varepsilon, \eta) > 0\) for all \(\varepsilon > 0\) and \(\eta \in [0, 1]\) if and only if

\[
\begin{cases}
\inf_{|\alpha| \leq \gamma} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta) & \text{for all } \eta \geq \frac{1}{2} \text{ such that } 2\eta - 1 < \varepsilon \leq \eta, \\
\inf_{\alpha \in A_F : \alpha \leq \gamma} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta) & \text{for all } \eta \geq \frac{1}{2} \text{ such that } \varepsilon \leq 2\eta - 1,
\end{cases}
\]

for all \(\varepsilon > 0\), which is equivalent to

\[
\begin{cases}
\inf_{|\alpha| \leq \gamma} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta) & \text{for all } \eta \geq \frac{1}{2} \text{ such that } \varepsilon \leq \eta < \frac{1 + \varepsilon}{2}, \\
\inf_{\alpha \in A_F : \alpha \leq \gamma} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta) & \text{for all } \eta \geq \frac{1}{2} \text{ such that } \frac{1 + \varepsilon}{2} \leq \eta \leq 1,
\end{cases}
\]

for all \(\varepsilon > 0\).

We immediately observe that

\[
\begin{align*}
\{ & \eta \geq \frac{1}{2}, \varepsilon \leq \eta < \frac{1 + \varepsilon}{2}, \varepsilon > 0 \} = \left\{ \frac{1}{2} \leq \eta \leq 1 \right\}, & \text{and} \\
\{ & \eta \geq \frac{1}{2}, \frac{1 + \varepsilon}{2} \leq \eta \leq 1, \varepsilon > 0 \} = \left\{ \frac{1}{2} < \eta \leq 1 \right\}.
\end{align*}
\]

Therefore, we reduce the above conditions as

\[
\begin{cases}
\inf_{|\alpha| \leq \gamma} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta) & \text{if } \frac{1}{2} \leq \eta \leq 1, \\
\inf_{\alpha \in A_F : \alpha \leq \gamma} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta) & \text{if } \frac{1}{2} < \eta \leq 1.
\end{cases}
\]

Note that \(\inf_{|\alpha| \leq \gamma} C_{\phi}(\alpha, \eta) \geq \inf_{\alpha \in A_F : \alpha \leq \gamma} C_{\phi}(\alpha, \eta)\) for all \(\eta\). Since the first case is included in the second case except when \(\eta = \frac{1}{2}\), this is equivalent to

\[
\begin{align*}
\inf_{|\alpha| \leq \gamma} C_{\phi}(\alpha, \frac{1}{2}) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \frac{1}{2}), & \text{ and } \inf_{\alpha \in A_F : \alpha \leq \gamma} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta) \text{ for } \eta \in \left(\frac{1}{2}, 1\right].
\end{align*}
\]

Finally, we check lower semicontinuity of \(\delta(\varepsilon, \eta)\) in \(\eta\). Fix a fixed \(\alpha, C_{\phi}(\alpha, \eta)\) is lower semicontinuous in \(\eta\) since \(C_{\phi}(\alpha, \eta)\) is linear in \(\eta\). Because pointwise infimum preserves lower semicontinuity, \(\inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta)\), \(\inf_{|\alpha| \leq \gamma} C_{\phi}(\alpha, \eta)\), and \(\inf_{\alpha \in A_F : |\alpha| \leq \gamma \text{ or } (2\eta - 1)\alpha \leq 0} C_{\phi}(\alpha, \eta)\) are lower semicontinuous in \(\eta\). Hence, \(\delta(\varepsilon, \eta)\) is lower semicontinuous in \(\eta\). This concludes the proof of part 3.
Part 4: We follow the same direction as part 3. If we take $\gamma \to 0$,

$$\bar{\delta}(\varepsilon, \eta) = \inf_{\alpha \in A_F : (2\eta-1)\alpha \leq 0} \Delta C_{\phi, F}(\alpha, \eta) \quad \text{such that } \varepsilon \leq |2\eta - 1|.$$  

Hence, by Proposition 4 and lower semicontinuity of $\bar{\delta}(\varepsilon, \eta)$ in $\eta$ (proven in part 3), $\phi$ is $(\phi_0, F)$-calibrated if and only if

$$\inf_{\alpha \in A_F : (2\eta-1)\alpha \leq 0} \Delta C_{\phi, F}(\alpha, \eta) > 0$$

for all $\varepsilon > 0$ and $\eta \in [0, 1]$ such that $\varepsilon \leq |2\eta - 1|$. In the same way as part 3 of Lemma 12, this is equivalent to

$$\inf_{\alpha \in A_F : \alpha \leq 0} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta) \quad \text{for all } \eta \geq \frac{1}{2} \text{ such that } \frac{1 + \varepsilon}{2} \leq \eta \leq 1,$$

for all $\varepsilon > 0$, by using part 1 of Lemma 12 and symmetry of $F$. In the same way as part 3 of Lemma 12, simple observations on ranges $\varepsilon$ and $\eta$ reduce the above joint conditions on $\varepsilon$ and $\eta$ to $\eta$ alone:

$$\inf_{\alpha \in A_F : \alpha \leq 0} C_{\phi}(\alpha, \eta) > \inf_{\alpha \in A_F} C_{\phi}(\alpha, \eta) \quad \text{for all } \eta \text{ such that } \frac{1}{2} < \eta \leq 1.$$

This is the lemma statement. \hfill \blacksquare

B.7. Proof of Lemma 13

Denote $\bar{\phi}(\alpha) \overset{\text{def}}{=} \phi(\alpha) + \phi(-\alpha)$. $\bar{\phi}$ is quasiconcave and even. To prove part 1, we use the following lemmas.

Lemma 14 A function $h : \mathbb{R} \to \mathbb{R}$ is quasiconcave if and only if $\min\{x_1, x_2, (x_2 - x_1)\} \leq 0$ for all $x_1, x_2 \in \text{dom}(h)$, and $x_1, x_2 \in \partial h(x_1)$ and $x_2, x_1 \in \partial h(x_2)$.

Proof If $h$ is quasimonotone, Theorem 4.1 in Aussel et al. (1994) implies that $-\partial h$ is a quasimonotone operator, i.e.,

$$x_1, x_2, (x_2 - x_1) < 0 \implies x_2, x_1, (x_1 - x_2) \leq 0$$

for all $x_1, x_2 \in \text{dom}(h)$ and $x_1, x_2 \in \partial h(x_1)$ and $x_2, x_1 \in \partial h(x_2)$.

This is clearly equivalent to $\min\{x_1, x_2, (x_2 - x_1)\} \leq 0$. \hfill \blacksquare

Lemma 15 Any element in $\partial \bar{\phi}(\alpha_0)$ can be represented by $\alpha^\ast_+ - \alpha^\ast_-$ for some $\alpha^\ast_+ \in \partial \phi(\alpha_0)$ and $\alpha^\ast_- \in \partial \phi(-\alpha_0)$. For any $\eta \in \left[\frac{1}{2}, 1\right]$, $\alpha^\ast_+ \in \partial \phi(\alpha_0)$, and $\alpha^\ast_- \in \partial \phi(-\alpha_0)$, if $\alpha^\ast_+ - \alpha^\ast_- \in \partial \bar{\phi}(\alpha_0)$, then $\eta \alpha^\ast_+ - (1 - \eta) \alpha^\ast_- \in \partial C_{\phi}(\alpha_0, \eta)$.
Proof By calculus of subderivative, we have \( \partial \phi'(\alpha_0) \subseteq \partial \phi(\alpha_0) - \partial \phi(-\alpha_0) \). The first statement follows from this fact. In order to prove the second statement, note that left-/right-derivatives of \( \phi \) exists because \( \phi \) is nonincreasing. We first observe that

\[
\begin{align*}
(i) \quad & \partial_- \phi(\alpha_0) \leq \alpha_+ \leq \partial_+ \phi(\alpha_0), \\
(ii) \quad & \partial_- \phi(-\alpha_0) \leq \alpha_- \leq \partial_+ \phi(-\alpha_0), \text{ and} \\
(iii) \quad & \partial_- \phi(\alpha_0) - \partial_- \phi(-\alpha_0) \leq \alpha_+ - \alpha_- \leq \partial_+ \phi(\alpha_0) - \partial_+ \phi(-\alpha_0).
\end{align*}
\]

We have (iii) because

\[
\alpha_+ - \alpha_- \in \partial \phi(\alpha_0) = [\partial_- \phi(\alpha_0), \partial_+ \phi(\alpha_0)] = [\partial_- \phi(\alpha_0) - \partial_- \phi(-\alpha_0), \partial_+ \phi(\alpha_0) - \partial_+ \phi(-\alpha_0)].
\]

Then, for \( \eta \in [\frac{1}{2}, 1] \), \( (\eta - \frac{1}{2}) \times (i) + (\eta - \frac{1}{2}) \times (ii) + \frac{1}{2} \times (iii) \) gives

\[
\eta \partial_- \phi(\alpha_0) - (1 - \eta) \partial_- \phi(-\alpha_0) \leq \eta \alpha_+ - (1 - \eta) \alpha_+ \leq \eta \partial_- \phi(\alpha_0) - (1 - \eta) \partial_+ \phi(-\alpha_0),
\]

which is equivalent to

\[
\eta \alpha_+ - (1 - \eta) \alpha_- \in \eta \partial_- \phi(\alpha_0) - (1 - \eta) \partial_- \phi(-\alpha_0), \eta \partial_+ \phi(\alpha_0) - (1 - \eta) \partial_+ \phi(-\alpha_0)
\]

\[
= [\partial_- C_\phi(\alpha_0, \eta), \partial_+ C_\phi(\alpha_0, \eta)]
\]

\[
= \partial C_\phi(\alpha_0, \eta).
\]

Now we proceed with the proof of Lemma 13.

Proof (of Lemma 13)

Part 1: Fix \( \eta \in \left[ \frac{1}{2}, 1 \right] \). Since \( \tilde{\phi} \) is quasiconcave, by Lemma 14, \( \min \{\alpha_1, \alpha_2, (\alpha_2 - \alpha_1)\} \leq 0 \) for any \( \alpha_1, \alpha_2 \in [-1, 1] \) and \( \alpha_1 \in \partial \phi(\alpha_1) \) and \( \alpha_2 \in \partial \phi(\alpha_2) \). Let us fix \( \alpha_1, \alpha_2 \in \mathbb{R} \) such that \( \alpha_1 \geq \alpha_2 \), which can be assumed without loss of generality. Since \( \partial C_\phi(\alpha, \eta) \subseteq \eta \partial \phi(\alpha) - (1 - \eta) \partial \phi(-\alpha) \) (the subdifferentiation is taken on \( \alpha \)) and \( \partial \phi(\alpha) \subseteq \partial \phi(\alpha) - \partial \phi(-\alpha) \), we can pick \( \alpha_+ \in \partial \phi(\alpha_1) \) and \( \alpha_- \in \partial \phi(-\alpha_1) \) such that \( \alpha_+ - \alpha_- \in \partial \phi(\alpha_1) \) and \( \eta \alpha_+ - (1 - \eta) \alpha_- \in \partial C_\phi(\alpha_1, \eta) \) by Lemma 15; in the same way, we can pick \( \alpha_+ \in \partial \phi(\alpha_2) \) and \( \alpha_- \in \partial \phi(-\alpha_2) \) such that \( \alpha_+ - \alpha_- \in \partial \phi(\alpha_2) \) and \( \eta \alpha_+ - (1 - \eta) \alpha_- \in \partial C_\phi(\alpha_2, \eta) \).

Here, let us divide \( \min \{\alpha_+ - \alpha_-, (\alpha_1 - \alpha_2), (\alpha_2 - \alpha_1)\} \leq 0 \), the necessary condition of quasiconcavity of \( \phi \), into two cases. Consider the case \( (\alpha_1 - \alpha_-) \leq 0 \) first. In this case,

\[
\begin{align*}
\{\eta \alpha_1 - (1 - \eta) \alpha_2\} = \eta \alpha_1 - (1 - \eta) \alpha_2 \\
\leq \eta \alpha_1 - (1 - \eta) \alpha_2 \\
= (2\eta - 1)(\alpha_1 - \alpha_2) \alpha_2 \leq 0,
\end{align*}
\]

where the last inequality holds because \( \phi \) is nonincreasing (hence \( \alpha_+ \leq 0 \)). Note again that \( \eta \alpha_+ - (1 - \eta) \alpha_- \in \partial C_\phi(\alpha_1, \eta) \).

26
In another case \((\alpha_2^+ - \alpha_2^-)(\alpha_2 - \alpha_1) \leq 0\), we can show \(\{\eta\alpha_2^+ - (1 - \eta)\alpha_2^-\}(\alpha_2 - \alpha_1) \leq 0\) in the same way, and note that \(\eta\alpha_2^+ - (1 - \eta)\alpha_2^- \in \partial C_\phi(\alpha_2, \eta)\). Since we take \(\alpha_1\) and \(\alpha_2\) arbitrarily, now we have \(\min\{\alpha_1^\eta(\alpha_1 - \alpha_2), \alpha_2^\eta(\alpha_2 - \alpha_1)\} \leq 0\) for all \(\alpha_1^\eta \in \partial C_\phi(\alpha_1, \eta)\) and \(\alpha_2^\eta \in \partial C_\phi(\alpha_2, \eta)\). This is the sufficient condition of quasiconcavity of \(C_\phi(\alpha, \eta)\) in \(\alpha \in \mathbb{R}\) by Lemma 14. Therefore, we confirm quasiconcavity of \(C_\phi(\alpha, \eta)\) in \(\alpha \in \frac{1}{2}\) given quasiconcavity of \(\phi\).

Finally, if \(\eta \in [0, \frac{1}{2}]\), we know by part 1 of Lemma 12 that \(C_\phi(\alpha, \eta) = C_\phi(\alpha, 1 - \eta)\) for \(\alpha \in \mathbb{R}\). Then, quasiconcavity of \(C_\phi(\alpha, \eta)\) in \(\alpha \in \mathbb{R}\) follows since \(C_\phi(\alpha, 1 - \eta)\) is quasiconcave in \(-\alpha \in \mathbb{R}\).

**Part 2:** Fix a \(\eta \in \left(\frac{1}{2}, 1\right]\) and \(\alpha_1, \alpha_2 \geq 0\) such that \(\alpha_1 < \alpha_2\). By the fact that \(\phi\) is nonincreasing, we have

\[
\phi(\alpha_1) - \phi(-\alpha_1) - \phi(\alpha_2) + \phi(-\alpha_2) = (\phi(\alpha_1) - \phi(\alpha_2)) + (\phi(-\alpha_2) - \phi(-\alpha_1)) \\
\geq 0.
\]

Then,

\[
C_\phi(\alpha_1, \eta) - C_\phi(\alpha_2, \eta) = (\phi(\alpha_1) - \phi(-\alpha_1) - \phi(\alpha_2) + \phi(-\alpha_2))\eta + \phi(-\alpha_1) - \phi(-\alpha_2) \\
\geq (\phi(\alpha_1) - \phi(-\alpha_1) - \phi(\alpha_2) + \phi(-\alpha_2))\frac{1}{2} + \phi(-\alpha_1) - \phi(-\alpha_2) \\
= \frac{\phi(\alpha_1) + \phi(-\alpha_1) - \phi(\alpha_2) - \phi(-\alpha_2)}{2} \\
\geq 0,
\]

where the last inequality holds because \(\phi(\alpha) + \phi(-\alpha)\) is nonincreasing when \(\alpha \geq 0\) by part 4. Therefore, \(C_\phi(\alpha, \eta)\) is nonincreasing in \(\alpha \geq 0\).

**Part 3:** Fix a \(\eta \in \left(\frac{1}{2}, 1\right]\). Then,

\[
C_\phi(-1, \eta) - C_\phi(1, \eta) = (2\eta - 1)(\phi(-1) - \phi(1)) \\
> 0,
\]

where the inequality holds due to \(\eta > \frac{1}{2}\) and \(\phi \neq 0\) and \(\phi\) is non-increasing.

**Part 4:** \(\phi\) is an even function, so it is symmetric in \(\alpha = 0\). Since \(\phi\) is quasiconcave even, i.e., \(\phi\) is quasiconcave. Every quasiconcave function is nondecreasing, or nonincreasing, or there is global maxima in its domain (Boyd and Vandenberghe, 2004). If \(\bar{\phi}\) is either nondecreasing or nonincreasing in \(\alpha \in [-1, 1]\), it is a constant function in \(\alpha \in [-1, 1]\) and clearly nonincreasing in \(\alpha \geq 0\). If \(\phi\) has global maxima, i.e., there is a point \(\alpha_* \in [-1, 1]\) such that \(\phi\) is nondecreasing for \(\alpha \leq \alpha_*\) and nonincreasing for \(\alpha \geq \alpha_*\), it is still nonincreasing in \(\alpha \geq 0\). This is clear when \(\alpha_* \leq 0\). When \(\alpha_* > 0\), \(\bar{\phi}\) may only be a constant function in \(\alpha \in [0, \alpha_*]\) otherwise we have a point \(\bar{\alpha} \in [0, \alpha_*]\) such that \(\bar{\phi}(\bar{\alpha}) < \bar{\phi}(\alpha_*)\); hence \(\bar{\phi}(\alpha_*) = \bar{\phi}(\alpha_*)\) (write this value as \(\bar{\phi}_\ast\) by the symmetry and \(\bar{\phi}_0 \overset{\text{def}}{=} \bar{\phi}(\bar{\alpha}) < \bar{\phi}_\ast\), which means there is no convex superlevel sets for \(\bar{\phi}\) within the range \((\bar{\phi}_0, \bar{\phi}_\ast)\).

For example, pick \(t \in (\bar{\phi}_0, \bar{\phi}_\ast)\) and consider \(t\)-superlevel set of \(\bar{\phi}\). If \(t\)-superlevel set is convex, it must contain every point in \([-\alpha_*, \alpha_*]\) since \(t < \bar{\phi}_\ast = \bar{\phi}(-\alpha_*) = \bar{\phi}(\alpha_*)\). However, \(t\)-superlevel set would not contain \(\bar{\alpha} \in [-\alpha_*, \alpha_*]\) since \(t > \bar{\phi}_0 = \bar{\phi}(\bar{\alpha})\). This contradicts with quasiconcavity of \(\phi\). In either case, \(\bar{\phi}\) is nonincreasing in \(\alpha \geq 0\).

**Part 5:** Fix \(\eta \in [0, 1]\). This is an immediate consequence of quasiconcavity of \(C_\phi(\alpha, \eta)\) (part 1). Boyd and Vandenberghe (2004) states that there are three cases for a quasiconcave function. If
DVERSARIALLY ROBUST CLASSIFICATION CALIBRATION

\[ C_\phi(\alpha, \eta) \] is nondecreasing or nonincreasing, the statement is clear. If there is a point \( \alpha_* \in [l, u] \) such that \( C_\phi(\alpha, \eta) \) is nondecreasing for \( \alpha \leq \alpha_* \) and nonincreasing for \( \alpha \geq \alpha_* \), the statement is clear again. In all cases, we have \( \inf_{\alpha \in [l, u]} C_\phi(\alpha, \eta) = \min \{ C_\phi(l, \eta), C_\phi(u, \eta) \} \).

Appendix C. Derivation of Calibration Functions

C.1. Ramp Loss

The ramp loss is \( \phi(\alpha) = \min \{ 1, \max \{ 0, \frac{1-\alpha}{2} \} \} \). We consider the shifted ramp loss: \( \phi_\beta(\alpha) = \phi(\alpha - \beta) \):

\[
\phi_\beta(\alpha) = \begin{cases} 
1 & \text{if } \alpha \leq -1 + \beta, \\
\frac{1 - \alpha + \beta}{2} & \text{if } -1 + \beta < \alpha \leq 1 + \beta, \\
0 & \text{if } 1 + \beta < \alpha. 
\end{cases}
\]

C.1.1. Calibration Function

We analyze \( \phi_\beta\)-CCR \( C_{\phi_\beta}(\alpha, \eta) \) = \( \eta \phi_\beta(\alpha) + (1 - \eta) \phi_\beta(-\alpha) \), and restrict \( \eta > \frac{1}{2} \) by virtue of the symmetry of \( C_{\phi_\beta} \) (part 1 in Lemma 12). \( C_{\phi_\beta}(\alpha, \eta) \) is plotted in Figure 13. By part 5 of Lemma 13, it is easy to see

\[
C^*_{\phi_\beta, \text{lin}}(\eta) = \min\{C_{\phi_\beta}(-1, \eta), C_{\phi_\beta}(1, \eta)\} = C_{\phi_\beta}(1, \eta). 
\]

Subsequently, we divide into cases depending on the relationship among \( C_{\phi_\beta}(-1, \eta), C_{\phi_\beta}(\gamma, \eta), \) and \( C_{\phi_\beta}(-\gamma, \eta) \).
(A) When \(0 \leq \beta < 1 - \gamma\):

\[
C_{\phi,\beta}(1, \eta) = \frac{\beta}{2} \eta + (1 - \eta),
\]

\[
C_{\phi,\beta}(-1, \eta) = \eta + \frac{\beta}{2} (1 - \eta),
\]

\[
C_{\phi,\beta}(\gamma, \eta) = \frac{1 - \gamma + \beta}{2} \eta + \frac{1 + \gamma + \beta}{2} (1 - \eta),
\]

\[
C_{\phi,\beta}(-\gamma, \eta) = \frac{1 + \gamma + \beta}{2} \eta + \frac{1 - \gamma + \beta}{2} (1 - \eta),
\]

from which it follows that \(C_{\phi,\beta}(-\gamma, \eta) - C_{\phi,\beta}(\gamma, \eta) = \frac{\gamma}{2} (2\eta - 1) > 0\), that is, \(C_{\phi,\beta}(\gamma, \eta) > C_{\phi,\beta}(\gamma, \eta)\) for all \(\eta > \frac{1}{2}\). In addition, since

\[
C_{\phi,\beta}(\gamma, \eta) - C_{\phi,\beta}(-1, \eta) = \left(1 + \gamma - \frac{\beta}{2}\right) (\eta - \eta_0) \quad \text{where} \quad \eta_0 \overset{\text{def}}{=} \frac{1 + \gamma}{2 (1 + \gamma - \frac{\beta}{2})},
\]

we have \(C_{\phi,\beta}(\gamma, \eta) > C_{\phi,\beta}(-1, \eta)\) if \(\eta < \eta_0\) and \(C_{\phi,\beta}(\gamma, \eta) \leq C_{\phi,\beta}(-1, \eta)\) if \(\eta \geq \eta_0\).

- If \(\frac{1}{2} < \eta < \eta_0\): By part 5 in Lemma 13, it follows that

\[
\inf_{|\alpha| \leq \gamma} C_{\phi,\beta}(\alpha, \eta) = C_{\phi,\beta}(\gamma, \eta) \quad \text{and} \quad \inf_{|\alpha| \leq 1: |\alpha| \leq \gamma \text{ or } (2\eta - 1) \alpha \leq 0} C_{\phi,\beta}(\alpha, \eta) = C_{\phi,\beta}(-1, \eta).
\]

Thus, by Lemma 7,

\[
\delta(\varepsilon, \eta) = \begin{cases} \infty & \text{if } \eta < \varepsilon, \\ C_{\phi,\beta}(\gamma, \eta) - C_{\phi,\beta, \mathcal{F}_{\text{lin}}}(\eta) &= \left(1 - \gamma - \frac{\beta}{2}\right) (\eta - \frac{1}{2}) + \frac{\beta}{2} & \text{if } \varepsilon \leq \eta < \frac{1 + \varepsilon}{2}, \\ C_{\phi,\beta}(-1, \eta) - C_{\phi,\beta, \mathcal{F}_{\text{lin}}}(\eta) &= (2 - \beta) (\eta - \frac{1}{2}) & \text{if } \frac{1 + \varepsilon}{2} \leq \eta. \end{cases}
\]

Hence we obtain

\[
\inf_{\eta \in \left(\frac{1}{2}, \eta_0\right]} \delta(\varepsilon, \eta) = \begin{cases} \left(1 - \frac{\beta}{2}\right) \varepsilon & \text{if } 0 < \varepsilon \leq \frac{\beta}{2 - \beta}, \\ \frac{\beta}{2} & \text{if } \frac{\beta}{2 - \beta} < \varepsilon \leq \frac{1}{2}, \\ \left(1 - \gamma - \frac{\beta}{2}\right) \left(\varepsilon - \frac{1}{2}\right) + \frac{\beta}{2} & \text{if } \frac{1}{2} < \varepsilon. \end{cases}
\]

Note that \(0 \leq \frac{\beta}{2 - \beta} \leq \frac{1}{2} < \frac{1}{2}\).

- If \(\eta_0 \leq \eta \leq 1\): By part 5 in Lemma 13, it follows that

\[
\inf_{|\alpha| \leq \gamma} C_{\phi,\beta}(\alpha, \eta) = \inf_{|\alpha| \leq 1: |\alpha| \leq \gamma \text{ or } (2\eta - 1) \alpha \leq 0} C_{\phi,\beta}(\alpha, \eta) = C_{\phi,\beta}(\gamma, \eta).
\]

Thus, by Lemma 7,

\[
\delta(\varepsilon, \eta) = \begin{cases} \infty & \text{if } \eta < \varepsilon, \\ C_{\phi,\beta}(\gamma, \eta) - C_{\phi,\beta, \mathcal{F}_{\text{lin}}}(\eta) &= \left(1 - \gamma - \frac{\beta}{2}\right) (\eta - \frac{1}{2}) + \frac{\beta}{2} & \text{if } \varepsilon \leq \eta, \end{cases}
\]

Hence we obtain

\[
\inf_{\eta \in [\eta_0, 1]} \delta(\varepsilon, \eta) = \begin{cases} \frac{\beta}{2} & \text{if } 0 < \varepsilon \leq \frac{1}{2}, \\ \left(1 - \gamma - \frac{\beta}{2}\right) \left(\varepsilon - \frac{1}{2}\right) + \frac{\beta}{2} & \text{if } \frac{1}{2} < \varepsilon. \end{cases}
\]
Combining the above, we obtain the $\phi_{\gamma}$-calibration function from Lemma 7:

$$
\delta(\varepsilon) = \begin{cases} 
\left(1 - \frac{\beta}{2}\right)\varepsilon & \text{if } 0 < \varepsilon \leq \frac{\beta}{2-\beta}, \\
\frac{\beta}{2} & \text{if } \frac{\beta}{2-\beta} < \varepsilon \leq \frac{1}{2}, \\
\left(1 - \gamma - \frac{\beta}{2}\right)(\varepsilon - \frac{1}{2}) + \frac{\beta}{2} & \text{if } \frac{1}{2} < \varepsilon.
\end{cases}
$$

(B) When $1 - \gamma \leq \beta < 1 + \gamma$:

$$
C_{\phi_{\beta}}(1, \eta) = \frac{\beta}{2}\eta + (1 - \eta),
$$

$$
C_{\phi_{\beta}}(-1, \eta) = \eta + \frac{\beta}{2}(1 - \eta),
$$

$$
C_{\phi_{\beta}}(\gamma, \eta) = \frac{1 - \gamma + \beta}{2}\eta + (1 - \eta),
$$

$$
C_{\phi_{\beta}}(-\gamma, \eta) = \eta + \frac{1 - \gamma + \beta}{2}(1 - \eta),
$$

from which it follows that $C_{\phi_{\beta}}(-\gamma, \eta) - C_{\phi_{\beta}}(\gamma, \eta) = \frac{1 + \gamma - \beta}{2}(2\eta - 1) > 0$, that is, $C_{\phi_{\beta}}(-\gamma, \eta) > C_{\phi_{\beta}}(\gamma, \eta)$ for all $\eta > \frac{1}{2}$. In addition, since

$$
C_{\phi_{\beta}}(\gamma, \eta) - C_{\phi_{\beta}}(-1, \eta) = -\frac{3 + \gamma - 2\beta}{2}(\eta - \eta_0), \quad \left(\eta_0 \overset{\text{def}}{=} \frac{2}{3 + \gamma - 2\beta}\right)
$$

we have $C_{\phi_{\beta}}(\gamma, \eta) > C_{\phi_{\beta}}(-1, \eta)$ if $\eta < \eta_0$ and $C_{\phi_{\beta}}(\gamma, \eta) \leq C_{\phi_{\beta}}(-1, \eta)$ if $\eta \geq \eta_0$.

- If $\frac{1}{2} < \eta < \eta_0$: By part 5 in Lemma 13, it follows that

$$
\inf_{|\alpha| \leq \gamma} C_{\phi_{\beta}}(\alpha, \eta) = C_{\phi_{\beta}}(\gamma, \eta) \quad \text{and} \quad \inf_{|\alpha| \leq \gamma (2\eta-1) \alpha \leq 0} C_{\phi_{\beta}}(\alpha, \eta) = C_{\phi_{\beta}}(-1, \eta).
$$

Thus, by Lemma 7,

$$
\delta(\varepsilon, \eta) = \begin{cases} 
\infty & \text{if } \eta < \varepsilon, \\
C_{\phi_{\beta}}(\gamma, \eta) - C_{\phi_{\beta},\mathcal{F}_{\text{lin}}}^{*}(\eta) = \frac{1 - \gamma}{2}\eta & \text{if } \varepsilon \leq \eta < \frac{1 + \varepsilon}{2}, \\
C_{\phi_{\beta}}(-1, \eta) - C_{\phi_{\beta},\mathcal{F}_{\text{lin}}}^{*}(\eta) = (2 - \beta)(\eta - \frac{1}{2}) & \text{if } \frac{1 + \varepsilon}{2} \leq \eta.
\end{cases}
$$

Hence we obtain

$$
\inf_{\eta \in \left(\frac{1}{2}, \eta_0\right]} \delta(\varepsilon, \eta) = \begin{cases} 
\left(1 - \frac{\beta}{2}\right)\varepsilon & \text{if } 0 < \varepsilon \leq \frac{1 - \gamma}{2(2 - \beta)}, \\
\frac{1 - \gamma}{4} & \text{if } \frac{1 - \gamma}{2(2 - \beta)} < \varepsilon \leq \frac{1}{2}, \\
\frac{1 - \gamma}{2} & \text{if } \frac{1}{2} < \varepsilon.
\end{cases}
$$

- If $\eta_0 \leq \eta \leq \frac{1}{2}$: By part 5 in Lemma 13, it follows that

$$
\inf_{|\alpha| \leq \gamma} C_{\phi_{\beta}}(\alpha, \eta) = \inf_{|\alpha| \leq \gamma (2\eta-1) \alpha \leq 0} C_{\phi_{\beta}}(\alpha, \eta) = C_{\phi_{\beta}}(\gamma, \eta).
$$
Thus, by Lemma 7,
\[
\delta(\varepsilon, \eta) = \begin{cases} 
\infty & \text{if } \eta < \varepsilon, \\
C_{\phi, \beta}(\gamma, \eta) - C^*_{\phi, \beta, \mathcal{F}_{\text{lin}}}(\eta) = \frac{1-\gamma}{2} \eta & \text{if } \varepsilon \leq \eta.
\end{cases}
\]
Hence we obtain
\[
\inf_{\eta \in (\eta_0, 1]} \delta(\varepsilon, \eta) = \begin{cases} 
\varepsilon' & \text{if } 0 < \varepsilon \leq \varepsilon', \\
1 - \frac{\gamma}{2} \varepsilon & \text{if } \varepsilon' < \varepsilon,
\end{cases}
\]
where \(\varepsilon' \overset{\text{def}}{=} \frac{1-\gamma}{2} \eta_0 (\geq \frac{1-\gamma}{4})\).

Combining the above, we obtain the \(\phi_\gamma\)-calibration function from Lemma 7:
\[
\delta(\varepsilon) = \begin{cases} 
(1 - \frac{\beta}{2}) \varepsilon & \text{if } 0 < \varepsilon \leq \frac{1-\gamma}{2(2-\beta)}, \\
1 - \frac{\gamma}{4} & \text{if } \frac{1-\gamma}{2(2-\beta)} < \varepsilon \leq \frac{1}{2}, \\
1 - \frac{\gamma}{2} \varepsilon & \text{if } \frac{1}{2} < \varepsilon.
\end{cases}
\]
Note that \(\frac{1-\gamma}{2(2-\beta)} \leq \frac{1-\gamma}{2(1+\gamma)} < \frac{1}{2}\) when \(1 - \gamma \leq \beta < 1 + \gamma\). This means the second case would not degenerate.

(C) When \(1 + \gamma \leq \beta < 2\): It is easy to see
\[
\inf_{|\alpha| \leq \gamma} C_{\phi, \beta}(\alpha, \eta) = 1,
\]
\[
\inf_{|\alpha| \leq 1| |\alpha| \leq \gamma \text{ or } (2n-1)|\alpha| \leq 0} C_{\phi, \beta}(\alpha, \eta) = C_{\phi, \beta}(-1, \eta) = \eta + \frac{\beta}{2} (1 - \eta),
\]
\[
C^*_{\phi, \beta, \mathcal{F}_{\text{lin}}}(\eta) = C_{\phi, \beta}(1, \eta) = \frac{\beta}{2} \eta + (1 - \eta).
\]
Hence, by part 5 in Lemma 13, it follows that
\[
\delta(\varepsilon, \eta) = \begin{cases} 
\infty & \text{if } \eta < \varepsilon, \\
1 - C_{\phi, \beta}(1, \eta) = (1 - \frac{\beta}{2}) \eta & \text{if } \varepsilon \leq \eta < \frac{1+\varepsilon}{2}, \\
C_{\phi, \beta}(-1, \eta) - C_{\phi, \beta}(1, \eta) = (2 - \beta) (\eta - \frac{1}{2}) & \text{if } \frac{1+\varepsilon}{2} \leq \eta.
\end{cases}
\]
Thus, by Lemma 7, \(\delta(\varepsilon) = \inf_{\eta \in (\frac{1}{2}, 1]} \delta(\varepsilon, \eta) = (1 - \frac{\beta}{2}) \varepsilon\).

(D) When \(2 \leq \beta\): In this case, \(C_{\phi, \beta}(\alpha, \eta) = 1\) for all \(\eta \in [0, 1]\) and \(\alpha \in [-1, 1]\). Hence, \(\Delta C_{\phi, \beta, \mathcal{F}_{\text{lin}}}(\alpha, \eta) = 0\) and \(\delta(\varepsilon) = 0\).

To sum up, the \((\phi_\gamma, \mathcal{F}_{\text{lin}})\)-calibration function and its Fenchel-Legendre biconjugate of the ramp loss is as follows:
\begin{itemize}
  \item If \(0 \leq \beta < 1 - \gamma\),
  \[
  \delta(\varepsilon) = \begin{cases} 
  (1 - \frac{\beta}{2}) \varepsilon & \text{if } 0 < \varepsilon \leq \frac{\beta}{2-\beta}, \\
  \frac{\beta}{2} & \text{if } \frac{\beta}{2-\beta} < \varepsilon \leq \frac{1}{2}, \\
  (1 - \gamma - \frac{\beta}{2}) (\varepsilon - \frac{1}{2}) + \frac{\beta}{2} & \text{if } \frac{1}{2} < \varepsilon,
  \end{cases}
  \]
\end{itemize}
and

\[ \delta^{**}(\varepsilon) = \begin{cases} 
\beta \varepsilon & \text{if } 0 < \varepsilon \leq \frac{1-\gamma}{2}, \\
\left(1 - \frac{\beta}{2}\right) \varepsilon & \text{if } \frac{1-\gamma}{2} < \varepsilon \leq \frac{1}{2}, \quad \text{and} \quad \delta^{**}(\varepsilon) = \left(1 - \frac{\gamma}{2}\right) \varepsilon.
\end{cases} \]

- If \(1 - \gamma \leq \beta < 1 + \gamma\),
  \[ \delta(\varepsilon) = \begin{cases} 
\left(1 - \frac{\beta}{2}\right) \varepsilon & \text{if } 0 < \varepsilon \leq \frac{1-\gamma}{2}, \\
\frac{1-\gamma}{4} & \text{if } \frac{1-\gamma}{2} < \varepsilon \leq \frac{1}{2}, \\
\frac{1-\gamma}{2} \varepsilon & \text{if } \frac{1}{2} < \varepsilon,
\end{cases} \]

- If \(1 + \gamma \leq \beta < 2\), \(\delta(\varepsilon) = \delta^{**}(\varepsilon) = \left(1 - \frac{\beta}{2}\right) \varepsilon\).
- If \(2 \leq \beta\), \(\delta(\varepsilon) = \delta^{**}(\varepsilon) = 0\).

We can see that the ramp loss is calibrated wrt \((\phi_\gamma, F_{\text{lin}})\) when \(0 < \beta < 2\).

C.1.2. Quasiconcavity of Even Part

We confirm that \(\phi_\beta(\alpha) + \phi_\beta(-\alpha)\) is quasiconcave when \(\beta \geq 0\). In each case, \(\phi_\beta(\alpha) + \phi_\beta(-\alpha)\) is plotted in Figure 14.

(A) When \(0 \leq \beta < 1\):

\[ \phi_\beta(\alpha) + \phi_\beta(-\alpha) = \begin{cases} 
1 & \alpha \leq -1 - \beta, \\
\frac{3+\alpha+\beta}{2} & -1 - \beta \leq \alpha < -1 + \beta, \\
\frac{-\alpha+\beta}{2} & -1 + \beta \leq \alpha < 1 - \beta, \\
\frac{3-\alpha+\beta}{2} & 1 - \beta \leq \alpha < 1 + \beta, \\
1 & 1 + \beta \leq \alpha.
\end{cases} \]

The \(t\)-superlevel set of \(\phi_\beta(\alpha) + \phi_\beta(-\alpha)\) (denote \(S_t\)) is as follows.

- If \(t < 1\), \(S_t = \mathbb{R}\).
- If \(1 \leq t \leq 1 + \beta\), \(S_t = \{\alpha \mid |\alpha| \leq 3 + \beta - 2t\}\).
- If \(1 + \beta < t\), \(S_t = \emptyset\).
In all cases, $S_t$ is convex.

**(B) When** $1 \leq \beta$:

$\phi_\beta(\alpha) + \phi_\beta(-\alpha) = \begin{cases} 
1 & \alpha \leq -1 - \beta, \\
\frac{3+\alpha+\beta}{2} & -1 - \beta \leq \alpha < 1 - \beta, \\
2 & 1 - \beta \leq \alpha < -1 + \beta, \\
\frac{3-\alpha+\beta}{2} & -1 + \beta \leq \alpha < 1 + \beta, \\
1 & 1 + \beta \leq \alpha. 
\end{cases}$

The $t$-superlevel set of $\phi_\beta(\alpha) + \phi_\beta(-\alpha)$ (denote $S_t$) is as follows.

- If $t < 1$, $S_t = \mathbb{R}$.
- If $1 \leq t \leq 2$, $S_t = \{\alpha \mid |\alpha| \leq 3 + \beta - 2t\}$.
- If $2 < t$, $S_t = \emptyset$.

In all cases, $S_t$ is convex.

### C.2. Sigmoid Loss

The sigmoid loss is $\phi(\alpha) = \frac{1}{1+e^{\alpha}}$. We consider the shifted sigmoid loss: $\phi_\beta(\alpha) = \frac{1}{1+e^{\alpha-\beta}}$ for $\beta \geq 0$. $\phi_\beta$-CCR is

$C_{\phi_\beta}(\alpha, \eta) = \frac{\eta}{1+e^{\alpha-\beta}} + \frac{1-\eta}{1+e^{-\alpha-\beta}}$.

$C_{\phi_\beta}$ is plotted in Figure 15.

#### C.2.1. Calibration Function

We focus on the case $\eta > \frac{1}{2}$ due to the symmetry of $C_{\phi_\beta}$. By part 5 of Lemma 13, it is easy to check

$C_{\phi_\beta, \text{lin}}(\eta) = \min\{C_{\phi_\beta}(-1, \eta), C_{\phi_\beta}(1, \eta)\} = C_{\phi_\beta}(1, \eta) = \frac{\eta}{1+e^{1-\beta}} + \frac{1-\eta}{1+e^{-1-\beta}}$.

Since

$C_{\phi_\beta}(-\gamma, \eta) - C_{\phi_\beta}(\gamma, \eta) = \left(\frac{\eta}{1+e^{-\gamma-\beta}} + \frac{1-\eta}{1+e^{\gamma-\beta}}\right) - \left(\frac{\eta}{1+e^{-\gamma-\beta}} + \frac{1-\eta}{1+e^{\gamma-\beta}}\right)$

$= (2\eta - 1) \left(\frac{1}{1+e^{-\gamma-\beta}} - \frac{1}{1+e^{\gamma-\beta}}\right)$

$> 0$, (since $-\gamma - \beta < \gamma - \beta$)

we have $C_{\phi_\beta}(\gamma, \eta) < C_{\phi_\beta}(-\gamma, \eta)$ for all $\eta > \frac{1}{2}$. On the other hand, since

$C_{\phi_\beta}(-1, \eta) - C_{\phi_\beta}(\gamma, \eta)$

$= \left(\frac{1}{1+e^{1-\beta}} - \frac{1}{1+e^{\gamma-\beta}} - \frac{1}{1+e^{\gamma-\beta}} + \frac{1}{1+e^{-\gamma-\beta}}\right) \eta + \left(\frac{1}{1+e^{1-\beta}} - \frac{1}{1+e^{-\gamma-\beta}}\right)$

$> \frac{1}{2} \left(\frac{1}{1+e^{1-\beta}} + \frac{1}{1+e^{1-\beta}} + \frac{1}{1+e^{\gamma-\beta}} + \frac{1}{1+e^{\gamma-\beta}}\right)$

$> 0$, 

33
we have \( C_\phi(\gamma, \eta) < C_\phi(-1, \eta) \) for all \( \eta \in [0, 1] \). By part 5 in Lemma 13, it follows that

\[
\inf_{|\alpha| \leq 1; |\alpha| \leq \gamma \text{ or } (2\gamma - 1)\alpha \leq 0} C_\phi(\alpha, \eta) = \inf_{|\alpha| \leq \gamma} C_\phi(\alpha, \eta) = C_\phi(\gamma, \eta).
\]

Thus, by Lemma 7, \( \bar{\delta}(\varepsilon, \eta) = A_0(\eta - \eta_0) \) if \( \varepsilon \leq \eta \), where

\[
A_0 \overset{\text{def}}{=} \phi_\beta(\gamma) - \phi_\beta(-\gamma) - \phi_\beta(1) + \phi_\beta(-1), \quad \eta_0 \overset{\text{def}}{=} \frac{\phi_\beta(-1) + \phi_\beta(-\gamma)}{A_0},
\]

and \( \bar{\delta}(\varepsilon, \eta) = \infty \) if \( \varepsilon > \eta \). Note that \( A_0 > 0 \), \( \eta_0 \leq \frac{1}{2} \), and \( \eta_0 = \frac{1}{2} \Leftrightarrow \beta = 0 \). Hence we obtain

\[
\delta(\varepsilon) = \inf_{\eta \in [\frac{1}{2}, 1]} \bar{\delta}(\varepsilon, \eta) = \begin{cases} A_1 & \text{if } 0 < \varepsilon \leq \frac{1}{2}, \\ A_0(\varepsilon - \eta_0) & \text{if } \frac{1}{2} < \varepsilon, \end{cases}
\]

where \( A_1 \overset{\text{def}}{=} A_0 \left( \frac{1}{2} - \eta_0 \right) = (\phi_\beta(\gamma) + \phi_\beta(-\gamma) - \phi_\beta(1) - \phi_\beta(-1))/2. \)

Thus, the sigmoid loss is calibrated wrt \( (\phi_\gamma, F_{\text{lin}}) \) when \( A_1 > 0 \), which is equivalent to \( \beta > 0 \).

Let \( \tilde{\delta} : [0, 1] \to \mathbb{R}_{\geq 0} \) be a function such that \( \tilde{\delta}(0) = \delta(0) \) and \( \tilde{\delta}(\varepsilon) = \delta(\varepsilon) \) for all \( \varepsilon > 0 \). Then, the Fenchel-Legendre biconjugate of \( \delta \) is

\[
\tilde{\delta}^{**}(\varepsilon) = \begin{cases} 2A_1\varepsilon & \text{if } 0 \leq \varepsilon \leq \frac{1}{2}, \\ A_0(\varepsilon - \eta_0) & \text{if } \frac{1}{2} < \varepsilon. \end{cases}
\]

**C.2.2. Quasiconcavity of Even Part**

We confirm that \( \phi_\beta(\alpha) + \phi_\beta(-\alpha) \) is quasiconcave when \( \beta \geq 0 \). \( \phi_\beta(\alpha) + \phi_\beta(-\alpha) \) is plotted in Figure 16, and

\[
\phi_\beta(\alpha) + \phi_\beta(-\alpha) = \frac{1}{1 + e^{\alpha - \beta}} + \frac{1}{1 + e^{-\alpha - \beta}} \quad (\overset{\text{def}}{=} \tilde{\phi}(\alpha)).
\]

Here, we use quasimonotonicity of \( -\tilde{\phi}' \) to show its quasiconcavity (see Appendix A). The derivative of \( \tilde{\phi} \) is

\[
\tilde{\phi}'(\alpha) = 4 \cosh^{-2} \left( \frac{\alpha + \beta}{2} \right) - 4 \cosh^{-2} \left( \frac{\alpha - \beta}{2} \right),
\]

which is plotted in Figure 17. Take \( \alpha_1, \alpha_2 \in \mathbb{R} \) such that \( \alpha_1 < \alpha_2 \) without loss of generality. Assume that \( (-\tilde{\phi}'(\alpha_1))(\alpha_2 - \alpha_1) > 0 \), which implies \( \tilde{\phi}'(\alpha_1) < 0 \Leftrightarrow \alpha_1 > 0 \). Hence, \( \alpha_2 > 0 \) and \( \tilde{\phi}'(\alpha_2) < 0 \). We obtain \( (-\tilde{\phi}'(\alpha_2))(\alpha_2 - \alpha_1) > 0 \) straightforwardly, which implies quasimonotonicity of \( -\tilde{\phi}' \). Therefore, \( \tilde{\phi} \) is quasiconcave.
C.3. Modified Squared Loss

We design a bounded and nonincreasing surrogate loss by modifying the squared loss, which we call modified squared loss here:

\[
\phi(\alpha) = \begin{cases} 
1 & \text{if } \alpha \leq 0, \\
(1 - \alpha)^2 & \text{if } 0 < \alpha \leq 1, \\
0 & \text{if } 1 < \alpha,
\end{cases}
\]

and consider the shifted version \(\phi_\beta(\alpha) \equiv \phi(\alpha - \beta)\):

\[
\phi_\beta(\alpha) = \begin{cases} 
1 & \text{if } \alpha \leq \beta, \\
(1 - \alpha + \beta)^2 & \text{if } \beta < \alpha \leq 1 + \beta, \\
0 & \text{if } 1 + \beta < \alpha.
\end{cases}
\]

C.3.1. Calibration Function

Now we consider \(\phi_\beta\)-CCR \(C_{\phi_\beta}(\alpha, \eta) = \eta \phi(\alpha) + (1 - \eta)\phi(-\alpha)\) and focus on the case \(\eta > \frac{1}{2}\) due to the symmetry of \(C_{\phi_\beta}\). \(C_{\phi_\beta}\) is plotted in Figure 18. By part 5 of Lemma 13, it is easy to see \(C_{\phi_\beta, \mathcal{F}_{\text{lin}}}(\eta) = \min\{C_{\phi_\beta}(-1, \eta), C_{\phi_\beta}(1, \eta)\} = C_{\phi_\beta}(1, \eta)\). We divide into three cases depending on the relationship among \(C_{\phi_\beta}(-1, \eta), C_{\phi_\beta}(-\gamma, \eta)\), and \(C_{\phi_\beta}(\gamma, \eta)\),

(A) When \(0 \leq \beta < \gamma\): Since

\[
C_{\phi_\beta}(-\gamma, \eta) - C_{\phi_\beta}(\gamma, \eta) = \left\{ \eta \cdot 1 + (1 - \eta)(1 - \gamma + \beta)^2 \right\} - \left\{ \eta(1 - \gamma + \beta)^2 + (1 - \eta) \cdot 1 \right\} = (2\eta - 1)(\gamma - \beta) \{2 - (\gamma - \beta)\} \geq 0,
\]

we have \(C_{\phi_\beta}(\gamma, \eta) < C_{\phi_\beta}(-\gamma, \eta)\) for all \(\eta > \frac{1}{2}\). On the other hand, since

\[
C_{\phi_\beta}(\gamma, \eta) - C_{\phi_\beta}(-1, \eta) = -\{(2 - \gamma + \beta)(\gamma - \beta) + (1 - \beta)^2\} (\eta - \eta_0)
\]

where \(\eta_0 \equiv \frac{1 - \beta^2}{(2 - \gamma + \beta)(\gamma - \beta) + (1 - \beta^2)}\),

and \(\frac{1}{2} < \eta_0 \leq 1\), we have \(C_{\phi_\beta}(\gamma, \eta) \geq C_{\phi_\beta}(-1, \eta)\) if \(\frac{1}{2} < \eta \leq \eta_0\) and \(C_{\phi_\beta}(\gamma, \eta) < C_{\phi_\beta}(-1, \eta)\) if \(\eta > \eta_0\).

- If \(\frac{1}{2} < \eta \leq \eta_0\): By part 5 in Lemma 13,

\[
\inf_{|\alpha| \leq 1: |\alpha| \leq \gamma \text{ or } (2\eta - 1)|\alpha| \leq 0} C_{\phi_\beta}(\alpha, \eta) = C_{\phi_\beta}(-1, \eta) \quad \text{and} \quad \inf_{|\alpha| \leq \gamma} C_{\phi_\beta}(\alpha, \eta) = C_{\phi_\beta}(\gamma, \eta).
\]

Thus, by Lemma 7,

\[
\bar{\delta}(\varepsilon, \eta) = \begin{cases} 
\infty & \text{if } \eta < \varepsilon, \\
C_{\phi_\beta}(\gamma, \eta) - C_{\phi_\beta, \mathcal{F}_{\text{lin}}}^*(\eta) = (1 - \gamma)(1 - \gamma + 2\beta)\eta & \text{if } \varepsilon \leq \eta < \frac{1 + \varepsilon}{2}, \\
C_{\phi_\beta}(-1, \eta) - C_{\phi_\beta, \mathcal{F}_{\text{lin}}}^*(\eta) = (1 - \beta^2)(2\eta - 1) & \text{if } \frac{1 + \varepsilon}{2} \leq \eta.
\end{cases}
\]
Hence we obtain
\[ \inf_{\eta \in (\frac{1}{2}, \eta_0]} \tilde{\delta}(\varepsilon, \eta) = \begin{cases} (1 - \beta^2)\varepsilon & \text{if } 0 < \varepsilon \leq \varepsilon_0, \\ \frac{(1-\gamma+2\beta)(1-\gamma)}{2} & \text{if } \varepsilon_0 < \varepsilon \leq \frac{1}{2}, \\ (1 - \gamma)(1 - \gamma + 2\beta)\varepsilon & \text{if } \frac{1}{2} < \varepsilon, \end{cases} \]
where \( \varepsilon_0 \overset{\text{def}}{=} \frac{(1-\gamma)(1-\gamma+2\beta)}{2(1-\beta^2)}. \)

\[ \cdot \text{ If } \eta_0 < \eta \leq 1: \text{ By part 5 in Lemma 13, it follows that} \]
\[ \inf_{|\alpha| \leq 1: \alpha \leq \gamma \text{ or } (2\eta-1)\alpha \leq 0} C_{\phi_\beta}(\alpha, \eta) = \inf_{|\alpha| \leq \gamma} C_{\phi_\beta}(\alpha, \eta) = C_{\phi_\beta}(\gamma, \eta). \]

Thus, by Lemma 7, \( \tilde{\delta}(\varepsilon, \eta) = C_{\phi_\beta}(\gamma, \eta) - C_{\phi_\beta,F_{\text{lin}}}(\eta) = (1 - \gamma)(1 - \gamma + 2\beta)\eta \) if \( \varepsilon \leq \eta \) and \( \tilde{\delta}(\varepsilon, \eta) = \infty \) if \( \eta < \varepsilon. \) Hence we obtain
\[ \inf_{\eta \in (\eta_0, 1]} \tilde{\delta}(\varepsilon, \eta) = \begin{cases} (1 - \gamma)(1 - \gamma + 2\beta) & \text{if } 0 < \varepsilon \leq \frac{1}{2}, \\ (1 - \gamma)(1 - \gamma + 2\beta)\varepsilon & \text{if } \frac{1}{2} < \varepsilon. \end{cases} \]

Combining the above, we obtain the \( \phi_\gamma \)-calibration function from Lemma 7:
\[ \delta(\varepsilon) = \begin{cases} (1 - \beta^2)\varepsilon & \text{if } 0 < \varepsilon \leq \varepsilon_0, \\ \frac{(1-\gamma+2\beta)(1-\gamma)}{2} & \text{if } \varepsilon_0 < \varepsilon \leq \frac{1}{2}, \\ (1 - \gamma)(1 - \gamma + 2\beta)\varepsilon & \text{if } \frac{1}{2} < \varepsilon. \end{cases} \]

Note that \( \varepsilon_0 \leq \frac{1}{2}, \) which means the second case is not vacuous.

\( \text{(B) When } \gamma \leq \beta < 1: \) It is easy to see
\[ \inf_{|\alpha| \leq \gamma} C_{\phi_\beta}(\alpha, \eta) = 1, \]
\[ \inf_{|\alpha| \leq 1: |\alpha| \leq \gamma \text{ or } (2\eta-1)\alpha \leq 0} C_{\phi_\beta}(\alpha, \eta) = C_{\phi_\beta}(-1, \eta). \]

Hence, by part 5 in Lemma 13, it follows that
\[ \tilde{\delta}(\varepsilon, \eta) = \begin{cases} \infty & \text{if } \eta < \varepsilon, \\ 1 - C_{\phi_\beta}(1, \eta) = (1 - \beta^2)\eta & \text{if } \varepsilon \leq \eta < \frac{1+\varepsilon}{2}, \\ C_{\phi_\beta}(-1, \eta) - C_{\phi_\beta}(1, \eta) = (1 - \beta^2)(2\eta - 1) & \text{if } \frac{1+\varepsilon}{2} \leq \eta. \end{cases} \]

Thus, by Lemma 7, \( \delta(\varepsilon) = \inf_{\eta \in (\frac{1}{2}, 1]} \tilde{\delta}(\varepsilon, \eta) = (1 - \beta^2)\varepsilon. \)

\( \text{(C) When } 1 \leq \beta: \) In this case, \( C_{\phi_\beta}(\alpha, \eta) = 1 \) for all \( \alpha \in [-1, 1]. \) Hence, \( \Delta C_{\phi_\beta,F_{\text{lin}}}(\alpha, \eta) = 0 \) and \( \delta(\varepsilon) = 0. \)

To sum up, the \((\phi_\gamma,F_{\text{lin}})\)-calibration function and its Fenchel-Legendre biconjugate of the modified squared loss are as follows:
\[ \delta(\varepsilon) = \begin{cases} 
(1 - \beta^2)\varepsilon & \text{if } 0 < \varepsilon \leq \varepsilon_0, \\
(1 - \gamma + 2\beta)(1 - \gamma) & \text{if } \varepsilon_0 < \varepsilon \leq \frac{1}{2}, \quad \text{and} \\
(1 - \gamma)(1 - \gamma + 2\beta)\varepsilon & \text{if } \frac{1}{2} < \varepsilon,
\end{cases} \]

where \( \varepsilon_0 \) is defined as

\[ \varepsilon_0 = \frac{(1 - \gamma)(1 - \gamma + 2\beta)}{2(1 - \beta^2)}. \]

- If \( \gamma \leq \beta < 1 \), \( \delta(\varepsilon) = \delta^{**}(\varepsilon) = (1 - \beta^2)\varepsilon. \)
- If \( 1 \leq \beta \), \( \delta(\varepsilon) = \delta^{**}(\varepsilon) = 0. \)

We deduce that the modified squared loss is calibrated wrt \( (\phi, F_{\text{lin}}) \) if \( 0 \leq \beta < 1 \).

C.3.2. Quasiconcavity of Even Part

We confirm that \( \phi_\beta(\alpha) + \phi_\beta(-\alpha) \) is quasiconcave when \( \beta \geq 0. \)

\[
\phi_\beta(\alpha) + \phi_\beta(-\alpha) = \begin{cases} 
1 & \alpha < -1 - \beta, \\
(1 + \alpha + \beta)^2 + 1 & -1 - \beta \leq \alpha < -\beta, \\
2 & -\beta \leq \alpha < \beta, \\
(1 - \alpha + \beta)^2 + 1 & \beta \leq \alpha < 1 + \beta, \\
1 & 1 + \beta \leq \alpha.
\end{cases}
\]

Its \( t \)-superlevel set \( S_t \) is as follows.

- If \( t < 1 \), \( S_t = \mathbb{R} \).
- If \( 1 \leq t \leq 2 \), \( S_t = \{ \alpha \mid |\alpha| \leq 1 + \beta - \sqrt{t - 1} \} \).
- If \( 2 < t \), \( S_t = \emptyset \).

In all cases, \( S_t \) is convex. Thus, \( \phi_\beta(\alpha) + \phi_\beta(-\alpha) \) is quasiconcave.
C.3.3. When $\beta < 0$

In this case, the modified squared loss is no longer quasiconcave even (see Figure 19 (b)). However, $\phi_\beta$ is still $(\phi_\gamma, \mathcal{F}_{\text{lin}})$-calibrated under some $\gamma$ and $\beta < 0$. Here, we show an example.

Assume that $\gamma < \frac{1}{4}$ and $-1 + \frac{1}{\sqrt{2}} < \beta < 0$. We focus on $\eta > \frac{1}{2}$ due to the symmetry of $C_{\phi_\beta}$. In these $\beta$ and $\gamma$, we still have $\eta_0 > \frac{1}{2}$, because

$$\eta_0 = \frac{1 - \beta^2}{(2 - \gamma + \beta)(\gamma - \beta) + (1 - \beta^2)} > \frac{1}{2}$$

$$\iff 2(1 - \beta^2) > (2 - \gamma + \beta)(\gamma - \beta) + (1 - \beta^2)$$

$$\iff \gamma^2 - 2(1 + \beta)\gamma + (2\beta + 1) > 0$$

$$\iff \gamma < 1 + 2\beta, \quad \frac{1}{2} < \gamma < \gamma$$

$$\iff \gamma < 1 + 2\beta,$$

and $1 + 2\beta > \sqrt{2} \left( 1 - \frac{1}{\sqrt{2}} \right) > \frac{1}{4} > \gamma$ always holds when $\gamma < \frac{1}{4}$. Then, we can confirm in the same way as the case (A) that

- $C_{\phi_\beta}(-\gamma, \eta) > C_{\phi_\beta}(\gamma, \eta)$ for all $\eta > \frac{1}{2}$.
- $C_{\phi_\beta}(\gamma, \eta) \geq C_{\phi_\beta}(-1, \eta)$ if $\frac{1}{2} < \eta \leq \eta_0$, and $C_{\phi_\beta}(\gamma, \eta) < C_{\phi_\beta}(-1, \eta)$ if $\eta_0 < \eta$.  

Figure 19: The class-conditional risk of the modified squared loss when $\gamma < \frac{1}{4}$ and $-1 + \frac{1}{\sqrt{2}} < \beta < 0$. 

38
In addition, we see that

\[
C_{\phi_\beta}(-1, \eta) - C_{\phi_\beta}(0, \eta) = \eta - (1 + \beta)^2,
\]

\[
C_{\phi_\beta}(0, \eta) - C_{\phi_\beta}(1, \eta) = (1 + \beta)^2 - (1 - \eta) > \frac{1}{2} - (1 - \eta) > 0,
\]

\[
C_{\phi_\beta}(\gamma, \eta) - C_{\phi_\beta}(1, \eta) = \left(1 - 4\gamma(1 + \beta)\right)\eta + (1 + \gamma + \beta)^2 - 1
\]

\[
> 1 - 4\gamma > 0
\]

\[
= \gamma^2 + \beta^2 + 2\beta + \frac{1}{2}
\]

\[
= (\beta + 1)^2 - \frac{1}{2} > 0.
\]

\[
C_{\phi_\beta}(0, \eta) - C_{\phi_\beta}(\gamma, \eta) = 4\gamma(\beta + 1) (\eta - \eta_1),
\]

where \(\eta_1 \overset{\text{def}}{=} \frac{2 + 2\beta + \gamma}{4(1 + \beta)} \in \left(\frac{1}{2}, 1\right]\). Then, we have

- \(C_{\phi_\beta}(-1, \eta) > C_{\phi_\beta}(0, \eta)\) for \(\eta > (1 + \beta)^2\), and \(C_{\phi_\beta}(-1, \eta) \leq C_{\phi_\beta}(0, \eta)\) for \(\frac{1}{2} < \eta \leq (1 + \beta)^2\).
- \(C_{\phi_\beta}(0, \eta) > C_{\phi_\beta}(1, \eta)\) for all \(\eta > \frac{1}{2}\).
- \(C_{\phi_\beta}(\gamma, \eta) > C_{\phi_\beta}(1, \eta)\) for all \(\eta > \frac{1}{2}\).
- \(C_{\phi_\beta}(\gamma, \eta) \geq C_{\phi_\beta}(0, \eta)\) if \(\frac{1}{2} < \eta \leq \eta_1\), and \(C_{\phi_\beta}(\gamma, \eta) < C_{\phi_\beta}(0, \eta)\) if \(\eta_1 < \eta\).

Figure 19 and the above comparisons give us

\[
C_{\phi_\beta, \text{lin}}^*(\eta) = \inf_{\alpha \in [-1, 1]} C_{\phi_\beta}\left(\alpha, \eta\right) = C_{\phi_\beta}(1, \eta),
\]

\[
\inf_{|\alpha| \leq \gamma} C_{\phi_\beta}(\alpha, \eta) = \min\{C_{\phi_\beta}(0, \eta), C_{\phi_\beta}(\gamma, \eta)\},
\]

\[
\inf_{\alpha \in [-\gamma, \gamma]} C_{\phi_\beta}(\alpha, \eta) = \min\{C_{\phi_\beta}(-1, \eta), C_{\phi_\beta}(0, \eta), C_{\phi_\beta}(\gamma, \eta)\}.
\]

By Lemma 7, when \(\varepsilon \leq \eta < \frac{1 + \varepsilon}{2}\),

\[
\bar{\delta}(\varepsilon, \eta) = \inf_{|\alpha| \leq \gamma} C_{\phi_\beta}(\alpha, \eta) - C_{\phi_\beta, \text{lin}}^*(\eta)
\]

\[
= \min\{C_{\phi_\beta}(0, \eta) - C_{\phi_\beta}(1, \eta), C_{\phi_\beta}(\gamma, \eta) - C_{\phi_\beta}(1, \eta)\}
\]

\[
= \min\{\eta + (\beta^2 + 2\beta)(1 - 4\gamma(1 + \beta))\eta + (1 + \gamma + \beta)^2 - 1\},
\]
ADVERSARILY ROBUST CLASSIFICATION CALIBRATION

and

\[
\inf_{\eta \in [\varepsilon, 1+\varepsilon^2] \cap \left(\frac{1}{2}, 1\right]} \delta(\varepsilon, \eta) = \min \left\{ \begin{array}{ll}
(\cdot) & 
\left( \varepsilon - \frac{1}{2} \right) + \frac{1}{2} + (\beta^2 + 2\beta), \\
(ii) & (1 - 4\gamma(1 + \beta)) \left( \varepsilon - \frac{1}{2} \right) + \frac{1}{2} + (1 + \gamma + \beta)^2 - 1
\end{array} \right. 
\]

When \(\frac{1+\varepsilon}{2} \leq \eta\),

\[
\delta(\varepsilon, \eta) = \inf_{\alpha \in [-1, \gamma]} C_{\phi_{1}}(\alpha, \eta) - C_{\phi_{1}, \mathcal{F}_{\text{lin}}}(\eta) = \min \{ C_{\phi_{1}}(-1, \eta) - C_{\phi_{1}}(1, \eta), C_{\phi_{1}}(0, \eta) - C_{\phi_{1}}(1, \eta), C_{\phi_{1}}(\gamma, \eta) - C_{\phi_{1}}(1, \eta) \} = \min \{ 2\eta - 1, \eta + (\beta^2 + 2\beta), (1 - 4\gamma(1 + \beta))\eta + (1 + \gamma + \beta)^2 - 1 \},
\]

and

\[
\inf_{\eta \in [\frac{1+\varepsilon}{2}, 1] \cap \left(\frac{1}{2}, 1\right]} \delta(\varepsilon, \eta)
= \min \left\{ \begin{array}{ll}
(iii) & \varepsilon \\
(iv) & \frac{1+\varepsilon}{2} + (\beta^2 + 2\beta), \\
(v) & (1 - 4\gamma(1 + \beta))\left( \frac{1+\varepsilon}{2} \right) + (1 + \gamma + \beta)^2 - 1
\end{array} \right. 
\]

Note that for any \(\gamma \in (0, \frac{1}{4})\), \(-1 + \frac{1}{\sqrt{2}} < \beta < 0\), and \(\varepsilon > 0\), we have (iv) \(\geq\) (i) and (v) \(\geq\) (ii), which means that \((\phi_{\gamma}, \mathcal{F}_{\text{lin}})\)-calibration function of \(\phi_{1}\) is

\[
\delta(\varepsilon) = \min \left\{ \begin{array}{ll}
\inf_{\eta \in [\varepsilon, \frac{1+\varepsilon}{2}] \cap \left(\frac{1}{2}, 1\right]} \delta(\varepsilon, \eta), \\
\inf_{\eta \in \left[\frac{1+\varepsilon}{2}, 1\right] \cap \left(\frac{1}{2}, 1\right]} \delta(\varepsilon, \eta)
\end{array} \right. 
\]

\[
= \min\{ (i), (ii), (iii) \}
= \begin{cases} 
\varepsilon & \text{if } 0 < \varepsilon \leq \varepsilon_0, \\
\varepsilon_0 & \text{if } \varepsilon_0 < \varepsilon \leq \frac{1}{2}, \\
\varepsilon + \beta^2 + 2\beta & \text{if } \frac{1}{2} < \varepsilon \leq \eta_1, \\
(1 - 4\gamma(1 + \beta))\varepsilon + (1 + \gamma + \beta)^2 - 1 & \text{if } \eta_1 < \varepsilon \leq 1,
\end{cases}
\]

where \(\varepsilon_0 \overset{\text{def}}{=} \beta^2 + 2\beta + \frac{1}{2}\). From this result, we see that the modified squared loss is still \((\phi_{\gamma}, \mathcal{F}_{\text{lin}})\) calibrated when \(0 < \gamma < \frac{1}{4}\) and \(-1 + \frac{1}{\sqrt{2}} < \beta < 0\), and it is no longer calibrated once \(\beta\) becomes \(-1 + \frac{1}{\sqrt{2}}\) (since \(\varepsilon_0 = 0\) at \(\beta = -1 + \frac{1}{\sqrt{2}}\)).
C.4. Hinge Loss

The $\phi_\beta$-CCR is

$$C_{\phi_\beta}(\alpha, \eta) = \begin{cases} -\eta\alpha + \eta(1 + \beta) & \text{if } \alpha < -(1 + \beta), \\ (1 - 2\eta)\alpha + (1 + \beta) & \text{if } -(1 + \beta) \leq \alpha < 1 + \beta, \\ (1 - \eta)\alpha + (1 - \eta)(1 + \beta) & \text{if } 1 + \beta < \alpha. \end{cases}$$

We restrict the range of $\eta$ to $\eta > \frac{1}{2}$ by virtue of part 1 of Lemma 12. Then, $C_{\phi_\beta, F_{\text{lin}}}(\eta) = C_{\phi_\beta}(1, \eta) = -2\eta + (2 + \beta)$. $C_{\phi_\beta}(\alpha, \eta)$ is plotted in Figure 20 in case of $\eta > \frac{1}{2}$. Then, it follows that

$$\inf_{|\alpha| \leq 1: |\alpha| \leq \gamma \text{ or } (2\eta - 1)\alpha \leq 0} C_{\phi_\beta}(\alpha, \eta) = \inf_{|\alpha| \leq \gamma} C_{\phi_\beta}(\alpha, \eta) = C_{\phi_\beta}(\gamma, \eta) = (1 - 2\eta)\gamma + (1 + \beta).$$

Hence, by Lemma 7,

$$\tilde{\delta}(\varepsilon, \eta) = \begin{cases} \infty & \text{if } \varepsilon < \eta, \\ C_{\phi_\beta}(\gamma, \eta) - C^*_{\phi_\beta, F_{\text{lin}}}(\eta) = (1 - \gamma)(2\eta - 1) & \text{if } \eta \leq \varepsilon, \end{cases}$$

and

$$\delta(\varepsilon) = \begin{cases} 0 & \text{if } 0 < \varepsilon \leq \frac{1}{2}, \\ (1 - \gamma)(2\varepsilon - 1) & \text{if } \frac{1}{2} < \varepsilon. \end{cases}$$

C.5. Squared Loss

The $\phi_\beta$-CCR is

$$C_{\phi_\beta}(\alpha, \eta) = \eta(1 - \alpha + \beta)^2 + (1 - \eta)(1 + \alpha + \beta)^2$$

$$= \{\alpha - (1 + \beta)(2\eta - 1)\}^2 + 4(1 + \beta)^2\eta(1 - \eta).$$

We restrict the range of $\eta$ to $\eta > \frac{1}{2}$ by virtue of part 1 of Lemma 12. $C_{\phi_\beta}(\alpha, \eta)$ is plotted in Figure 21 in case of $\eta > \frac{1}{2}$. By comparing $\alpha_\ast \overset{\text{def}}{=} (1 + \beta)(2\eta - 1)$ and 1, we have

$$C^*_{\phi_\beta, F_{\text{lin}}}(\eta) = \begin{cases} C_{\phi_\beta}(1, \eta) & \text{if } \alpha_\ast < 1, \\ C_{\phi_\beta}(\alpha_\ast, \eta) & \text{if } \alpha_\ast \geq 1. \end{cases}$$
functions of We can utilize (12) to numerically approximate the Bayes risk. Let
\[ \eta \]
where
\[ H \]
Hence, by Lemma 7,
\[ D.1. \text{Detail of Numerical Approximation of Bayes Risks} \]
\[ \text{Appendix D. Simulation Results} \]
\[ \text{For the concrete forms of} \ C, \ \text{we obtain in Appendix C except the logistic loss as follows.} \]
\[ \text{Appendix D. Simulation Results} \]
\[ \text{D.1. Detail of Numerical Approximation of Bayes Risks} \]
\[ \text{Consider to compute the Bayes} \ (\phi,F_{\text{lin}})\text{-risk for a loss} \ \phi. \]
\[ \inf_{f \in F_{\text{lin}}} R_\phi(f) = \mathbb{E}_X \left[ \inf_{\alpha \in A_{\text{lin}}} C_\phi(\alpha,\mathbb{P}(Y = +1|X)) \right] \]
\[ = \mathbb{E}_X \left[ C^*_{\phi,F_{\text{lin}}} (\mathbb{P}(Y = +1|X)) \right]. \quad (12) \]
\[ \text{We can utilize} \ (12) \ \text{to numerically approximate the Bayes risk. Let} \ q_+ \text{ and} \ q_- \ \text{be probability density} \]
\[ \text{functions of} \ \mathcal{N}([2\ 2]^T, I_2) \ \text{and} \ \mathcal{N}([-2 \ 2], I_2), \ \text{respectively. Then,} \]
\[ \mathbb{P}(Y = +1|X) = \frac{\mathbb{P}(Y = +1)\mathbb{P}(X|Y = +1)}{\mathbb{P}(Y = +1)\mathbb{P}(X|Y = +1) + \mathbb{P}(Y = -1)\mathbb{P}(X|Y = -1)} = \frac{\frac{1}{2}q_+(X)}{\frac{1}{2}q_+(X) + \frac{1}{2}q_-(X)}. \]
\[ \text{For the concrete forms of} \ C^*_{\phi,F_{\text{lin}}}, \ \text{we obtain in Appendix C except the logistic loss as follows.} \]
Adversarially Robust Classification Calibration

- Robust 0-1 loss: \( C^*_{\phi, F_{\text{lin}}} (\eta) = \min\{\eta, 1 - \eta\} \)
- Ramp loss: \( C^*_{\phi, F_{\text{lin}}} (\eta) = \min\left\{\frac{\beta}{2} \eta + (1 - \eta), \eta + \frac{\beta}{2} \right\} \)
- Sigmoid loss: \( C^*_{\phi, F_{\text{lin}}} (\eta) = \min\left\{ \frac{\eta}{1 + e^{1 - \eta}}, \frac{1 - \eta}{1 + e^{1 - \eta}} \right\} \)
- Hinge loss: \( C^*_{\phi, F_{\text{lin}}} (\eta) = 2 \min\{\eta, 1 - \eta\} + \beta \)

For the logistic loss, it is not difficult to see

\[
C^*_{\phi, F_{\text{lin}}} (\eta) = \begin{cases} \eta \log(1 + e^{-\alpha^* + \beta}) + (1 - \eta) \log(1 + e^{\alpha^* + \beta}) & \text{if } \eta > \frac{1}{2}, \\ \eta \log(1 + e^{\alpha^* + \beta}) + (1 - \eta) \log(1 + e^{-\alpha^* + \beta}) & \text{if } \eta \leq \frac{1}{2}, \end{cases}
\]

where \( \alpha^* = \text{clip}_{[-1, 1]} \left( \log \left( \frac{\eta}{1 - \eta} \right) \right) \). By plugging these expressions into (12) and performing numerical integration, we can approximate the Bayes risks. In the simulation results, we performed numerical integration for the range \([-10, 10] \times [-10, 10]\) split by 200 \(\times\) 200 segments. The partitioning quadrature method was used. The approximated Bayes risks are as follows.

- Robust 0-1 loss: 0.0023474
- Ramp loss: 0.10211
- Sigmoid loss: 0.31110
- Hinge loss: 0.20469
- Logistic loss: 0.37356

D.2. Full Simulation Results of Benchmark Dataset

We show the full simulation results of MNIST in Tables 2 and 3. Simulation details are as follows.

- Dataset: MNIST extracted with two digits (7,000 instances for each digit).
- Preprocessing: Reduced to 2-dimension with the principal component analysis.
- Train-test split: 14,000 instances are randomly split into training and test data with the ratio 4 to 1.
- Model: Linear models \( f(x) = \theta^T x + \theta_0 \) (\( \theta \) and \( \theta_0 \) are learnable parameters)
- Surrogate loss: The ramp, sigmoid, hinge, and logistic losses with shift \( \beta = +0.5 \).
- Target loss: the \( \gamma \)-adversarially robust 0-1 loss with \( \gamma = 0.1 \).
- Optimization: Batch gradient ascent with 1,000 iterations.
Table 2: The simulation results of the $\gamma$-adversarially robust 0-1 loss with $\gamma = 0.1$ and $\beta = 0.5$. 50 trials are conducted for each pair of a method and dataset. Standard errors (multiplied by $10^4$) are shown in parentheses. Bold-faces indicate outperforming methods, chosen by one-sided t-test with the significant level 5%.

<table>
<thead>
<tr>
<th></th>
<th>Ramp</th>
<th>Sigmoid</th>
<th>Hinge</th>
<th>Logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 vs 1</td>
<td>0.034</td>
<td>0.017</td>
<td>0.087</td>
<td>0.321</td>
</tr>
<tr>
<td>0 vs 2</td>
<td>0.111</td>
<td>0.133</td>
<td>0.109</td>
<td>0.281</td>
</tr>
<tr>
<td>0 vs 3</td>
<td>0.107</td>
<td>0.126</td>
<td>0.120</td>
<td>0.307</td>
</tr>
<tr>
<td>0 vs 4</td>
<td>0.069</td>
<td>0.093</td>
<td>0.072</td>
<td>0.269</td>
</tr>
<tr>
<td>0 vs 5</td>
<td>0.233</td>
<td>0.340</td>
<td>0.233</td>
<td>0.269</td>
</tr>
<tr>
<td>0 vs 6</td>
<td>0.129</td>
<td>0.167</td>
<td>0.127</td>
<td>0.287</td>
</tr>
<tr>
<td>0 vs 7</td>
<td>0.067</td>
<td>0.073</td>
<td>0.090</td>
<td>0.302</td>
</tr>
<tr>
<td>0 vs 8</td>
<td>0.096</td>
<td>0.123</td>
<td>0.100</td>
<td>0.263</td>
</tr>
<tr>
<td>0 vs 9</td>
<td>0.082</td>
<td>0.101</td>
<td>0.092</td>
<td>0.279</td>
</tr>
</tbody>
</table>

Table 3: The simulation results of the 0-1 loss with $\beta = 0.5$. 50 trials are conducted for each pair of a method and dataset. Standard errors (multiplied by $10^4$) are shown in parentheses. Bold-faces indicate outperforming methods, chosen by one-sided t-test with the significant level 5%.

<table>
<thead>
<tr>
<th></th>
<th>Ramp</th>
<th>Sigmoid</th>
<th>Hinge</th>
<th>Logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 vs 1</td>
<td>0.012</td>
<td>0.005</td>
<td>0.038</td>
<td>0.228</td>
</tr>
<tr>
<td>0 vs 2</td>
<td>0.050</td>
<td>0.059</td>
<td>0.058</td>
<td>0.206</td>
</tr>
<tr>
<td>0 vs 3</td>
<td>0.047</td>
<td>0.054</td>
<td>0.064</td>
<td>0.229</td>
</tr>
<tr>
<td>0 vs 4</td>
<td>0.028</td>
<td>0.029</td>
<td>0.032</td>
<td>0.184</td>
</tr>
<tr>
<td>0 vs 5</td>
<td>0.117</td>
<td>0.185</td>
<td>0.117</td>
<td>0.193</td>
</tr>
<tr>
<td>0 vs 6</td>
<td>0.060</td>
<td>0.080</td>
<td>0.063</td>
<td>0.206</td>
</tr>
<tr>
<td>0 vs 7</td>
<td>0.027</td>
<td>0.027</td>
<td>0.045</td>
<td>0.214</td>
</tr>
<tr>
<td>0 vs 8</td>
<td>0.050</td>
<td>0.054</td>
<td>0.054</td>
<td>0.186</td>
</tr>
<tr>
<td>0 vs 9</td>
<td>0.040</td>
<td>0.044</td>
<td>0.046</td>
<td>0.192</td>
</tr>
</tbody>
</table>