MAGISTERARBEIT | MASTER'S THESIS

Titel | Title

Properties and Applications of the Levy Gauge

verfasst von | submitted by

Tihomir Todorov BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Magister der Sozial- und Wirtschaftswissenschaften (Mag.rer.soc.oec.)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 951
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Magisterstudium Statistik
Betreut von | Supervisor:	Univ.-Prof. Mag. Dr. Hannes Leeb

Acknowledgements

Completing your Master's thesis feels like one of those milestones that prompt you to make some kind of retrospection and to reassure yourself that the choices you have made along the way have led you to the right place. This occasion invites sentimentality; however, I will do my best to keep it within reasonable bounds while expressing my gratitude to those who truly deserve it.

First and foremost, I would like to thank my supervisor, Prof. Hannes Leeb, who suggested this fascinating topic. I highly appreciate his steady involvement, resulting in numerous useful comments, in-depth discussions and recommendations for improvement. His motivating attitude made the entire process of writing this thesis both a responsibility and a pleasure.

There are two more people to whom I owe a great deal: Nicolai Amann, whose research is closely related to the Levy gauge and who provided invaluable guidance when I needed it; and Georg Köstenberger, who gave me constructive advice on my optional subjects and my first presentations of scientific papers. Thank you both for being a true inspiration!

I cannot move forward without mentioning my fellow students. My sincere thanks go to Cornelius for being the best ally in all study-related activities in this Master, expressed in thousands of written blackboards on the 6th floor. To Lorenz, whose interest in the Levy gauge led to a late-night conversation about some tiny details. To Yavor, with whom you can talk about literally anything whenever you need it, as long as some pieces of cake are included. To my study buddies from the Bachelor studies with whom we are still in touch, Mikail, Moritz, and Seungchan, for their genuine friendship. I am grateful for the opportunity to share wonderful moments with each of you, and wish you all the best in your future endeavours!

Though a bit distant in time, I would like to thank my school maths teacher, Ivelina Romanova. I was fortunate to have such an open-minded teacher for eight years, whose liberal approach but high demands laid the mathematical foundations I still rely on.

Finally, my deepest thanks go to my parents, Katya and Tihomir, for their unconditional support and for showing me the importance of standing by my principles, shaped by independent and conscious decisions.

Abstract

The Levy gauge was first introduced in Amann, Leeb and Steinberger 2025 as a useful tool for the proofs of statistical guarantees for methods of uncertainty quantification, such as the Jackknife and the Jackknife+, with a focus on providing asymptotically valid prediction intervals conditional on the training data. Despite its connection to the well-known Levy metric and the Kolmogorov distance between two cumulative distribution functions, the Levy gauge itself is not a metric. Nevertheless, it possesses a broad range of properties - some related to the Levy metric and others specific to the Levy gauge - which provide it with sufficient flexibility. In this sense, the Levy gauge can be applied to the framework of distorted prediction intervals, and, under mild assumptions, be related to the classical concept of convergence in probability. The latter is used to derive the asymptotic conditional conservativeness of prediction intervals obtained via the Jackknife and the Jackknife+. This Master's thesis considers all these concepts in detail, providing rigorous proofs, illustrative examples, and clarifications and extensions of ideas and connections implicitly mentioned in the original paper.

Kurzfassung

Das Levy Gauge wurde erstmals im Paper von Amann, Leeb and Steinberger 2025 als nützliches Beweiszeug zur statistischen Garantien der Methoden für Unsicherheitsquantifizierung, wie Jackknife und Jackknife+, eingeführt, mit dem Fokus auf die Bereitstellung asymptotisch gültiger Vorhersageintervallen, gegeben die Trainingsdaten. Trotz seiner Verbindung zur bekannten Levy-Metrik und der Kolmogorovschen Distanz zwischen zwei Verteilungsfunktionen, ist das Levy Gauge selbst keine Metrik. Dennoch besitzt dies eine Vielzahl von Eigenschaften, sowohl welche, die mit der Levy-Metrik verbunden sind, als auch spezifische, die ihm ausreichende Flexibilität verleihen. In diesem Sinne kann man das Levy Gauge zur Einstellung von verzerrten Vorhersageintervallen anwenden, und, unter milden Annahmen, mit dem klassischen Konzept der Konvergenz in Wahrscheinlichkeit verbinden. Letzteres wird verwendet, um die asymptotische bedingte Konservativität von Vorhersageintervallen abzuleiten, die durch Jackknife und Jackknife+ erhalten werden. Diese Masterarbeit untersucht alle diese Konzepte eingehend; sie liefert detailierte Beweise, anschauliche Besipiele sowie Klarstellung und Erweiterungen von Ideen und Zusammenhängen, die im Originalpaper implizit erwähnt wurden.

Contents

1 Introduction

Given two cumulative distribution functions, one may be interested in measuring the distance between them in a given sense. Useful and well-known tools in this direction are the Levy metric and the Kolmogorov distance. However, they are not flexible enough to provide the statistician with sufficient information in the framework of a prediction interval for a future outcome, based on some known training data. For example, if we consider the distribution function F of the prediction error conditional on the training data, as well as some estimator $\hat{F}$, the Kolmogorov distance between F and $\hat{F}$ may become large if F has points of discontinuity that are not well-estimated, as discussed in Amann 2023.

A solution to this problem turns out to be the Levy gauge, which can be considered as a modification of the Levy metric. It is not a metric itself; however, it possesses a lot of useful properties, some preserved from the original object, others entirely new. Thus, the desired flexibility is guaranteed, especially if one aims to distort the prediction intervals in a plausible way to derive asymptotically valid results for methods of uncertainty quantification, such as the Jackknife and the Jackknife+.

One of the goals of this Master's thesis is to provide a deeper understanding to the Levy gauge, also by extending some ideas from Amann, Leeb and Steinberger 2025. It is important to point out that this thesis *does* build mostly on the concepts from the original paper. In particular, some of the presented results here have already occurred in the latter, as long as this coincides with the above-mentioned goal. To sum it up, some of the definitions and properties here may have already been stated and proved (or at least, sketched) in Amann, Leeb and Steinberger 2025, but this thesis extends the sketches by clarifying the steps of the proofs; in addition, it provides approachable examples and optimizes the proof techniques, where necessary.

The thesis is structured as follows. Chapter 2 introduces the Levy gauge and is divided into three sections. Section 2.1. defines the Levy gauge and shows its connection to the well-known Levy metric based on both their common properties and explicit differences. Section 2.2. gives more information about the relation between the Levy gauge and the convergence in distribution of given probability measures. Section 2.3. concentrates on the modification of the Levy metric that was made to obtain the Levy gauge and examines more precisely its influence on the latter. Chapter 3 directs the attention to the setting of prediction intervals and the respective general application of the Levy gauge there. Chapter 4 builds on the theoretical foundations of Chapter 3 and applies them to the Jackknife and Jackknife+ methods, in such a way providing the connection to Amann, Leeb and Steinberger 2025.

1.1 Notation

We introduce here some of the general notation which is repeatedly used throughout the Master's thesis. The specific notation is given in the beginning of the respective chapters or sections.

If we have some sequence, we denote by $\uparrow$ convergence from below and by $\downarrow$ convergence from above. If this convergence occurs in a specific setting, then the condition is stated under the arrow, otherwise it is assumed to hold when $n \to \infty$. The empty set is denoted by $\emptyset$. For a function $f : \mathbb{R} \to \mathbb{R}$ we will write $||f||_\infty$ for its supremum in absolute values, i.e., $||f||_\infty := \sup_{t\in\mathbb{R}} |f(t)|$. The indicator functions are denoted by $\mathbb{1}_A(t) = \mathbb{1}\{t \in A\}$, for $t \in \mathbb{R}$ and some set A. We prefer the first version mostly when A is a real-valued interval. We always write $\overline{\mathbb{R}}$ for the extended real line $[-\infty, +\infty]$. We use $[n]$ as a shorthand for the set of all natural numbers between 1 and n, i.e., $[n] = \{1, 2, \ldots, n\}$. The ceiling and floor functions are denoted by $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$, respectively.

If we have a cumulative distribution function F, we will denote by $F(t-)$ the limit from the left. That is, for some deterministic sequence $(t_n)_{n\in\mathbb{N}}$, with $t_n \uparrow t$, we have $F(t_n) \uparrow_{t_n\uparrow t} F(t-)$. Furthermore, if we have a sequence of random variables X_n that converges in probability to a random variable X, we will write $X_n \xrightarrow{p} X$. The abbreviation $X_n = o_p(t_n)$ would mean that the sequence $(X_n t_n^{-1})_{n\in\mathbb{N}}$ converges to 0 in probability.

2 The Levy Gauge

2.1 Definitions and important properties

Definition 2.1 (Levy metric)**.** *Let F, G be cumulative distribution functions. Then the Levy metric is defined as*

$$\mathcal{L}(F,G) := \inf \tau(F,G),$$

where

$$\tau(F,G) := \{\varepsilon \geq 0 : \forall t \in \mathbb{R} : F(t-\varepsilon) - \varepsilon \leq G(t) \leq F(t+\varepsilon) + \varepsilon\}.$$

Definition 2.2 (Levy gauge)**.** *Let F, G be cumulative distribution functions and $\delta \geq 0$. Then the Levy gauge is defined as*

$$\mathcal{L}_\delta(F,G) := \inf \tau_\delta(F,G),$$

where

$$\tau_\delta(F,G) := \{\varepsilon \geq 0 : \forall t \in \mathbb{R} : F(t-\delta) - \varepsilon \leq G(t) \leq F(t+\delta) + \varepsilon\}.$$

From that point on, if the cumulative distribution functions are clear from the context, we will write only $\mathcal{L}$ and τ, or $\mathcal{L}_\delta$ and τ_δ, respectively.

Remark 2.3. The sets τ and τ_δ are non-empty for any choice of cumulative distribution functions, since $1 \in \tau$ and $1 \in \tau_\delta$. It follows now from Definitions 2.1 and 2.2 that $\mathcal{L} \in [0,1]$ and $\mathcal{L}_\delta \in [0,1]$. Furthermore, if some $\varepsilon^* \in \tau_\delta$, then for any $\varepsilon \geq \varepsilon^*$ it holds that $\varepsilon \in \tau_\delta$ (or τ, respectively). Thus, τ_δ and τ are non-empty intervals bounded on the left by 0 and unbounded on the right.

Lemma 2.4 (Symmetry of the Levy gauge)**.** *For any cumulative distribution functions F, G and $\delta \geq 0$ we have*

$$\mathcal{L}_\delta(F,G) = \mathcal{L}_\delta(G,F).$$

Proof. Let $\varepsilon \geq 0$ such that

$$\forall t \in \mathbb{R} : F(t-\delta) - \varepsilon \overset{(i)}{\leq} G(t) \overset{(ii)}{\leq} F(t+\delta) + \varepsilon.$$

In addition, fix $t_0 \in \mathbb{R}$ and consider the following:

Set $t = t_0 + \delta$ and note that $F(t_0) \leq G(t_0+\delta) + \varepsilon$ in view of (i). Similarly, set $t = t_0 - \delta$, implying $G(t_0 - \delta) - \varepsilon \leq F(t_0)$ by (ii).

Since this holds true for any choice of $t_0 \in \mathbb{R}$, we have

$$\forall t_0 \in \mathbb{R} : G(t_0 - \delta) - \varepsilon \leq F(t_0) \leq G(t_0 + \delta) + \varepsilon.$$

This shows that $\tau_\delta(F,G) \subseteq \tau_\delta(G,F)$. Exchanging F and G gives now $\tau_\delta(G,F) \subseteq \tau_\delta(F,G)$ and the claim follows. □

Corollary 2.5 (Symmetry of the Levy metric)**.** *The Levy metric is symmetric in the sense that* $\mathcal{L}(F,G) = \mathcal{L}(G,F)$ *for any cumulative distribution functions* F *and* G*.*

Proof. Repeat the proof of Lemma 2.4 with δ replaced by ε and τ_δ by τ, respectively. □

Remark 2.6. The proofs of Lemma 2.4 and Corollary 2.5 give alternative representations of $\tau(F,G)$ and $\tau_\delta(F,G)$:

$$\tau(F,G) = \{\varepsilon \geq 0 : \forall t \in \mathbb{R} : G(t) \leq F(t+\varepsilon) + \varepsilon, F(t) \leq G(t+\varepsilon) + \varepsilon\},$$
$$\tau_\delta(F,G) = \{\varepsilon \geq 0 : \forall t \in \mathbb{R} : F(t) \leq G(t+\delta) + \varepsilon, G(t) \leq F(t+\delta) + \varepsilon\}.$$

The next two results regard the attainment of the infimum both for the Levy metric and the Levy gauge.

Lemma 2.7. *The infimum in the Levy gauge* $\mathcal{L}_\delta(F,G) = \inf \tau_\delta(F,G)$ *is attained.*

Proof. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence converging from above to the Levy gauge, i.e., $\varepsilon_n \downarrow \mathcal{L}_\delta$. Now we have:

$$\forall t \in \mathbb{R} : \underbrace{F(t-\delta) - \varepsilon_n}_{\uparrow F(t-\delta)-\mathcal{L}_\delta} \leq G(t) \leq \underbrace{F(t+\delta) + \varepsilon_n}_{\downarrow F(t+\delta)+\mathcal{L}_\delta},$$

proving the claim. □

Lemma 2.8. *The infimum in the Levy metric* $\mathcal{L}(F,G) = \inf \tau(F,G)$ *is attained.*

Proof. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence such that $\varepsilon_n \downarrow \mathcal{L}$. Use the alternative representation from Remark 2.6 and the right continuity of the cumulative distribution functions F and G to obtain

$$\forall t \in \mathbb{R} : F(t) \leq \underbrace{G(t+\varepsilon_n) + \varepsilon_n}_{\downarrow G(t+\mathcal{L})+\mathcal{L}} \quad \text{and} \quad \forall t \in \mathbb{R} : G(t) \leq \underbrace{F(t+\varepsilon_n) + \varepsilon_n}_{\downarrow F(t+\mathcal{L})+\mathcal{L}}.$$

This proves the claim. □

Together with Remark 2.3, we thus have $\tau_\delta(F,G) = [\mathcal{L}_\delta(F,G), +\infty)$ and $\tau(F,G) = [\mathcal{L}(F,G), +\infty)$.

We can now formulate an alternative definition of the Levy gauge which turns out to be more useful in certain settings.

Lemma 2.9 (Alternative definition of the Levy gauge)**.** *For given* $\delta \geq 0$ *and cumulative distribution functions* F *and* G *it holds that*

$$\mathcal{L}_\delta(F,G) = \sup_{t\in\mathbb{R}} \max\{F(t) - G(t+\delta), G(t) - F(t+\delta)\}.$$

Proof. It follows from Remarks 2.3, 2.6 and Lemma 2.7 that

$$\begin{aligned}\tau_\delta(F,G) &= \{\varepsilon \geq 0 : \forall t \in \mathbb{R} : F(t) \leq G(t+\delta)+\varepsilon, G(t) \leq F(t+\delta)+\varepsilon\}\\ &= \{\varepsilon \geq 0 : \forall t \in \mathbb{R} : \max\{F(t)-G(t+\delta), G(t)-F(t+\delta)\} \leq \varepsilon\}\\ &= \{\varepsilon \geq 0 : \sup_{t\in\mathbb{R}} \max\{F(t)-G(t+\delta), G(t)-F(t+\delta)\} \leq \varepsilon\}\\ &= \left[\max\left\{0, \sup_{t\in\mathbb{R}} \max\{F(t)-G(t+\delta), G(t)-F(t+\delta)\}\right\}, +\infty\right).\end{aligned}$$

Note that for $t \to \infty$ we have $F(t) - G(t+\delta) \underset{t\to\infty}{\longrightarrow} 0$, implying that

$$\sup_{t\in\mathbb{R}} \max\{F(t)-G(t+\delta), G(t)-F(t+\delta)\} \geq 0.$$

Hence,

$$\tau_\delta(F,G) = \left[\sup_{t\in\mathbb{R}} \max\{F(t)-F(t+\delta), G(t)-F(t+\delta)\}, +\infty\right),$$

and since the Levy gauge is the attained infimum of $\tau_\delta(F,G)$, the claim follows. □

Remark 2.10. For any fixed δ it holds that $\max\{F(t)-G(t+\delta), G(t)-F(t+\delta)\}$ is right-continuous in $t \in \mathbb{R}$. In addition, $\mathbb{Q}$ is dense in $\mathbb{R}$, which means that we can take in Lemma 2.9 the supremum over $\mathbb{Q}$, in other words, over a set consisting of countably many elements.

Up to now only similar properties of the Levy metric and the Levy gauge have been considered. For the purposes of the further comparison between the two objects, it is essential to explore the explicit connection between them.

Lemma 2.11. *Let $\mathcal{L}(F,G)$ and $\mathcal{L}_\delta(F,G)$ be defined as in Definitions 2.1 and 2.2. Then the following statement holds:*

$$\min\{\delta, \mathcal{L}_\delta\} \leq \mathcal{L} \leq \max\{\delta, \mathcal{L}_\delta\}. \tag{2.1}$$

In particular, if $\delta \geq \mathcal{L}$, then $\mathcal{L}_\delta \leq \mathcal{L}$.

Proof. In the sense of Lemmata 2.7 and 2.8, the infimum is attained, that is,

$$\forall t \in \mathbb{R} : F(t-\delta) - \mathcal{L}_\delta \leq G(t) \leq F(t+\delta) + \mathcal{L}_\delta, \tag{2.2}$$

$$\forall t \in \mathbb{R} : F(t-\mathcal{L}) - \mathcal{L} \leq G(t) \leq F(t+\mathcal{L}) + \mathcal{L}. \tag{2.3}$$

We will first prove the statement $\mathcal{L} \leq \max\{\delta, \mathcal{L}_\delta\}$. To do this, set $\varepsilon = \max\{\delta, \mathcal{L}_\delta\}$. Since F and G are monotonically non-decreasing, we have

$$\forall t \in \mathbb{R} : F(t-\varepsilon) - \varepsilon \leq F(t-\delta) - \mathcal{L}_\delta \leq G(t) \leq F(t+\delta) + \mathcal{L}_\delta \leq F(t+\varepsilon) + \varepsilon, \tag{2.4}$$

implying that $\varepsilon \in \tau$. Now it follows that

$$\mathcal{L} = \inf \tau \leq \varepsilon = \max\{\delta, \mathcal{L}_\delta\},$$

which yields the result.
The first inequality in (2.1), i.e., $\min\{\delta, \mathcal{L}_\delta\} \leq \mathcal{L}$, holds trivially in the case $\delta < \mathcal{L}$. Now assume that $\delta \geq \mathcal{L}$. We can use an argumentation similar to (2.4) to conclude that

$$\forall t \in \mathbb{R} : F(t-\delta) - \mathcal{L} \leq F(t-\mathcal{L}) - \mathcal{L} \leq G(t) \leq F(t+\mathcal{L}) + \mathcal{L} \leq F(t+\delta) + \mathcal{L}.$$

Then $\mathcal{L} \in \tau_\delta$, meaning that $\mathcal{L}_\delta \leq \mathcal{L}$. This finishes the proof. □

Lemma 2.11 allows us to derive an alternative representation of the Levy metric, based primarily on the Levy gauge.

Corollary 2.12. *For arbitrary cumulative distribution functions F and G it holds that*

$$\mathcal{L}(F,G) = \inf\{\delta \geq 0 : \delta \geq \mathcal{L}_\delta(F,G)\}.$$

Proof. In view of Remark 2.3, it holds $\mathcal{L}_1 \leq 1$ for any choice of F and G, implying that $\{\delta \geq 0 : \delta \geq \mathcal{L}_\delta(F,G)\} \neq \emptyset$. In addition, this set is also bounded from below by 0.
It is clear that the condition $\delta \geq \mathcal{L}_\delta$ is equivalent to $\max\{\delta, \mathcal{L}_\delta\} = \delta$. Now use the second inequality from (2.1) to get

$$\mathcal{L} \leq \delta, \quad \forall \delta \geq 0 : \delta \geq \mathcal{L}_\delta.$$

This yields $\mathcal{L} \leq \inf\{\delta \geq 0 : \delta \geq \mathcal{L}_\delta\}$.
It remains to prove that $\mathcal{L} \geq \inf\{\delta \geq 0 : \delta \geq \mathcal{L}_\delta\}$. To see this, note that the additional result of Lemma 2.11 gives $\mathcal{L}_\mathcal{L} \leq \mathcal{L}$. Hence,

$$\mathcal{L} \in \{\delta \geq 0 : \delta \geq \mathcal{L}_\delta\},$$

which proves the claim. □

Remark 2.13. Corollary 2.12 gives another interpretation of the Levy metric in view of fixed point equations. In general, the existence of a solution of the fixed point equation $\delta = \mathcal{L}_\delta(F,G)$ is not guaranteed. However, under the assumption that there exist some solutions, it holds that the Levy metric $\mathcal{L}(F,G)$ is the smallest among them. More details to this topic are given in Section 2.3.

Corollary 2.12 shows that the Levy metric $\mathcal{L}(F,G)$ can be easily computed if we have the information about the Levy gauge. Furthermore, it turns out that based on $\mathcal{L}_\delta(F,G)$, we can also derive the Kolmogorov distance between F and G.

Definition 2.14 (Kolmogorov distance)**.** *The Kolmogorov distance between two cumulative distribution functions F and G is defined as*

$$||F-G||_\infty := \sup_{t\in\mathbb{R}} |F(t) - G(t)|.$$

Lemma 2.15. $\mathcal{L}_0(F,G) = ||F-G||_\infty$.

Proof. Use Definitions 2.2 and 2.14 to get

$$\begin{aligned}\mathcal{L}_0(F,G) &= \inf\{\varepsilon \geq 0 : \forall t \in \mathbb{R} : F(t) - \varepsilon \leq G(t) \leq F(t) + \varepsilon\} \\ &= \inf\{\varepsilon \geq 0 : \forall t \in \mathbb{R} : |F(t) - G(t)| \leq \varepsilon\} \\ &= \inf\{\varepsilon \geq 0 : \sup_{t\in\mathbb{R}} |F(t) - G(t)| \leq \varepsilon\} \\ &= ||F - G||_\infty.\end{aligned}$$

□

Example 2.16. *Let $F(t) = \mathbb{1}_{[\theta,+\infty)}(t)$ and $\hat{F}(t) = \mathbb{1}_{[\hat{\theta},+\infty)}(t)$, where $\hat{\theta}$ is some estimator for θ, such that for some $\varepsilon > 0$ it holds almost surely that $0 < |\hat{\theta} - \theta| \leq \varepsilon$. In other words, $\hat{\theta}$ is a "good" estimator for θ but does not obtain it exactly (a.s.). Then it holds that*

$$||F - \hat{F}||_\infty = \sup_{t\in\mathbb{R}} |F(t) - \hat{F}(t)| = 1.$$

Thus, the Kolmogorov distance can be large even if the point of discontinuity of F is relatively well estimated.

It has already been pointed out that the Levy gauge is a modification of the Levy metric. In this context, fixing the argument shift to be some non-negative δ results in the violating of one of the conditions for a metric, more precisely the triangle inequality. However, we can find a relaxed version of the triangle inequality. These observations are demonstrated below.

Lemma 2.17 (Relaxed triangle inequality)**.** *For given $\delta_1, \delta_2 \geq 0$ and cumulative distribution functions F, G and H, the following inequality is satisfied:*

$$\mathcal{L}_{\delta_1+\delta_2}(F,G) \leq \mathcal{L}_{\delta_1}(F,H) + \mathcal{L}_{\delta_2}(H,G).$$

Proof. Recall that the infimum is attained in view of Lemma 2.7. Fix $t \in \mathbb{R}$ and use the alternative representation of τ_δ, given in Remark 2.6 at t and $t + \delta_1$, respectively, to conclude

$$F(t) \leq H(t+\delta_1) + \mathcal{L}_{\delta_1}(H,F) \quad \text{and} \quad H(t+\delta_1) \leq G(t+\delta_1+\delta_2) + \mathcal{L}_{\delta_2}(G,H).$$

Applying the symmetry of the Levy gauge (Lemma 2.4) yields

$$F(t) \leq H(t+\delta_1) + \mathcal{L}_{\delta_1}(F,H) \leq G(t+\delta_1+\delta_2) + \mathcal{L}_{\delta_1}(F,H) + \mathcal{L}_{\delta_2}(H,G). \tag{2.5}$$

With the same arguments as in (2.5), we obtain

$$G(t) \leq H(t+\delta_2) + \mathcal{L}_{\delta_2}(H,G) \leq F(t+\delta_2+\delta_1) + \mathcal{L}_{\delta_1}(F,H) + \mathcal{L}_{\delta_2}(H,G). \tag{2.6}$$

It remains to combine (2.5) and (2.6) to see that

$$\mathcal{L}_{\delta_1}(F,H) + \mathcal{L}_{\delta_2}(H,G) \in \tau_{\delta_1+\delta_2}(F,G).$$

This proves the statement. □

Remark 2.18. Note that for any $\delta \geq 0$ it holds

$$\mathcal{L}_{2\delta}(F, G) \leq \mathcal{L}_{\delta}(F, H) + \mathcal{L}_{\delta}(H, G).$$

Moreover, the inequality given in Lemma 2.17 is sharp in the sense that there does not exist $\alpha \geq 0$, such that $\alpha < \delta_1 + \delta_2$ and

$$\mathcal{L}_{\alpha}(F, G) \leq \mathcal{L}_{\delta_1}(F, H) + \mathcal{L}_{\delta_2}(H, G),$$

for any F, G and H.

We can illustrate this in a simple setting provided by the example below.

Example 2.19. *Let $F(t) = \mathbb{1}_{[0,+\infty)}(t), H(t) = F(t + \delta_1) = \mathbb{1}_{[-\delta_1,+\infty)}(t)$, and $G(t) = H(t + \delta_2) = \mathbb{1}_{[-\delta_1-\delta_2,+\infty)}(t)$, for some $\delta_1, \delta_2 > 0$. It follows from the alternative definition of the Levy gauge given in Lemma 2.9 that $\mathcal{L}_{\delta_1}(F, H) = 0$ and $\mathcal{L}_{\delta_2}(H, G) = 0$.*
Let now $\alpha < \delta_1 + \delta_2$. Consider

$$\begin{aligned}
\mathcal{L}_{\alpha}(F, G) &= \sup_{t\in\mathbb{R}} \max\{F(t) - G(t + \alpha), G(t) - F(t + \alpha)\} \\
&= \sup_{t\in\mathbb{R}} \max\{F(t) - F(t + \delta_1 + \delta_2 + \alpha), F(t + \delta_1 + \delta_2) - F(t + \alpha)\} \\
&= \sup_{t\in\mathbb{R}} \{F(t + \delta_1 + \delta_2) - F(t + \alpha)\},
\end{aligned}$$

due to the monotonicity of F. Choosing some $t^* \in (-\delta_1 - \delta_2, -\alpha)$ gives

$$F(t^* + \delta_1 + \delta_2) - F(t^* + \alpha) = 1,$$

entailing $\mathcal{L}_{\alpha}(F, G) = 1$. Since $\mathcal{L}_{\alpha}(F, G) > \mathcal{L}_{\delta_1}(F, H) + \mathcal{L}_{\delta_2}(H, G)$ for an arbitrary $\alpha < \delta_1 + \delta_2$, the result aligns with Remark 2.18.

We will examine further the Levy gauge in the sense of some properties related to the choice of the cumulative distribution functions.

Lemma 2.20 (Scaling invariance)**.** *The Levy gauge is scaling invariant, that is,*

$$\mathcal{L}_{\delta}(F, G) = \mathcal{L}_{\frac{\delta}{\alpha}}(F(\cdot\alpha), G(\cdot\alpha)),$$

for given $\alpha > 0, \delta \geq 0$ and cumulative distribution functions F and G.

Proof. First, we will show that for $\alpha > 0$ it holds

$$\mathcal{L}_{\frac{\delta}{\alpha}}(F(\cdot\alpha), G(\cdot\alpha)) \leq \mathcal{L}_{\delta}(F(\cdot), G(\cdot)).$$

Note that $F(t \pm \delta) = F\left(\frac{\alpha}{\alpha}(t \pm \delta)\right) = F\left(\left(\frac{t}{\alpha} \pm \frac{\delta}{\alpha}\right)\alpha\right)$ and $G(t) = G\left(\frac{t}{\alpha}\alpha\right)$. Set $\tilde{t} = \frac{t}{\alpha}$ for any $t \in \mathbb{R}$. In view of (2.2) we conclude

$$\forall \tilde{t} \in \mathbb{R} : F\left(\left(\tilde{t} - \frac{\delta}{\alpha}\right)\alpha\right) - \mathcal{L}_{\delta}(F, G) \leq G(\tilde{t}\alpha) \leq F\left(\left(\tilde{t} + \frac{\delta}{\alpha}\right)\alpha\right) + \mathcal{L}_{\delta}(F, G).$$

Hence,

$$\mathcal{L}_\delta(F(\cdot),G(\cdot)) \in \tau_{\frac{\delta}{\alpha}}(F(\cdot\alpha),G(\cdot\alpha)),$$

or, equivalently,

$$\mathcal{L}_{\frac{\delta}{\alpha}}(F(\cdot\alpha),G(\cdot\alpha)) \leq \mathcal{L}_\delta(F(\cdot),G(\cdot)).$$

The result above holds for the functions $F(\cdot),G(\cdot)$, and a scale factor $\alpha > 0$. Consider instead the functions $F(\cdot\alpha),G(\cdot\alpha)$ and the scale factor $\frac{1}{\alpha} > 0$. Setting $\tilde{\delta} = \frac{\delta}{\alpha} \geq 0$ yields

$$\begin{aligned}\mathcal{L}_{\frac{\tilde{\delta}}{1/\alpha}}\left(F\left(\cdot\alpha\frac{1}{\alpha}\right),G\left(\cdot\alpha\frac{1}{\alpha}\right)\right) &\leq \mathcal{L}_{\tilde{\delta}}(F(\cdot\alpha),G(\cdot\alpha))\\ \mathcal{L}_\delta(F(\cdot),G(\cdot)) &\leq \mathcal{L}_{\frac{\delta}{\alpha}}(F(\cdot\alpha),G(\cdot\alpha)),\end{aligned}$$

and the statement follows. □

It remains to see what happens when we do not take two arbitrary cumulative distribution functions F and G but observe a more general structure. That raises questions regarding convexity, which are summarized in the following.

Lemma 2.21 (Convexity)**.** *The Levy gauge $\mathcal{L}_\delta(\cdot,\cdot),\delta \geq 0$ is convex.*

Proof. Consider the pairs of distribution functions $(F_i,G_i), i = 1,2$, as well as the convex combination

$$\alpha(F_1,G_1) + (1-\alpha)(F_2,G_2) = \Big(\alpha F_1 + (1-\alpha)F_2, \alpha G_1 + (1-\alpha)G_2\Big), \alpha \in [0,1].$$

We have to show

$$\mathcal{L}_\delta\Big(\alpha F_1 + (1-\alpha)F_2, \alpha G_1 + (1-\alpha)G_2\Big) \leq \alpha\mathcal{L}_\delta(F_1,G_1) + (1-\alpha)\mathcal{L}_\delta(F_2,G_2).$$

In view of (2.2) we get

$$\forall t \in \mathbb{R}: \quad F_1(t-\delta) - \mathcal{L}_\delta(F_1,G_1) \leq G_1(t) \leq F_1(t+\delta) + \mathcal{L}_\delta(F_1,G_1) \tag{2.7}$$

$$\forall t \in \mathbb{R}: \quad F_2(t-\delta) - \mathcal{L}_\delta(F_2,G_2) \leq G_2(t) \leq F_2(t+\delta) + \mathcal{L}_\delta(F_2,G_2). \tag{2.8}$$

Consider the expression $\alpha\cdot$ (2.7) + $(1-\alpha)\cdot$ (2.8). Preserving all sides of the inequalities yields

$$\begin{aligned}\forall t \in \mathbb{R}: \quad &\alpha F_1(t-\delta) + (1-\alpha)F_2(t-\delta) - [\alpha\cdot\mathcal{L}_\delta(F_1,G_1) + (1-\alpha)\mathcal{L}_\delta(F_2,G_2)]\\ &\leq \alpha G_1(t) + (1-\alpha)G_2(t)\\ &\leq \alpha F_1(t+\delta) + (1-\alpha)F_2(t+\delta) + [\alpha\cdot\mathcal{L}_\delta(F_1,G_1) + (1-\alpha)\mathcal{L}_\delta(F_2,G_2)].\end{aligned}$$

We then have

$$\alpha\mathcal{L}_\delta(F_1,G_1) + (1-\alpha)\mathcal{L}_\delta(F_2,G_2) \in \tau_\delta(\alpha F_1 + (1-\alpha)F_2, \alpha G_1 + (1-\alpha)G_2).$$

This gives the claim. □

Corollary 2.22. *Consider balls of the type*

$$B(F,r) := \{G : \mathfrak{L}_\delta(F,G) \le r\},$$

with $\delta \ge 0$, radius $r \ge 0$, and cumulative distribution functions F and G. Then $B(F,r)$ is convex.

Proof. Consider the pairs of cumulative distribution functions (F, G_1) and (F, G_2), with $G_1, G_2 \in B(F,r)$. Applying Lemma 2.21 yields

$$\mathfrak{L}_\delta(F, \alpha G_1 + (1-\alpha)G_2) \le \alpha \mathfrak{L}_\delta(F, G_1) + (1-\alpha)\mathfrak{L}_\delta(F, G_2), \quad \alpha \in [0,1]. \tag{2.9}$$

Since $G_1, G_2 \in B(F,r)$, it follows that $\mathfrak{L}_\delta(F,G_1) \le r$ and $\mathfrak{L}_\delta(F,G_2) \le r$. Use (2.9) to get

$$\mathfrak{L}_\delta(F, \alpha G_1 + (1-\alpha)G_2) \le \alpha r + (1-\alpha)r = r, \quad \alpha \in [0,1]$$

or, equivalently,

$$\alpha G_1 + (1-\alpha)G_2 \in B(F,r), \quad \alpha \in [0,1],$$

which proves the claim. □

2.2 Relation of the Levy gauge to convergence concepts

Until now, only the right continuity property and the boundedness of F and G have been used. However, they are not arbitrarily chosen functions but cumulative distribution ones. In this context, it is essential to explore the meaning of the Levy metric and the Levy gauge with respect to major concepts of probability theory.

Definition 2.23 (Convergence in distribution)**.** *Consider a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ with corresponding cumulative distribution functions $(F_n)_{n\in\mathbb{N}}$ and another random variable X with the cumulative distribution function F. In addition, let $C \subseteq \mathbb{R}$ be the set of continuity points of F. One says that X_n converge in distribution to X if it holds that*

$$\forall t \in C : \lim_{n\to\infty} F_n(t) = F(t).$$

We will denote this by $X_n \xrightarrow{w} X$.

First, we will show the relation of the Levy metric to the convergence in distribution and we will use this result to prove the similar statement for the Levy gauge.

Theorem 2.24. *Let $(X_n)_{n\in\mathbb{N}}, (F_n)_{n\in\mathbb{N}}, X, F$ and C be as in Definition 2.23. Then we have the following equivalence:*

$$\mathfrak{L}(F_n, F) \underset{n\to\infty}{\longrightarrow} 0 \quad \Longleftrightarrow \quad X_n \xrightarrow{w} X.$$

Proof. Clearly, the Levy metric of the functions F_n and F depends on n. For the sake of readability, in the current proof we will denote $\mathcal{L}(F_n, F) = \mathcal{L}(F, F_n)$ only by $\mathcal{L}$.
$\boxed{\Rightarrow}$: Let t be a point of continuity of F (i.e., $t \in C \subseteq \mathbb{R}$). Since $\mathcal{L} \underset{n\to\infty}{\longrightarrow} 0$ by assumption, it is clear that $t \pm \mathcal{L} \underset{n\to\infty}{\longrightarrow} t$. In view of (2.3), we obtain

$$\forall t \in C : \underbrace{F(t-\mathcal{L})}_{\underset{n\to\infty}{\longrightarrow} F(t)} - \underbrace{\mathcal{L}}_{\underset{n\to\infty}{\longrightarrow} 0} \le F_n(t) \le \underbrace{F(t+\mathcal{L})}_{\underset{n\to\infty}{\longrightarrow} F(t)} + \underbrace{\mathcal{L}}_{\underset{n\to\infty}{\longrightarrow} 0}.$$

In other words, $\lim_{n\to\infty} F_n(t) = F(t)$ for any point of continuity t, which yields the first direction of the statement.
$\boxed{\Leftarrow}$: Let $\varepsilon > 0$ and extend the domain of F_n and F to $\overline{\mathbb{R}}$ under the conventions $F(-\infty) = F_n(-\infty) = 0$ and $F(\infty) = F_n(\infty) = 1$.
Choose real-valued points $\tilde{t}_0, \dots \tilde{t}_k, k < \infty$, such that

$$-\infty = \tilde{t}_0 < \tilde{t}_1 < \cdots < \tilde{t}_k = +\infty, \quad \text{with} \quad F(\tilde{t}_p-) - F(\tilde{t}_{p-1}) \le \frac{\varepsilon}{4}, \forall p \in [k].$$

The points $\tilde{t}_p$ are defined iteratively as follows:
Assume we have chosen $j-1$ points such that $F(\tilde{t}_p-) - F(\tilde{t}_{p-1}) \le \frac{\varepsilon}{4}, \forall p \in [j-1]$. Let

$$\tilde{t}_j := \sup\left\{t > \tilde{t}_{j-1} : F(t-) - F(\tilde{t}_{j-1}) < \frac{\varepsilon}{4}\right\}.$$

Note that $\tilde{t}_j$ is well-defined, since

$$F(t-) - F(\tilde{t}_{j-1}) = \mathbb{P}(\tilde{t}_{j-1} < X < t) \underset{t\downarrow\tilde{t}_{j-1}}{\downarrow} \mathbb{P}(X \in \emptyset) = 0.$$

If $\tilde{t}_j = \infty$, then we are done. Otherwise, we continue with the iterations.
It is not clear from the approach above that the procedure ends in finitely many steps, i.e., $k < \infty$. To prove this, we will consider the following
Claim: For $\tilde{t}_j < \infty$ it holds that $F(\tilde{t}_j) - F(\tilde{t}_{j-1}) \ge \frac{\varepsilon}{4}$.
The claim can be derived by contradiction. Assume that $F(\tilde{t}_j) - F(\tilde{t}_{j-1}) < \frac{\varepsilon}{4}$, for $\tilde{t}_j < \infty$. By the right continuity of F we have

$$\exists t > \tilde{t}_j : \quad F(t) - F(\tilde{t}_{j-1}) < \frac{\varepsilon}{4} \quad \text{and} \quad F(t-) - F(\tilde{t}_{j-1}) < \frac{\varepsilon}{4}.$$

Now we conclude

$$\tilde{t}_j < t < \sup\left\{t > \tilde{t}_{j-1} : F(t-) - F(\tilde{t}_{j-1}) < \frac{\varepsilon}{4}\right\} = \tilde{t}_j,$$

which is a contradiction.
Note further that $\tilde{t}_p, \forall p \in [k]$ are not necessarily continuity points of F. Due to the separability of $\mathbb{R}$ (C is a dense subset of $\mathbb{R}$), we can consider a new partition, as follows:

- $t_0 = \tilde{t}_0, t_k = \tilde{t}_k$.

- If $\tilde{t}_p \in C, p \in [k-1]$, set $t_p = \tilde{t}_p$.
- If $\tilde{t}_p \notin C$, choose $t_p > \tilde{t}_p$, such that $t_p \in C$ and $F(t_p) - F(\tilde{t}_{p-1}) \leq \frac{\varepsilon}{2}$.

Moreover, this construction guarantees that

$$F(\tilde{t}_{p+1}) - F(t_p) \leq F(\tilde{t}_{p+1}) - F(\tilde{t}_p) \leq \frac{\varepsilon}{4} < \frac{\varepsilon}{2}.$$

Thus, we have a new partition $t_0, \ldots, t_k$, $k < \infty$, such that

$$t_p \in C, \forall p \in 0 \cup [k] \quad \text{and} \quad F(t_p) - F(t_{p-1}) \leq \frac{\varepsilon}{2}.$$

In addition, it holds that the finite number of points k depends on ε, for instance, $k \leq \left\lfloor \frac{2}{\varepsilon} \right\rfloor$. Let now $x \in \mathbb{R}$, meaning that there exists $t_j, j \in [k]$, such that $x \in [t_{j-1}, t_j)$. In other words,

$$F_n(t_{j-1}) \leq F_n(x) \leq F_n(t_j) \quad \text{and} \quad F(t_{j-1}) \leq F(x) \leq F(t_j). \tag{2.10}$$

Furthermore, the assumption of convergence in distribution implies that

$$\exists n_0 = n_0(j, \varepsilon) \in \mathbb{N} : \forall n \geq n_0 : \quad F_n(t_j) - F(t_j) < \frac{\varepsilon}{2} \quad \text{and} \quad F_n(t_{j-1}) - F(t_{j-1}) > -\frac{\varepsilon}{2}. \tag{2.11}$$

Combining all these results, we conclude

$$\begin{aligned} \forall n \geq n_0 : F_n(x) - F(x+\varepsilon) \leq F_n(x) - F(x) &\overset{(2.10)}{\leq} F_n(t_j) - F(t_{j-1}) \\ &= \underbrace{F_n(t_j) - F(t_j)}_{\overset{(2.11)}{<} \frac{\varepsilon}{2}} + \underbrace{F(t_j) - F(t_{j-1})}_{\leq \frac{\varepsilon}{2}} < \varepsilon. \end{aligned}$$

Proceeding in a similar way, we can show that $\forall n \geq n_0 : F_n(x) - F(x-\varepsilon) > -\varepsilon$. To sum it up, we conclude

$$\forall n \geq n_0 : F(x-\varepsilon) - \varepsilon < F_n(x) < F(x+\varepsilon).$$

Since $k < \infty$, we can choose an appropriate $n_0(j, \varepsilon)$ for any $j \in [k]$ and take

$$N = \max_{j \in k} n_0(j, \varepsilon).$$

Then we have

$$\forall x \in \mathbb{R}, \forall n \geq N : F(x-\varepsilon) - \varepsilon < F_n(x) < F(x+\varepsilon) + \varepsilon,$$

giving that $\forall n \geq N : \varepsilon \in \tau(F_n, F)$, or $\mathcal{L}(F_n, F) \leq \varepsilon$. This gives $\limsup_{n\to\infty} \mathcal{L}(F_n, F) \leq \varepsilon$. Letting $\varepsilon \downarrow 0$ proves the claim. $\square$

Lemma 2.25. *Let $(X_n)_{n\in\mathbb{N}}, (F_n)_{n\in\mathbb{N}}, F$ and X and C be as in Definition 2.23. Then X_n converges to X in distribution (i.e., $X_n \xrightarrow{w} X$) if and only if $\mathcal{L}_\delta(F_n, F) \underset{n\to\infty}{\longrightarrow} 0$ for any $\delta > 0$.*

Proof. With Theorem 2.24 it is enough to prove the following statement:

$$\mathcal{L}(F_n, F) \underset{n\to\infty}{\longrightarrow} 0 \quad \Longleftrightarrow \quad \forall \delta > 0 : \mathcal{L}_\delta(F_n, F) \underset{n\to\infty}{\longrightarrow} 0.$$

$\boxed{\Rightarrow}$: By the first inequality of Lemma 2.11, we have

$$0 \le \min\{\delta, \mathcal{L}_\delta(F_n, F)\} \le \mathcal{L}(F_n, F).$$

Hence,

$$0 \le \limsup_{n\to\infty} \min\{\delta, \mathcal{L}_\delta(F_n, F)\} \le \limsup_{n\to\infty} \mathcal{L}(F_n, F) = \lim_{n\to\infty} \mathcal{L}(F_n, F) = 0.$$

This holds true for any $\delta > 0$, yielding the claim.
$\boxed{\Leftarrow}$: For the reverse direction, use the second inequality of Lemma 2.11 and take lim sup on both sides to conclude

$$0 \le \limsup_{n\to\infty} \mathcal{L}(F_n, F) \le \limsup_{n\to\infty} \max\{\delta, \mathcal{L}_\delta(F_n, F)\} = \delta.$$

This holds for all $\delta > 0$ and under the assumption that $\mathcal{L}_\delta(F_n, F) \underset{n\to\infty}{\longrightarrow} 0$. This finishes the proof. □

2.3 Influence of δ on the Levy Gauge $\mathcal{L}_\delta$

The results in this section are related to the influence of the shift δ on $\mathcal{L}_\delta(F, G)$. In other words, here we consider the map $\delta \mapsto \mathcal{L}_\delta(F, G)$ for arbitrary cumulative distribution functions F and G.

Lemma 2.26 (Monotonicity of the Levy gauge)**.** *The map $\delta \mapsto \mathcal{L}_\delta(F, G)$ is monotonically non-increasing. That is, for some nonnegative shifts $0 \le \delta \le \delta'$ it holds true that $\mathcal{L}_{\delta'}(F, G) \le \mathcal{L}_\delta(F, G)$.*

Proof. Recalling the fact that the infimum is attained (in view of Lemma 2.7 and (2.2)) and with the right continuity of F we conclude

$$\forall t \in \mathbb{R} : F(t - \delta') - \mathcal{L}_\delta \le F(t - \delta) - \mathcal{L}_\delta \le G(t) \le F(t + \delta) + \mathcal{L}_\delta \le F(t + \delta') + \mathcal{L}_\delta.$$

Thus, $\mathcal{L}_\delta \in \tau_{\delta'}$, proving the statement. □

The properties of the map $\delta \mapsto \mathcal{L}_\delta(F, G)$ can be further examined regarding continuity. It turns out that right continuity always occurs, as illustrated by the following.

Lemma 2.27. *It holds that $\mathcal{L}_\delta(F, G) = \lim_{\delta_n \downarrow \delta} \mathcal{L}_{\delta_n}(F, G)$, for any $\delta \ge 0$ and a sequence $(\delta_n)_{n\in\mathbb{N}}$, such that $\delta_n \downarrow \delta$.*

Proof. $\lim_{\delta_n \downarrow \delta} \mathcal{L}_{\delta_n}$ exists because of the monotonicity of the Levy gauge (Lemma 2.26) and together with the boundedness, given in Remark 2.3.
Let $\lambda = \lim_{\delta_n \downarrow \delta} \mathcal{L}_{\delta_n}$, with $\lambda \leq \mathcal{L}_\delta$, due to Lemma 2.26. It remains to show that $\mathcal{L}_\delta \leq \lambda$. We use the symmetry of $\mathcal{L}_\delta$ (Lemma 2.4) and (2.2) to get

$$\forall n \in \mathbb{N}, \forall t \in \mathbb{R}: \quad G(t) \leq F(t+\delta_n) + \mathcal{L}_{\delta_n}(F,G) \quad \text{and} \quad F(t) \leq G(t+\delta_n) + \mathcal{L}_{\delta_n}(F,G).$$

Letting $n \to \infty$ and using the right continuity of F and G now yield

$$\forall t \in \mathbb{R}: G(t) \leq F(t+\delta) + \lambda \quad \text{and} \quad F(t) \leq G(t+\delta) + \lambda.$$

Thus, we have $\lambda \in \tau_\delta$, or $\mathcal{L}_\delta \leq \lambda$, proving the statement. □

Remark 2.28. Provided that the solution of the fixed point equation $\delta = \mathcal{L}_\delta(F,G)$ exists (in view of Remark 2.13), and together with Lemmata 2.26 and 2.27, it is easy to see that this solution is unique. Examples 2.29 and 2.33 illustrate both cases (existence and non-existence of the solution, respectively) for given cumulative distribution functions F and G.

Until now, we have seen that the map $\delta \mapsto \mathcal{L}_\delta(F,G)$ is non-increasing and right-continuous. However, it is not left-continuous in general.

Example 2.29. *Let $F(t) = \mathbb{1}_{[0,+\infty)}(t)$ and $G(t) = \frac{1}{2}\mathbb{1}_{[-\frac{1}{4},+\infty)}(t) + \frac{1}{2}\mathbb{1}_{[\frac{1}{4},+\infty)}(t)$. We will consider two cases.*

- <u>*Case 1*</u>*: $\delta \in \left[0, \frac{1}{4}\right)$. For $\varepsilon \geq 0$ we have $\varepsilon \in \tau_\delta$ if and only if the inequalities from*

Table 2.1: Values to compare (Case 1)

t	$-\infty$	$-\frac{1}{4}$	$-\delta$	δ	$\frac{1}{4}$ $+\infty$
$F(t-\delta)-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$1-\varepsilon$	$1-\varepsilon$
$G(t)$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
$F(t+\delta)+\varepsilon$	ε	ε	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$

Definition 2.2 are satisfied. The yellow cells in Table 2.1 indicate the places where they do not follow directly. Hence, the Levy gauge is the smallest $\varepsilon \geq 0$, such that $\frac{1}{2} \leq \varepsilon, \forall t \in \left[-\frac{1}{4}, -\delta\right)$ and $1-\varepsilon \leq \frac{1}{2}, \forall t \in \left[\delta, \frac{1}{4}\right)$. In other words, we obtain $\mathcal{L}_\delta(F,G) = \frac{1}{2}$.

- <u>*Case 2*</u>*: $\delta \in \left[\frac{1}{4}, +\infty\right)$. We proceed in a similar way as in Case 1.*
  Since all inequalities are satisfied for any choice of $\varepsilon \geq 0$ (Table 2.2), it follows $\mathcal{L}_\delta(F,G) = 0$.

Combining both cases gives the following expression for the Levy gauge:

$$\mathcal{L}_\delta(F,G) = \begin{cases} \frac{1}{2}, & 0 \leq \delta < \frac{1}{4} \\ 0, & \delta \geq \frac{1}{4} \end{cases}.$$

Table 2.2: Values to compare (Case 2)

t	$-\infty$	$-\delta$	$-\frac{1}{4}$	$\frac{1}{4}$	δ	$+\infty$
$F(t-\delta)-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$1-\varepsilon$
$G(t)$	0	0	$\frac{1}{2}$	1	1
$F(t+\delta)+\varepsilon$	ε	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$

To sum it up, we have a map $\delta \mapsto \mathcal{L}_\delta(F,G)$, which is non-increasing and right-continuous (in the sense of Lemmata 2.26 and 2.27) but not left-continuous. Moreover, the fixed point equation $\delta = \mathcal{L}_\delta(F,G)$ does not have a solution, as one can see from the graphical representation, given in Figure 2.1.

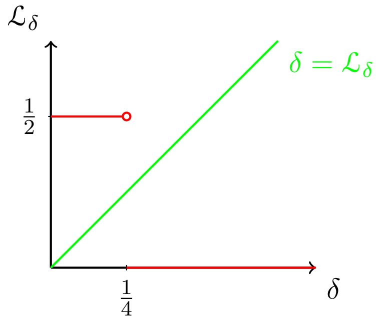

Figure 2.1: Visualisation of the map $\delta \mapsto \mathcal{L}_\delta$ (in red)

However, the Levy metric $\mathcal{L}(F,G)$ is well-defined. To see this, use the representation of the Levy metric from Corollary 2.12 to get

$$\mathcal{L}(F,G) = \inf\{\delta \geq 0 : \delta \geq \mathcal{L}_\delta(F,G)\} = \frac{1}{4}.$$

Example 2.29 raises the question whether we can find some mild assumptions which are necessary to guarantee the continuity of the map $\delta \mapsto \mathcal{L}_\delta(F,G)$. Moreover, in view of Remarks 2.13 and 2.28, this would also guarantee the existence of the solution of the fixed point equation $\delta = \mathcal{L}_\delta(F,G)$.
One possible assumption is the continuity of at least one of the distribution functions F and G. The exact choice of the function does not play a role due to the symmetry of the Levy gauge. We will summarize this in the following result.

Lemma 2.30. *Consider $\mathcal{L}_\delta(F,G)$, where $\delta \geq 0$ and F is continuous. Show that the map $\delta \mapsto \mathcal{L}_\delta(F,G)$ is also continuous.*

Proof. First, we will show that for any sequence $(\delta_n)_{n\in\mathbb{N}}$, with $\delta_n \underset{n\to\infty}{\longrightarrow} \delta, \delta \in \mathbb{R}$ it holds

$$\sup_{t\in\mathbb{R}} |F(t+\delta_n) - F(t+\delta)| \underset{n\to\infty}{\longrightarrow} 0. \tag{2.12}$$

We will further use the convention that $F(-\infty) = 0, F(+\infty) = 1$.
We will prove the claim by contradiction. Assume that there exist a sequence $(t_n)_{n\in\mathbb{N}}$, a subsequence of indices $(n_k)_{k\in\mathbb{N}}$, and a constant $c > 0$, such that

$$|F(t_{n_k} + \delta_{n_k}) - F(t_{n_k} + \delta)| > c, \quad \forall n_k \geq N, N \in \mathbb{N}.$$

Under the assumption that $(t_{n_k})_{k\in\mathbb{N}}$ converges in $\overline{\mathbb{R}}$, we can distinguish between two cases.

- Case 1: $t_{n_k} \underset{k\to\infty}{\longrightarrow} t, t \in \mathbb{R}$. Then we get

$$|F(t_{n_k} + \delta_{n_k}) - F(t_{n_k} + \delta)| \underset{k\to\infty}{\longrightarrow} |F(t + \delta) - F(t + \delta)| = 0,$$

  which is a contradiction to the assumption.

- Case 2: $t_{n_k} \underset{k\to\infty}{\longrightarrow} \pm\infty$. Then we have

$$|F(t_{n_k} + \delta_{n_k}) - F(t_{n_k} + \delta)| \underset{k\to\infty}{\longrightarrow} |F(\pm\infty) - F(\pm\infty)| = 0,$$

  which is again a contradiction.

Note that $(t_{n_k})_{k\in\mathbb{N}}$ does not necessarily converge. However, $\overline{\mathbb{R}}$ is compact and there exists a converging subsequence $(t_{n_{k_i}})_{i\in\mathbb{N}}$ (Bolzano-Weierstrass setting). Therefore, we can refer to the cases above by taking $t_{n_{k_i}}$ instead of t_{n_k}.
It is now enough to prove the left continuity of the map $\delta \mapsto \mathcal{L}_\delta(F, G)$. Let $(\delta_n^-)_{n\in\mathbb{N}}$ be a sequence that converges from below to δ, i.e., $\delta_n^- \uparrow \delta$. In addition, fix some $\varepsilon > 0$. Since (2.12) holds for any $\delta \in \mathbb{R}$, we can derive the following two statements.

$$\exists n_0^- \in \mathbb{N} : \forall n \geq n_0^-, \quad \forall t \in \mathbb{R} : \quad F(t + \delta) \leq F(t + \delta_n^-) + \varepsilon \tag{2.13}$$
$$\exists n_0^+ \in \mathbb{N} : \forall n \geq n_0^+, \quad \forall t \in \mathbb{R} : \quad F(t - \delta) \geq F(t - \delta_n^-) - \varepsilon. \tag{2.14}$$

Take now $N = \max\{n_0^-, n_0^+\}$. Combining all results leads to the following:

$$\begin{aligned}\forall n \geq N, \quad \forall t \in \mathbb{R} : F(t - \delta_n^-) - (\mathcal{L}_\delta + \varepsilon) &\overset{(2.14)}{\leq} F(t - \delta) - \mathcal{L}_\delta \overset{(2.2)}{\leq} G(t) \\ &\overset{(2.2)}{\leq} F(t + \delta) + \mathcal{L}_\delta \overset{(2.13)}{\leq} F(t + \delta_n) + (\mathcal{L}_\delta + \varepsilon).\end{aligned}$$

Hence, $\forall n \geq N : \mathcal{L}_\delta + \varepsilon \in \tau_{\delta_n^-}$. Together with the monotonicity of the map (Lemma 2.26), we obtain

$$\forall n \geq N : |\mathcal{L}_{\delta_n^-} - \mathcal{L}_\delta| \leq \varepsilon.$$

Since ε was arbitrarily chosen, the result follows. □

Remark 2.31. In Lemma 2.30 we proved that if we have a continuous cumulative distribution function F, then (2.12) follows. Note further that the other direction holds trivially: if a function satisfies (2.12), then it is continuous. In other words, continuity of a cumulative distribution function F is equivalent to condition (2.12).

Remark 2.32. The assumption that at least one of the underlying cumulative distribution functions is continuous is only necessary but not sufficient. We can find examples where two cumulative distribution functions F and G with different points of discontinuity may still produce a continuous map $\delta \mapsto \mathcal{L}_\delta(F, G)$.

Example 2.33. *Let* $F(t) = \mathbb{1}_{[\frac{3}{4}, +\infty)}(t)$ *and* $G(t) = \begin{cases} 0, & t < 0 \\ t, & 0 \le t < \frac{4}{5} \\ 1, & t \ge \frac{4}{5} \end{cases}$.

We will consider three cases, as follows:

- *Case 1:* $\delta \in \left[0, \frac{1}{20}\right)$. *As in Example 2.29, we will summarize the values in a table.*

Table 2.3: Values to compare (Case 1)

t	$-\infty$	0	$\frac{3}{4}-\delta$	$\frac{3}{4}$	$\frac{3}{4}+\delta$	$\frac{4}{5}$ $+\infty$
$F(t-\delta)-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$1-\varepsilon$	$1-\varepsilon$
$G(t)$	0	t	t	t	t	1
$F(t+\delta)+\varepsilon$	ε	ε	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$

Recall that the Levy gauge is the smallest $\varepsilon \ge 0$, such that all inequalities, derived from the columns of Table 2.3, are satisfied. This can be reduced to the following cases (in yellow):

$\boxed{*}$ $t \le \varepsilon, \forall t \in \left[0, \frac{3}{4} - \delta\right)$*, giving the candidate* $\varepsilon_1 = \frac{3}{4} - \delta$.

$\boxed{*}$ $1 - \varepsilon \le t, \forall t \in \left[\frac{3}{4} + \delta, \frac{4}{5}\right)$. *Setting* $t^* = 1 - t$ *gives now* $\varepsilon \ge t^*, \forall t^* \in \left(\frac{1}{5}, \frac{1}{4} - \delta\right]$, *giving the candidate* $\varepsilon_2 = \frac{1}{4} - \delta$.

Since $\varepsilon_1 > \varepsilon_2$ *and both yellow cases must be satisfied, we obtain*

$$\mathcal{L}_\delta(F, G) = \varepsilon_1 = \frac{3}{4} - \delta.$$

- *Case 2:* $\delta \in \left[\frac{1}{20}, \frac{3}{4}\right)$.

Table 2.4: Values to compare (Case 2)

t	$-\infty$	0	$\frac{3}{4}-\delta$	$\frac{3}{4}$	$\frac{4}{5}$	$\frac{3}{4}+\delta$ $+\infty$
$F(t-\delta)-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$1-\varepsilon$
$G(t)$	0	t	t	t	1	1
$F(t+\delta)+\varepsilon$	ε	ε	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$

In a similar fashion, note that all inequalities in Table 2.4 always hold, up to the yellow case $t \le \varepsilon, \forall t \in [0, \frac{3}{4} - \delta)$. *In other words, in this case we obtain*

$$\mathcal{L}_\delta(F, G) = \frac{3}{4} - \delta.$$

- <u>*Case 3*</u>*:* $\delta \in \left[\frac{3}{4}, +\infty\right)$.

Table 2.5: Values to compare (Case 3)

t	$-\infty$	$\frac{3}{4}-\delta$	0	$\frac{3}{4}$	$\frac{4}{5}$	$\frac{3}{4}+\delta$	$+\infty$
$F(t-\delta)-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$-\varepsilon$	$1-\varepsilon$	
$G(t)$	0	0	t	t	1	1	
$F(t+\delta)+\varepsilon$	ε	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$	$1+\varepsilon$	

Since all inequalities in Table 2.5 are satisfied for any choice of ε*, it follows that*

$$\mathcal{L}_\delta(F,G) = 0.$$

We can now summarize the map $\delta \mapsto \mathcal{L}_\delta(F,G)$*, as follows:*

$$\mathcal{L}_\delta(F,G) = \begin{cases} \frac{3}{4}-\delta, & 0 \le \delta < \frac{3}{4} \\ 0, & \delta \ge \frac{3}{4} \end{cases}.$$

The graphical representation is provided in Figure 2.2. Moreover, the fixed point equation $\delta = \mathcal{L}_\delta(F,G)$ *has a unique solution, so we obtain that* $\mathcal{L}(F,G) = \frac{3}{8}$.

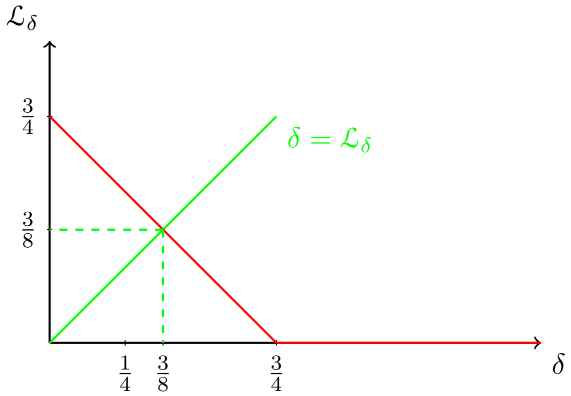

Figure 2.2: Visualisation of the map $\delta \mapsto \mathcal{L}_\delta$ (in red)

3 The Levy Gauge in the Setting of Prediction Intervals

This chapter considers the relation of the Levy gauge to prediction intervals. In doing so, it is essential to introduce some definitions and specific notations.

Definition 3.1 (Quantile)**.** *Consider a cumulative distribution function F and some $\alpha \in \mathbb{R}$. Then the quantile of F is defined as*

$$Q_\alpha^F := \inf\{t \in \mathbb{R} : F(t) \geq \alpha\}.$$

The quantile Q_α^F takes values in the extended real line $\overline{\mathbb{R}}$.

Remark 3.2. For convenience, we will use the conventions $F((-\infty)-) = F(-\infty) = 0$ and $F(+\infty) = 1$. In addition, we will always take $\inf \emptyset = +\infty$.

In addition, we need some results that clarify the connection between α and the corresponding quantile.

Lemma 3.3. *Consider the setting of Definition 3.1. Then the following claims hold true:*

- *If $\alpha < 0$, then $Q_\alpha^F = -\infty$.*
- *If $\alpha > 1$, then $Q_\alpha^F = +\infty$.*
- *If $\alpha \in [0,1]$, then $F(Q_\alpha^F -) \leq \alpha \leq F(Q_\alpha^F)$.*

Proof. We will consider the three cases separately.

- $\underline{\alpha < 0}$: Since $F(t) \geq \alpha, \forall t \in \mathbb{R}$, the claim follows directly from the definition.
- $\underline{\alpha > 1}$: Since $F(t) < \alpha, \forall t \in \mathbb{R}$, it holds $Q_\alpha^F = \inf \emptyset = +\infty$.
- $\underline{\alpha \in [0,1]}$: Note that $\alpha = 0$ is the only possibility to obtain $Q_\alpha^F = -\infty$. In view of Remark 3.2 we have $F((-\infty)-) = 0 = \alpha = F(-\infty)$. In addition, $Q_\alpha^F = +\infty$ can be achieved if and only if $\alpha = 1$ and the claim follows directly.
  Now assume that $Q_\alpha^F \in \mathbb{R}$ and consider some sequence $(t_n^+)_{n \in \mathbb{N}}$, such that $t_n^+ \downarrow Q_\alpha^F$. That is,
  $$Q_\alpha^F < t_n^+, \forall n \in \mathbb{N} \quad \text{and} \quad F(t_n^+) \geq \alpha, \forall n \in \mathbb{N}.$$
  The right continuity of F now yields
  $$\alpha \leq \lim_{n \to \infty} F(t_n^+) = F(Q_\alpha^F).$$

Similarly, take some sequence $(t_n^-)_{n\in\mathbb{N}}$ that converges to Q_α^F from below, i.e., $t_n^- \uparrow Q_\alpha^F$. Therefore,

$$F(t_n^-) < \alpha, \forall n \in \mathbb{N} \quad \text{and} \quad F(Q_\alpha^F-) = \lim_{n\to\infty} F(t_n^-) \leq \alpha.$$

Combining both results gives the statement.

□

Corollary 3.4. *If $\alpha_1 \geq 0$ and $\alpha_2 \leq 1$, we have $F(Q_{\alpha_2}^F) - F(Q_{\alpha_1}^F-) \geq \alpha_2 - \alpha_1$. Similarly, for $\alpha_1 \leq 1$ and $\alpha_2 \geq 0$, we have $F(Q_{\alpha_2}^F-) - F(Q_{\alpha_1}^F) \leq \alpha_2 - \alpha_1$.*

Proof. Consider the case $\alpha_1 \geq 0, \alpha_2 \leq 1$. We have to apply Lemma 3.3, treating α_1 and α_2 separately:

- If $\alpha_1 \in [0,1]$, we have $F(Q_{\alpha_1}^F-) \leq \alpha_1 \leq F(Q_{\alpha_1}^F)$, therefore,

  $$-F(Q_{\alpha_1}^F-) \geq -\alpha_1. \tag{3.1}$$

  For $\alpha_1 > 1$ it is easy to see that

  $$F(Q_{\alpha_1}^F-) \leq F(Q_{\alpha_1}^F) = F(+\infty) = 1 < \alpha_1$$

  and (3.1) follows immediately.

- If $\alpha_2 \in [0,1]$, we obtain

  $$F(Q_{\alpha_2}^F) \geq \alpha_2. \tag{3.2}$$

  For $\alpha_2 < 0$ we have $Q_{\alpha_2}^F = -\infty$ and (3.2) follows.

Adding (3.1) and (3.2) to one another gives the statement.
To obtain the second statement of the corollary, apply the first result with exchanged α_1 and α_2, that is,

$$F(Q_{\alpha_1}^F) - F(Q_{\alpha_2}^F-) \geq \alpha_1 - \alpha_2.$$

Multiplying both sides by -1 finishes the proof. □

Until now we discussed the set $\tau_\delta(F,G)$ from the perspective of the underlying cumulative distribution functions. Moreover, it was further noticed that $\tau_\delta(F,G) = [\mathcal{L}_\delta(F,G), +\infty)$. In this sense, it seems natural to find an alternative representation which is based on the quantiles.

Lemma 3.5. *It holds that*

$$\tau_\delta(F,G) = \{\varepsilon \geq 0 : \forall \alpha \in \mathbb{R} : Q_{\alpha-\varepsilon}^F - \delta \leq Q_\alpha^G \leq Q_{\alpha+\varepsilon}^F + \delta\}.$$

Proof. Let $E = \{\varepsilon \geq 0 : \forall\alpha \in \mathbb{R} : Q^F_{\alpha-\varepsilon} - \delta \leq Q^G_\alpha \leq Q^F_{\alpha+\varepsilon} + \delta\}$.
Let $\varepsilon \geq 0$. Assume first that $\varepsilon \in \tau_\delta(F, G)$. Note that for any $t \in \mathbb{R}$ with $G(t) \geq \alpha, \alpha \in \mathbb{R}$ we get

$$F(t + \delta) \geq G(t) - \varepsilon \geq \alpha - \varepsilon,$$

in view of Remark 2.6. We can rewrite this as events:

$$\{t \in \mathbb{R} : G(t) \geq \alpha\} \subseteq \{t \in \mathbb{R} : F(t + \delta) \geq \alpha - \varepsilon\}.$$

Taking infimum on both sides and Definition 3.1 yield

$$Q^G_\alpha \geq \inf\{t \in \mathbb{R} : F(t + \delta) \geq \alpha - \varepsilon\} = \inf\{t \in \mathbb{R} : F(t) \geq \alpha - \varepsilon\} - \delta = Q^F_{\alpha-\varepsilon} - \delta.$$

Swapping F and G gives now

$$\forall\alpha \in \mathbb{R} : Q^F_\alpha \geq Q^G_{\alpha-\varepsilon} - \delta.$$

Obviously, this is equivalent to

$$\forall\alpha \in \mathbb{R} : Q^G_\alpha \leq Q^F_{\alpha+\varepsilon} + \delta,$$

i.e., $\varepsilon \in E$.
Assume now that $\varepsilon \notin \tau_\delta(F, G)$. This may hold true when at least one of the conditions in $\tau_\delta(F, G)$ is violated, that is,

$$\exists t^* \in \mathbb{R} : G(t^*) > F(t^* + \delta) + \varepsilon \quad \text{or} \quad F(t^*) > G(t^* + \delta) + \varepsilon. \tag{3.3}$$

We will prove only the first case, since the second one is completely analogous.
Set $\alpha^* = G(t^*)$, which implies that $Q^G_{\alpha^*} \leq t^*$. In addition, (3.3) gives $F(t^* + \delta) < \alpha^* - \varepsilon$, meaning that $t^* + \delta < Q^F_{\alpha^*-\varepsilon}$. Hence, there exists $\alpha^* \in \mathbb{R}$, such that $Q^G_{\alpha^*} \leq t^* < Q^F_{\alpha^*-\varepsilon} - \delta$, therefore, $\varepsilon \notin E$.

□

Corollary 3.6. *Let $\alpha_1, \alpha_2 \in \mathbb{R}$. Fix $\langle$ to be either [or (, and $\rangle$ to be either] or). For $\varepsilon \in \tau_\delta(F, G)$, we then have*

$$\langle Q^F_{\alpha_1+\varepsilon} + \delta, Q^F_{\alpha_2-\varepsilon} - \delta\rangle \subseteq \langle Q^G_{\alpha_1}, Q^G_{\alpha_2}\rangle \subseteq \langle Q^F_{\alpha_1-\varepsilon} - \delta, Q^F_{\alpha_2+\varepsilon} + \delta\rangle. \tag{3.4}$$

Proof. Since $\varepsilon \in \tau_\delta$, we can apply Lemma 3.5 to the upper end points and conclude

$$Q^F_{\alpha_2-\varepsilon} - \delta \leq Q^G_{\alpha_2} \leq Q^F_{\alpha_2+\varepsilon} + \delta.$$

Similarly, for the lower end points, we obtain

$$Q^F_{\alpha_1+\varepsilon} + \delta \geq Q^G_{\alpha_1} \geq Q^F_{\alpha_1-\varepsilon} - \delta.$$

Combining both results gives the claim. □

Remark 3.7. As stated in Amann, Leeb and Steinberger 2025, the result from Corollary 3.6 "allows us to sandwich the coverage probability of a prediction interval based on a distribution function G by the coverage probabilities of prediction intervals, based on another distribution function F". This turns out to be useful for the results below.

Before we proceed to the main result in this chapter, we need additional notation for certain prediction intervals. Let $\alpha_1, \alpha_2, \delta \in \mathbb{R}$ and F be some cumulative distribution function. Then we will denote by

$$_{\alpha_1}I^F_{\alpha_2}(\delta) := [Q^F_{\alpha_1} - \delta, Q^F_{\alpha_2} + \delta], \quad {}_{\alpha_1}\mathring{I}^F_{\alpha_2}(\delta) := (Q^F_{\alpha_1} - \delta, Q^F_{\alpha_2} + \delta)$$

the corresponding distorted prediciton intervals.
If $\delta > 0$, we are talking about δ-inflated (or just inflated) prediction intervals. If $\delta < 0$, then the corresponding prediction interval is said to be δ-shrunken. If $\delta = 0$, we will simply write $_{\alpha_1}I^F_{\alpha_2}$ and $_{\alpha_1}\mathring{I}^F_{\alpha_2}$, respectively. Furthermore, if $Q^F_{\alpha_1} - \delta > Q^F_{\alpha_2} + \delta$, we will set the interval to be the empty set $\emptyset$.

Lemma 3.8. *Let F and G be cumulative distribution functions, and let U be a random variable. In addition, let $\varepsilon \geq 0, \delta \geq 0$. If $\mathcal{L}_\delta(F,G) \leq \varepsilon$ (or, equivalently, $\varepsilon \in \tau_\delta(F,G)$), then*

$$\inf_{\alpha_1 \leq \alpha_2} \left\{ \{\mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(\delta)\right) - \mathbb{P}\left(U \in {}_{\alpha_1+\varepsilon}I^F_{\alpha_2-\varepsilon}\right) \right\} \geq 0 \quad \textit{and} \tag{3.5}$$

$$\sup_{\alpha_1 \leq \alpha_2} \left\{ \mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(-\delta)\right) - \mathbb{P}\left(U \in {}_{\alpha_1-\varepsilon}I^F_{\alpha_1+\varepsilon}\right) \right\} \leq 0. \tag{3.6}$$

Proof. Let $\varepsilon \in \tau_\delta(F,G)$. Replace F, G, α_1 and α_2 in Corollary 3.6 with $G, F, \alpha_1 + \varepsilon$ and $\alpha_2 - \varepsilon$, respectively. Then the second part of (3.4) entails $_{\alpha_1+\varepsilon}I^F_{\alpha_2-\varepsilon} \subseteq {}_{\alpha_1}I^G_{\alpha_2}(\delta)$, therefore,

$$\mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(\delta)\right) - \mathbb{P}\left(U \in {}_{\alpha_1+\varepsilon}I^F_{\alpha_2-\varepsilon}\right) \geq 0,$$

for any $\alpha_1 \leq \alpha_2$, implying (3.5).
In a similar fashion, use Corollary 3.6 with $G, F, \alpha_1 - \varepsilon$ and $\alpha_2 + \varepsilon$, respectively, and obtain by the first part of (3.4) that $_{\alpha_1}I^G_{\alpha_2}(-\delta) \subseteq {}_{\alpha_1-\varepsilon}I^F_{\alpha_2+\varepsilon}$. This entails

$$\mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(-\delta)\right) - \mathbb{P}\left(U \in {}_{\alpha_1-\varepsilon}I^F_{\alpha_2+\varepsilon}\right) \leq 0,$$

for any $\alpha_1 \leq \alpha_2$, implying (3.6). Note that replacing $[$ with $\langle$, and $]$ with $\rangle$ gives the same results in view of Corollary 3.6.

□

We can now proceed to the main result in this chapter.

Proposition 3.9. *Let U be a random variable, F and G be cumulative distribution functions, with $F(t) = \mathbb{P}(U \leq t)$. Then it holds that*

$$\inf_{0 \leq \alpha_1 \leq \alpha_2 \leq 1} \left\{ \mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(\delta)\right) - (\alpha_2 - \alpha_1) \right\} \geq -2\mathcal{L}_\delta(F,G) \quad \textit{and} \tag{3.7}$$

$$\sup_{0 \leq \alpha_1 \leq \alpha_2 \leq 1} \left\{ \mathbb{P}\left(U \in {}_{\alpha_1}\mathring{I}^G_{\alpha_2}(-\delta)\right) - (\alpha_2 - \alpha_1) \right\} \leq 2\mathcal{L}_\delta(F,G). \tag{3.8}$$

Proof. Since $\alpha_1 + \mathcal{L}_\delta \geq 0$ and $\alpha_2 - \mathcal{L}_\delta \leq 1$, for any $0 \leq \alpha_1 \leq \alpha_2 \leq 1$, Corollary 3.4 yields

$$\mathbb{P}\left(U \in {}_{\alpha_1+\mathcal{L}_\delta}I^F_{\alpha_2-\mathcal{L}_\delta}\right) = F(Q^F_{\alpha_2-\mathcal{L}_\delta}) - F(Q^F_{\alpha_1+\mathcal{L}_\delta}-) \geq (\alpha_2-\mathcal{L}_\delta)-(\alpha_1+\mathcal{L}_\delta) = \alpha_2-\alpha_1-2\mathcal{L}_\delta.$$

Hence,

$$\mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(\delta)\right) - (\alpha_2 - \alpha_1 - 2\mathcal{L}_\delta) \geq \mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(\delta)\right) - \mathbb{P}\left(U \in {}_{\alpha_1+\mathcal{L}_\delta}I^F_{\alpha_2-\mathcal{L}_\delta}\right) \geq 0,$$

due to (3.5). Since this holds for any $0 \leq \alpha_1 \leq \alpha_2 \leq 1$, we obtain (3.7).
Similarly, recall that $\alpha_2 + \mathcal{L}_\delta \geq 0$ and $\alpha_1 - \mathcal{L}_\delta \leq 1$ and apply Corollary 3.4 to conclude

$$\mathbb{P}\left(U \in {}_{\alpha_1-\mathcal{L}_\delta}\mathring{I}^F_{\alpha_2+\mathcal{L}_\delta}\right) = F(Q^F_{\alpha_2+\mathcal{L}_\delta}-) - F(Q^F_{\alpha_1-\mathcal{L}_\delta}) \leq \alpha_2 - \alpha_1 + 2\mathcal{L}_\delta.$$

It remains to apply (3.6) to finish the proof. □

Remark 3.10. Based on Proposition 3.9, it is easy to see that for any $\varepsilon > \mathcal{L}_\delta$ it holds

$$\inf_{0\leq\alpha_1\leq\alpha_2\leq1}\left\{\mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(\delta)\right) - (\alpha_2-\alpha_1)\right\} \overset{(3.7)}{\geq} -2\mathcal{L}_\delta(F,G) > -2\varepsilon \quad \text{and}$$
$$\sup_{0\leq\alpha_1\leq\alpha_2\leq1}\left\{\mathbb{P}\left(U \in {}_{\alpha_1}\mathring{I}^G_{\alpha_2}(-\delta)\right) - (\alpha_2-\alpha_1)\right\} \overset{(3.8)}{\leq} 2\mathcal{L}_\delta(F,G) < 2\varepsilon.$$

Moreover, if we take some $\delta^* > \delta$, then $\mathbb{P}\left(U \in {}_{\alpha_1}\mathring{I}^G_{\alpha_2}(-\delta)\right) \geq \mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(-\delta^*)\right)$, which yields

$$\sup_{0\leq\alpha_1\leq\alpha_2\leq1}\left\{\mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(-\delta^*)\right) - (\alpha_2-\alpha_1)\right\} < 2\varepsilon.$$

The last statement from Remark 3.10 provides a result for the supremum case, which holds true for closed shrunken prediction intervals. The "price" for that is the tiny enlargement of δ to some δ^*, which is strictly positive. Since in the following chapter we will need the results in the other direction, we will take the negations of all results from Remark 3.10 and formulate this more precisely in the following.

Corollary 3.11. *Let $\delta \geq 0, \varepsilon > 0, F$ and G be cumulative distribution functions and U be a random variable such that $F(t) = \mathbb{P}(U \leq t)$. Then the following claims hold true:*

- *If* $\inf_{0\leq\alpha_1\leq\alpha_2\leq1}\left\{\mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(\delta)\right) - (\alpha_2-\alpha_1)\right\} \leq -2\varepsilon$*, then* $\mathcal{L}_\delta(F,G) \geq \varepsilon$.
- *If* $\sup_{0\leq\alpha_1\leq\alpha_2\leq1}\left\{\mathbb{P}\left(U \in {}_{\alpha_1}\mathring{I}^G_{\alpha_2}(-\delta)\right) - (\alpha_2-\alpha_1)\right\} \geq 2\varepsilon$*, then* $\mathcal{L}_\delta(F,G) \geq \varepsilon$.
- *Let* $0 \leq \delta < \delta^*$*. If* $\sup_{0\leq\alpha_1\leq\alpha_2\leq1}\left\{\mathbb{P}\left(U \in {}_{\alpha_1}I^G_{\alpha_2}(-\delta^*)\right) - (\alpha_2-\alpha_1)\right\} \geq 2\varepsilon$*, then* $\mathcal{L}_\delta(F,G) \geq \varepsilon$.

4 Application of the Levy Gauge in the Jackknife and Jackknife+ Methods

The Levy gauge was introduced as an object of independent interest in the paper of Amann, Leeb and Steinberger 2025. Its relation to distorted prediction intervals, as seen in the previous chapter, makes it a useful tool if one aims to derive asymptotic results for methods of uncertainty quantification, such as the Jackknife and the Jackknife+. The current chapter mostly follows the ideas, concepts and notation covered in Amann, Leeb and Steinberger 2025, but explores more precisely the exact usage of the Levy gauge. It turns out that under mild assumptions, the Levy gauge converges in probability, which can be applied in some of the proofs formulated in the original paper.

Before we proceed with the above-mentioned methods of uncertainty quantification, we need to specify the details regarding the data. In this chapter, we will mostly follow the notation given in Amann, Leeb and Steinberger 2025. Thus, consider the training data $T_n = (y_i, x_i)_{i=1}^n$, consisting of n independent and identically distributed (i.i.d.) response-feature pairs (y_i, x_i), with $y_i \in \mathbb{R}, x_i \in \mathbb{R}^p, \forall i \in [n]$, each of them being distributed according to some Borel measure $\mathcal{P}_n$ on the space $\mathbb{R} \times \mathbb{R}^p$. The dimension of the feature vector may grow with the sample size, i.e., $p = p(n)$. We will use $T_n^{\backslash i}$ as a shorthand for the training data, consisting of all response-feature pairs without the i-th one, that is, $T_n^{\backslash i} = (y_i, x_i)_j$, for $j \in [n]\backslash i$.

In addition, let $(y_{n+1}, x_{n+1}) \sim \mathcal{P}_n$ be a new response-feature pair, which is *independent* of the training data T_n. We will denote by $\hat{y}_{n+1}$ a predictor for the new response y_{n+1}, given the training data T_n and the new feature vector x_{n+1}. In other words, $\hat{y}_{n+1} := \mathcal{A}_{p,n}(x_{n+1}, T_n)$, where $\mathcal{A}_{p,n}$ is the Borel-measurable function describing the underlying prediction algorithm. Similarly, $\hat{y}_{n+1}^{\backslash i}$ would be the predictor, given $T_n^{\backslash i}$ and x_{n+1}.

4.1 Jackknife method

The Jackknife method can be considered as a special case of the cross validation (CV), where the number of folds equals the number of observations n. It was firstly introduced in 1950s (Quenouille 1956, Tukey 1958) as "a technique for reducing the bias of a serial correlation estimator based on splitting the sample", as stated in Miller 1974. A short review of the approach is given by Miller 1974, which investigates the role of the Jackknife in bias reduction and in robust interval estimation.

In this thesis, we are primarily interested in the application of the Jackknife method to the asymptotics of prediction intervals. A recent result in this direction is Steinberger and

Leeb 2023 which proves the asymptotic conditional validity of the prediction intervals, that is, their actual coverage probability given the observations in the training sample converges to the corresponding nominal level. This includes the high-dimensional cases, under certain assumptions. The paper of Amann, Leeb and Steinberger 2025 extends these results and focuses on the asymptotic conditional conservativeness of the prediction intervals based on the Jackknife (and more generally, the CV), under mild assumptions. This means that "the probability of the actual coverage probability, conditional on the training data, undershooting its nominal level vanishes asymptotically."

Taking this into consideration, we need to introduce some notation. Denote by $F_n(t)$ the distribution function of the prediction error given the training data T_n, i.e,

$$F_n(t) = \mathbb{P}(y_{n+1} - \hat{y}_{n+1} \leq t || T_n).$$

It is clear that we need some estimator for F_n. In the case of the Jackknife method, this is the empirical distribution function of the leave-one-out residuals $r_i = y_i - \hat{y}_i^{\setminus i}$, i.e.,

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{r_i \leq t\}.$$

Having defined $\hat{F}_n$ and in view of Lemma 3.3, we can rewrite the definition of the quantile of $\hat{F}_n$, as follows:

$$Q_\alpha^{\hat{F}_n} = \inf\{t \in \mathbb{R} : \hat{F}_n(t) \geq \alpha\} = \begin{cases} -\infty, & \alpha \leq 0 \\ r_{(\lceil \alpha n \rceil)}, & \alpha \in (0,1] \\ +\infty, & \alpha > 1, \end{cases} \tag{4.1}$$

where $r_{(i)}$ denotes the i-th ordered leave-one-out residual. Therefore, the quantile can obtain countably many values, which will be important for the results which are to be introduced in this chapter.

As already stated, we are interested in a prediction interval based on the training data T_n and the new feature vector x_{n+1}. Denote by

$$PI^J_{\alpha_1,\alpha_2}(\delta) := \hat{y}_{n+1} + [Q^{\hat{F}_n}_{\alpha_1} - \delta, Q^{\hat{F}_n}_{\alpha_2} + \delta] = \hat{y}_{n+1} + {}_{\alpha_1}I^{\hat{F}_n}_{\alpha_2}(\delta) \tag{4.2}$$

the δ-distorted prediction interval for the unknown new response y_{n+1}, based on the predictor $\hat{y}_{n+1}$ and the Jackknife method. This interval has a nominal coverage probability of $\alpha_2 - \alpha_1$. Note further that the second representation above refers to the well-known notation from Chapter 3. We will switch between them depending on the context. For the sake of readability, we will write $PI^J_{\alpha_1,\alpha_2}$ instead of $PI^J_{\alpha_1,\alpha_2}(0)$.

To proceed further, we need two more assumptions, which we will use for the rest of the current thesis.

Definition 4.1 (Asymptotic out-of-sample stability). *We will say that the predictor $\hat{y}_{n+1}$ is asymptotic-out-of sample stable (or just stable, if not stated otherwise), if*

$$\forall \varepsilon > 0 : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}(|\hat{y}_{n+1} - \hat{y}^{\setminus i}_{n+1}| \geq \varepsilon) = 0.$$

This is Definition 3.1 from Amann, Leeb and Steinberger 2025.

Definition 4.2 (CC-Assumption)**.** *The Continuous case assumption consists of two components.*

- *Let $\mathcal{P}_n^x$ be the marginal distribution of x_1. Then for all $n \in \mathbb{N}$ and $\mathcal{P}_n^x$-almost every x, it holds that the conditional distribution of $y_1|x_1 = x$ is absolutely continuous. Furthermore, the supremum norm of the density is finite, i.e., $||f_{y_1|x_1=x}||_\infty < \infty$.*

- *The sequence of random variables $(||f_{y_1|x_1}||_\infty)_{n\in\mathbb{N}}$ is stochastically bounded.*

This is Definition 2.1 from Amann, Leeb and Steinberger 2025.
We will now state a result regarding the convergence in probability of the Levy gauge.

Lemma 4.3 (Convergence in probability)**.** *Suppose that the prediction error $y_{n+1} - \hat{y}_{n+1}$ is stochastically bounded and the predictor $\hat{y}_{n+1}$ is stable in the sense of Definition 4.1. Then the Levy gauge between F_n and $\hat{F}_n$ converges to 0 in probability, i.e.,*

$$\mathcal{L}_\delta(\hat{F}_n, F_n) \xrightarrow{p} 0, \quad \forall \delta > 0.$$

Proof. Let $\varepsilon > 0, \eta_{n,\varepsilon} := \frac{1}{n}\sum_{i=1}^n \mathbb{P}(|\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}| > \varepsilon) \overset{n\to\infty}{\longrightarrow} 0$ in view of the asymptotic stability of the predictor. In addition, let $(L_n)_{n\in\mathbb{N}}$ be a sequence, satisfying

$$\lim_{n\to\infty} L_n = +\infty \quad \text{and} \quad L_n = o_p\left(\left(\eta_{n,\varepsilon} + \frac{1}{n}\right)^{-1}\right).$$

Lemma D.4 from Amann, Leeb and Steinberger 2025 gives the following upper bound for $\mu = 0, L = L_n$ and $k_n = n$:

$$\begin{aligned}\mathbb{P}(\mathcal{L}_\delta(\hat{F}_n, F_n) \geq \varepsilon) \leq & \frac{2\mathbb{P}(|y_{n+1} - \hat{y}_{n+1}| \geq L_n)}{\varepsilon} + \frac{8L_n + 12\delta}{n\delta\varepsilon^2} \\ & + \frac{4(5n+1)}{n^2\delta\varepsilon^2}\sum_{i=1}^n \mathbb{E}[\min\{2L_n + 3\delta, |\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}|\}]. \end{aligned} \tag{4.3}$$

The first two terms in (4.3) vanish asymptotically due to the stochastic boundedness of the prediction error. It remains to consider the third term, containing the expectation. By setting $Z_i := \min\{2L_n + 3\delta, |\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}|\}$ we get

$$\begin{aligned}\mathbb{E}[Z_i] &= \mathbb{E}[Z_i \mathbb{1}\{|\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}| > \varepsilon\} + Z_i \mathbb{1}\{|\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}| \leq \varepsilon\}] \\ &\leq \mathbb{E}\left[(2L_n + 3\delta)\mathbb{1}\{|\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}| > \varepsilon\} + \underbrace{|\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}|\mathbb{1}\{|\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}| \leq \varepsilon\}}_{\leq \varepsilon}\right] \\ &\leq (2L_n + 3\delta)\mathbb{P}(|\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}| > \varepsilon) + \varepsilon. \end{aligned} \tag{4.4}$$

Take now $\limsup_{n\to\infty}$ of the third term in (4.3) to conclude

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{4(5n+1)}{n\delta\varepsilon^2}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Z_i] \\
\overset{(4.4)}{\leq}\ &\limsup_{n\to\infty} \frac{4(5n+1)}{n\delta\varepsilon^2}\left[(2L_n+3\delta)\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}(|\hat{y}_{n+1}-\hat{y}_{n+1}^{\setminus i}|>\varepsilon)+\varepsilon\right] \\
\leq\ & const\cdot\left(\limsup_{n\to\infty}(2L_n+3\delta)\eta_{n,\varepsilon}+\varepsilon\right) = const\cdot\varepsilon,
\end{aligned}$$

due to the way we defined L_n and $\eta_{n,\varepsilon}$. Since $\varepsilon>0$ has been chosen arbitrarily, the claim follows. □

Theorem 4.4 (Asymptotically conservative/valid prediction intervals with the Jackknife)**.** *Suppose that the prediction error $y_{n+1}-\hat{y}_{n+1}$ is stochastically bounded and the predictor $\hat{y}_{n+1}$ is stable. Then the following statements hold true.*

1. ***General case:*** *For every $\delta>0$ the δ-inflated Jackknife prediction intervals are asymptotically uniformly conditionally conservative, that is,*

$$\lim_{n\to\infty}\mathbb{P}\left[\inf_{0\leq\alpha_1\leq\alpha_2\leq 1}\left\{\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}(\delta)||T_n)-(\alpha_2-\alpha_1)\right\}\geq-\varepsilon\right]=1,\forall\varepsilon>0.$$

Additionally, the conditional coverage probability of δ-shrunken prediction intervals uniformly does not overshoot its nominal level asymptotically, that is,

$$\lim_{n\to\infty}\mathbb{P}\left[\sup_{0\leq\alpha_1\leq\alpha_2\leq 1}\left\{\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}(-\delta)||T_n)-(\alpha_2-\alpha_1)\right\}\leq\varepsilon\right]=1,\forall\varepsilon>0.$$

2. ***Continuous case:*** *Assume further that the CC-Assumption (see Definition 4.2) is satisfied. Then the non-distorted prediction interval (in other words, the case $\delta=0$) is asymptotically uniformly conditionally valid, that is,*

$$\lim_{n\to\infty}\mathbb{E}\left[\sup_{0\leq\alpha_1\leq\alpha_2\leq 1}\left|\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}||T_n)-(\alpha_2-\alpha_1)\right|\right]=0.$$

This is Theorem 3.2 in Amann, Leeb and Steinberger 2025. The proof of the general case is based on the results that have already been presented in this thesis. The proof of the continuous case follows the paper of Amann, Leeb and Steinberger 2025 but together with some necessary additional clarifications.

Proof. Consider first the general case. Use Corollary 3.11 and take

$$U=y_{n+1}-\hat{y}_{n+1}, F=F_n, G=\hat{F}_n.$$

Note further that the infimum and the supremum are measurable, since they can be restricted to infimum and supremum over a finite set, respectively. To see this, recall from

(4.1) that the quantiles of $\hat{F}_n$ can take countably many values. The Levy gauge is also measurable for any cumulative distribution function due to Remark 2.10. This leads to

$$\mathbb{P}\left[\inf_{0\le\alpha_1\le\alpha_2\le 1}\left\{\mathbb{P}(y_{n+1}-\hat{y}_{n+1}\in {}_{\alpha_1}I_{\alpha_2}^{\hat{F}_n}(\delta)||T_n)-(\alpha_2-\alpha_1)\right\}\le -\varepsilon\right]\le \mathbb{P}\left(\mathcal{L}_\delta(\hat{F}_n,F_n)\ge\frac{\varepsilon}{2}\right),$$

which tends to 0 as $n\to\infty$ due to Lemma 4.3. The first claim from the general case follows directly.
We can proceed in a similar way towards the supremum claim in the general case. Since $\delta>0$, we can always find some positive $\delta'<\delta$. Corollary 3.11 and Lemma 4.3 now yield

$$\mathbb{P}\left[\sup_{0\le\alpha_1\le\alpha_2\le 1}\left\{\mathbb{P}(y_{n+1}-\hat{y}_{n+1}\in {}_{\alpha_1}I_{\alpha_2}^{\hat{F}_n}(-\delta)||T_n)-(\alpha_2-\alpha_1)\right\}\ge \varepsilon\right]\le \mathbb{P}\left(\mathcal{L}_{\delta'}(\hat{F}_n,F_n)\ge\frac{\varepsilon}{2}\right),$$

which tends to 0 for $n\to\infty$ and the claim follows.
Now we proceed with the continuous case. Let $\varepsilon>0$ be arbitrary. We will first prove that there exists some fixed $\delta_\varepsilon=\delta(\varepsilon)$, such that it holds almost surely that

$$|\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}(\delta_\varepsilon)||T_n)-\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}(-\delta_\varepsilon)||T_n)|\le\varepsilon. \tag{4.5}$$

Recall that the predictor $\hat{y}_{n+1}$ is based on the training data T_n and the new feature vector x_{n+1}. In view of the CC-Assumption (Definition 4.2) and for an arbitrary $\alpha\in[0,1],\delta>0$, it holds almost surely

$$\begin{aligned}
&\mathbb{P}(y_{n+1}\in\hat{y}_{n+1}+[Q_\alpha^{\hat{F}_n}-\delta,Q_\alpha^{\hat{F}_n}+\delta]||T_n)\\
&=\mathbb{E}[\mathbb{P}(y_{n+1}\in\underbrace{\hat{y}_{n+1}+[Q_\alpha^{\hat{F}_n}-\delta,Q_\alpha^{\hat{F}_n}+\delta]}_{\text{interval of length }2\delta}||T_n,x_{n+1})]\\
&\le\mathbb{E}[\min\{1,2\delta||f_{y_{n+1}|x_{n+1}}||_\infty\}].
\end{aligned}$$

Now it is easy to see that

$$\begin{aligned}
&|\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}(\delta)||T_n)-\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}(-\delta)||T_n)|\\
&\le\mathbb{P}(y_{n+1}\in\hat{y}_{n+1}+[Q_{\alpha_1}^{\hat{F}_n}-\delta,Q_{\alpha_1}^{\hat{F}_n}+\delta]||T_n)+\mathbb{P}(y_{n+1}\in\hat{y}_{n+1}+[Q_{\alpha_2}^{\hat{F}_n}-\delta,Q_{\alpha_2}^{\hat{F}_n}+\delta]||T_n)\\
&\le 2\mathbb{E}[\min\{1,2\delta||f_{y_{n+1}|x_{n+1}}||_\infty]\text{ a.s.}
\end{aligned}\tag{4.6}$$

Note that (4.6) is a general result, i.e., holding for an arbitrary $\delta>0$. We will derive the existence of the fixed δ_ε from the stochastic boundedness of $||f_{y_{n+1}|x_{n+1}}||_\infty$. It implies the existence of a positive $M>0$, such that

$$\sup_{n\in\mathbb{N}}\mathbb{P}(||f_{y_{n+1}|x_{n+1}}||_\infty>M)\le\frac{\varepsilon}{4}. \tag{4.7}$$

Choose $\delta_\varepsilon=\frac{\varepsilon}{8M}$ and denote $Z:=\min\{1,2\delta_\varepsilon||f_{y_{n+1}|x_{n+1}}||_\infty\}$. Applying the same approach as in (4.4) yields

$$
\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}\left[Z\mathbb{1}\left\{2\delta_\varepsilon||f_{y_{n+1}|x_{n+1}}||_\infty > \frac{\varepsilon}{4}\right\} + Z\mathbb{1}\left\{2\delta_\varepsilon||f_{y_{n+1}|x_{n+1}}||_\infty \leq \frac{\varepsilon}{4}\right\}\right] \\
&\leq \mathbb{E}\left[1\mathbb{1}\left\{||f_{y_{n+1}|x_{n+1}}||_\infty > \frac{\varepsilon}{8\delta_\varepsilon}\right\} + \underbrace{2\delta_\varepsilon||f_{y_{n+1}|x_{n+1}}||_\infty \mathbb{1}\left\{2\delta_\varepsilon||f_{y_{n+1}|x_{n+1}}||_\infty \leq \frac{\varepsilon}{4}\right\}}_{\leq \varepsilon/4}\right] \\
&\leq \mathbb{P}\left(||f_{y_{n+1}|x_{n+1}}||_\infty > M\right) + \frac{\varepsilon}{4}.
\end{aligned} \tag{4.8}
$$

To sum up, (4.6), (4.7) and (4.8) together yield

$$
\begin{aligned}
&|\mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}(\delta_\varepsilon)||T_n) - \mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}(-\delta_\varepsilon)||T_n)| \\
&\overset{(4.6)}{\leq} \sup_{n\in\mathbb{N}} 2\mathbb{E}[\min\{1, 2\delta_\varepsilon||f_{y_{n+1}|x_{n+1}}||_\infty] \\
&\overset{(4.8)}{\leq} 2\sup_{n\in\mathbb{N}} \mathbb{P}(||f_{y_{n+1}|x_{n+1}}||_\infty > M) + \frac{\varepsilon}{2} \\
&\overset{(4.7)}{\leq} \varepsilon \text{ a.s.}
\end{aligned}
$$

This proves (4.5). We can now write the following chain of inequalities:

$$
\begin{aligned}
\mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}(\delta_\varepsilon)||T_n) - \varepsilon &\overset{(4.5)}{\leq} \mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}(-\delta_\varepsilon)||T_n) \\
&\leq \mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}||T_n) \\
&\leq \mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}(\delta_e)||T_n) \\
&\overset{(4.5)}{\leq} \mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}(-\delta_\varepsilon)||T_n) + \varepsilon \text{ a.s.}
\end{aligned} \tag{4.9}
$$

Observe further that for any choice of α_1, α_2, such that $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ it holds almost surely

$$
\begin{aligned}
&\{|\mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}||T_n) - (\alpha_2 - \alpha_1)| \geq 2\varepsilon\} \\
=&\{\mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}||T_n) - (\alpha_2 - \alpha_1) \geq 2\varepsilon\} \cup \{\mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}||T_n) - (\alpha_2 - \alpha_1) \leq -2\varepsilon\} \\
\overset{(4.9)}{\subseteq}& \left\{\sup_{0\leq\alpha_1\leq\alpha_2\leq 1} \{\mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}(-\delta_\varepsilon)||T_n) - (\alpha_2 - \alpha_1)\} \geq \varepsilon\right\} \\
&\cup \left\{\inf_{0\leq\alpha_1\leq\alpha_2\leq 1} \{\mathbb{P}(y_{n+1} \in PI^J_{\alpha_1,\alpha_2}(\delta_\varepsilon)||T_n) - (\alpha_2 - \alpha_1)\} \leq -\varepsilon\right\}.
\end{aligned}
$$

Recalling the fact that the supremum over $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ is attained by construction as a maximum of a finite set (see (4.1)), we can apply the general case to conclude

$$\begin{aligned}
&\mathbb{P}\left[\sup_{0\le\alpha_1\le\alpha_2\le1}|\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}||T_n)-(\alpha_2-\alpha_1)|\ge 2\varepsilon\right]\\
\le&\underbrace{\mathbb{P}\left[\sup_{0\le\alpha_1\le\alpha_2\le1}\{\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}(-\delta_\varepsilon)||T_n)-(\alpha_2-\alpha_1)\}\ge\varepsilon\right]}_{\xrightarrow{n\to\infty}0}\\
&+\underbrace{\mathbb{P}\left[\inf_{0\le\alpha_1\le\alpha_2\le1}\{\mathbb{P}(y_{n+1}\in PI^J_{\alpha_1,\alpha_2}(\delta_\varepsilon)||T_n)-(\alpha_2-\alpha_1)\}\le-\varepsilon\right]}_{\xrightarrow{n\to\infty}0}\xrightarrow{n\to\infty}0.
\end{aligned}$$

Since this holds for an arbitrary $\varepsilon>0$, the convergence in probability follows. Now we can apply the Dominated convergence theorem due to the boundedness of the supremum to derive the convergence in L_1. This finishes the proof.

□

Remark 4.5. Without the two assumptions in Lemma 4.3, we can only say that if there exists $\delta^*>0$, such that $\mathcal{L}_{\delta^*}(\hat{F}_n,F_n)\xrightarrow{p}0$, then the convergence in probability of the Levy gauge continues to hold for any $\delta>\delta^*$. This follows directly from the monotonicity of the Levy gauge stated in Lemma 2.26. Nevertheless, Lemma 4.3 provides a statement for any $\delta>0$, allowing us to derive results for any δ-distorted prediction interval. This justifies the distortion of prediction intervals, including its application to the proof of the continuous case in Theorem 4.4.

4.2 Jackknife+ method

A common characteristics of the Jackknife prediction intervals (either distorted or not) is the fact that they are centered around the predictor $\hat{y}_{n+1}$, see (4.2). Foygel Barber et al. 2021 modifies this approach and uses instead all leave-one-out analogues of the predictor, i.e., $\hat{y}^{\setminus 1}_{n+1},\hat{y}^{\setminus 2}_{n+1},\dots\hat{y}^{\setminus n}_{n+1}$. Thus, Foygel Barber et al. 2021 justifies the interpretation of the Jackknife+ prediction intervals as intervals around the median prediction based on $\hat{y}^{\setminus i}_{n+1},\forall i\in[n]$.

The theoretical foundations of the Jackknife+ method demand certain modifications of the objects from the previous section, which allow us to use the leave-one-out analogues of the predictor $\hat{y}_{n+1}$. Denote by

$$G_n(t):=\mathbb{P}(y_{n+1}\le t||T_n)$$

the cumulative distribution function of y_{n+1}, given the training data T_n. It is easy to see that G_n is nothing else than the function F_n shifted by the predictor $\hat{y}_{n+1}$, i.e., $G_n(t)=F_n(t-\hat{y}_{n+1})$. Therefore, it makes sense to define the corresponding shifted empirical distribution function

$$\hat{G}_n(t) := \hat{F}_n(t - \hat{y}_{n+1}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{u_i \leq t\},$$

where $u_i = y_i - \hat{y}_i^{\setminus i} + \hat{y}_{n+1}$.
Note that all results for the Levy gauge $\mathcal{L}_\delta(\hat{F}_n, F_n)$ still hold true if we use the corresponding shifted functions and observe instead $\mathcal{L}_\delta(\hat{G}_n, G_n)$. Furthermore, in view of $\hat{G}_n$ we can rewrite the prediction interval, based on the Jackknife estimator, that is,

$$PI^{J}_{\alpha_1,\alpha_2}(\delta) = [\hat{y}_{n+1} + Q^{\hat{F}_n}_{\alpha_1} - \delta, \hat{y}_{n+1} + Q^{\hat{F}_n}_{\alpha_2} + \delta] = [Q^{\hat{G}_n}_{\alpha_1} - \delta, Q^{\hat{G}_n}_{\alpha_2} + \delta] = {}_{\alpha_1}I^{\hat{G}_n}_{\alpha_2}(\delta).$$

The definition of $\hat{G}_n$ allows us to derive its equivalent in the Jackknife+ setting, i.e.,

$$\hat{G}_n^{+}(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{v_i \leq t\},$$

with $v_i = y_i - \hat{y}_i^{\setminus i} + \hat{y}_{n+1}^{\setminus i}$, and, in such a way, the quantile

$$Q^{\hat{G}_n^{+}}_{\alpha} = \inf\{t \in \mathbb{R} : \hat{G}_n^{+}(t) \geq \alpha\} = \begin{cases} -\infty, & \alpha \leq 0 \\ v_{(\lceil \alpha n \rceil)}, & \alpha \in (0,1] \\ +\infty, & \alpha > 1, \end{cases}$$

meaning that the latter can obtain finitely many values. In addition, denote by

$$PI^{J+}_{\alpha_1,\alpha_2}(\delta) := {}_{\alpha_1}I^{\hat{G}_n^{+}}_{\alpha_2}(\delta),$$

the Jackknife+ distorted prediction interval. We will write again only $PI^{J+}_{\alpha_1,\alpha_2}$ instead of $PI^{J+}_{\alpha_1,\alpha_2}(0)$.

Our goal now is to find some relations between $G_n, \hat{G}_n$ and $\hat{G}_n^{+}$ based on the Levy gauge. To do this, we first need an auxiliary result for the Levy gauge between functions with specific structure.

Lemma 4.6. *Let $n \in \mathbb{N}$. Consider two n-dimensional vectors $a = (a_i)_{i=1}^n, b = (b_i)_{i=1}^n$, as well as a vector of weghts $p = (p_i)_{i=1}^n$, i.e., $p_i \in [0,1], \forall i \in [n]$ and $\sum_{i=1}^n p_i = 1$. Define further*

$$F_a(t) := \sum_{i=1}^{n} p_i \mathbb{1}\{a_i \leq t\}, \quad F_b(t) := \sum_{i=1}^{n} p_i \mathbb{1}\{b_i \leq t\}.$$

Then the corresponding Levy gauge $\mathcal{L}_\delta(F_a, F_b)$ has an upper bound in the sense that

$$\forall \delta \geq 0 : \mathcal{L}_\delta(F_a, F_b) \leq \sum_{i=1}^{n} p_i \mathbb{1}\{|a_i - b_i| > \delta\}.$$

Proof. Since F_a and F_b are weighted distribution functions, the Levy gauge is well-defined. By Lemma 2.9, we have

$$\mathcal{L}_\delta(F_a, F_b) = \sup_{t\in\mathbb{R}} \max\{F_a(t) - F_b(t+\delta), F_b(t) - F_a(t+\delta)\}.$$

We can derive the following result, which holds true for all $t \in \mathbb{R}$:

$$\begin{aligned} F_a(t) - F_b(t+\delta) &= \sum_{i=1}^{n} p_i(\mathbb{1}\{a_i \le t\} - \mathbb{1}\{b_i - \delta \le t\}) \\ &\le \sum_{i=1}^{n} p_i \mathbb{1}\{a_i \le t < b_i - \delta\} \\ &\le \sum_{i=1}^{n} p_i \mathbb{1}\{\delta < b_i - a_i\} \\ &\le \sum_{i=1}^{n} p_i \mathbb{1}\{|a_i - b_i| > \delta\}. \end{aligned}$$

In a similar way we obtain

$$F_b(t) - F_a(t+\delta) \le \sum_{i=1}^{n} p_i \mathbb{1}\{|a_i - b_i| > \delta\}, \forall t \in \mathbb{R}.$$

To sum up, we conclude

$$\mathcal{L}_\delta(F_a, F_b) \le \sum_{i=1}^{n} p_i \mathbb{1}\{|a_i - b_i| > \delta\}.$$

□

We can now proceed to the equivalence result regarding the behaviour of the Levy gauge in the Jackknife and the Jackknife+ setting, respectively.

Lemma 4.7. *The following two statements are equivalent.*

1. $\forall \delta > 0 : \mathcal{L}_\delta(\hat{G}_n, G_n) \xrightarrow{p} 0.$
2. $\forall \delta > 0 : \mathcal{L}_\delta(\hat{G}_n^+, G_n) \xrightarrow{p} 0.$

Proof. We will first derive an upper bound for $\mathcal{L}_\delta(\hat{G}_n, \hat{G}_n^+)$ which holds almost surely. To see this, apply Lemma 4.6 by taking $p_i = \frac{1}{n}, a_i = u_i, b_i = v_i$, $\forall i \in [n]$, with u_i and v_i defined as above, and obtain

$$\mathcal{L}_\delta(\hat{G}_n, \hat{G}_n^+) \le \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{|\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}| > \delta\}. \tag{4.10}$$

The Markov inequality and the stability assumption of the predictor $\hat{y}_{n+1}$ (in the sense of Definition 4.1) gives the convergence in probability of $\mathcal{L}_\delta(\hat{G}_n, \hat{G}_n^+)$. More precisely, for all $\varepsilon > 0$ it holds

$$\mathbb{P}(\mathcal{L}_\delta(\hat{G}_n, \hat{G}_n^+) \geq \varepsilon) \leq \frac{1}{\varepsilon}\mathbb{E}[\mathcal{L}_\delta(\hat{G}_n, \hat{G}_n^+)] \overset{(4.10)}{\leq} \frac{1}{\varepsilon n}\sum_{i=1}^{n} \mathbb{P}(|\hat{y}_{n+1} - \hat{y}_{n+1}^{\setminus i}| > \delta) \overset{n\to\infty}{\longrightarrow} 0.$$

Assume that $\mathcal{L}_\delta(\hat{G}_n, G_n) \xrightarrow{p} 0, \forall \delta > 0$. We can use the relaxed triangle inequality given in Lemma 2.17 to conclude

$$0 \leq \mathcal{L}_\delta(\hat{G}_n^+, G_n) \leq \underbrace{\mathcal{L}_{\frac{\delta}{2}}(\hat{G}_n^+, \hat{G}_n)}_{\xrightarrow{p} 0} + \underbrace{\mathcal{L}_{\frac{\delta}{2}}(\hat{G}_n, G_n)}_{\xrightarrow{p} 0} \xrightarrow{p} 0, \quad \forall \delta > 0. \tag{4.11}$$

The reverse direction can be proven analogously. Assume that $\mathcal{L}_\delta(\hat{G}_n^+, G_n) \xrightarrow{p} 0, \forall \delta > 0$ and swap $\hat{G}_n$ and $\hat{G}_n^+$ in (4.11) to finish the proof. □

This allows us to formulate a corollary with a similar statement as Theorem 4.4 but this time for the Jackknife+ method. This is Corollary 4.2 in Amann, Leeb and Steinberger 2025.

Corollary 4.8 (Asymptotically conservative/valid prediction intervals with the Jackknife+)**.** *Suppose that the prediction error $y_{n+1} - \hat{y}_{n+1}$ is stochastically bounded and the predictor $\hat{y}_{n+1}$ is stable. Then the following statements hold true.*

1. ***General case**: For every $\delta > 0$ the δ-inflated Jackknife+ prediction intervals are asymptotically uniformly conditionally conservative, that is,*

$$\lim_{n\to\infty} \mathbb{P}\left[\inf_{0\leq\alpha_1\leq\alpha_2\leq 1} \left\{\mathbb{P}(y_{n+1} \in PI_{\alpha_1,\alpha_2}^{J+}(\delta)||T_n) - (\alpha_2 - \alpha_1)\right\} \geq -\varepsilon\right] = 1, \forall \varepsilon > 0.$$

   Additionally, the conditional coverage probability of δ-shrunken prediction intervals uniformly does not overshoot its nominal level asymptotically, that is,

$$\lim_{n\to\infty} \mathbb{P}\left[\sup_{0\leq\alpha_1\leq\alpha_2\leq 1} \left\{\mathbb{P}(y_{n+1} \in PI_{\alpha_1,\alpha_2}^{J+}(-\delta)||T_n) - (\alpha_2 - \alpha_1)\right\} \leq \varepsilon\right] = 1, \forall \varepsilon > 0.$$

2. ***Continuous case:** Assume further that the CC-Assumption (see Definition 4.2) is satisfied. Then the non-distorted prediction interval (in other words, the case $\delta = 0$) is asymptotically uniformly conditionally valid, that is,*

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{0\leq\alpha_1\leq\alpha_2\leq 1} \left|\mathbb{P}(y_{n+1} \in PI_{\alpha_1,\alpha_2}^{J+}||T_n) - (\alpha_2 - \alpha_1)\right|\right] = 0.$$

Proof. Proceeding similarly to the proof of Theorem 4.4, take

$$U = y_{n+1}, F = G_n, G = \hat{G}_n^+,$$

as well as $0 < \delta' < \delta$ and use Corollary 3.11 to conclude

$$\mathbb{P}\left[\inf_{0\leq\alpha_1\leq\alpha_2\leq 1}\left\{\mathbb{P}(y_{n+1}\in {}_{\alpha_1}I_{\alpha_2}^{\hat{G}_n^+}(\delta)||T_n)-(\alpha_2-\alpha_1)\right\}\leq -\varepsilon\right]\leq \mathbb{P}\left(\mathcal{L}_\delta(G_n,\hat{G}_n^+)\geq\frac{\varepsilon}{2}\right)$$
$$\mathbb{P}\left[\sup_{0\leq\alpha_1\leq\alpha_2\leq 1}\left\{\mathbb{P}(y_{n+1}\in {}_{\alpha_1}I_{\alpha_2}^{\hat{G}_n^+}(-\delta)||T_n)-(\alpha_2-\alpha_1)\right\}\geq \varepsilon\right]\leq \mathbb{P}\left(\mathcal{L}_{\delta'}(G_n,\hat{G}_n^+)\geq\frac{\varepsilon}{2}\right).$$

In view of Lemma 4.7 $\mathcal{L}_\delta(G_n,\hat{G}_n^+)\xrightarrow{p}0\iff\mathcal{L}_\delta(G_n,\hat{G}_n)\xrightarrow{p}0$. It remains to apply Lemma 4.3 to derive the statements in the general case.
The proof of the continuous case is the same as in Theorem 4.4, we just have to consider the Jackknife+ prediction intervals instead of the Jackknife prediction intervals. The application of the general case finishes the proof. □

Bibliography

Amann, Nicolai (2023). "Assumption-lean conditional predictive inference via the Jackknife". PhD thesis. University of Vienna.

Amann, Nicolai, Hannes Leeb and Lukas Steinberger (2025). "Uncertainty quantification via cross-validation and its variants under algorithmic stability". arXiv: 2312.14596 [math.ST]. URL: https://arxiv.org/abs/2312.14596.

Foygel Barber, Rina et al. (2021). "Predictive inference with the jackknife+". In: *The Annals of Statistics* 49, pp. 486–507.

Miller, Rupert G. (1974). "The Jackknife–A Review". In: *Biometrika* 61, pp. 1–15. URL: http://www.jstor.org/stable/2334280.

Quenouille, Maurice H. (1956). "Notes on Bias in Estimation". In: *Biometrika* 43, pp. 353–360. URL: http://www.jstor.org/stable/2332914.

Steinberger, Lukas and Hannes Leeb (2023). "Conditional predictive inference for stable algorithms". In: *The Annals of Statistics* 51, pp. 290–311.

Tukey, John W. (1958). "Bias and confidence in not-quite large samples". In: *Annals of Mathematical Statistics* 29, p. 614.