STL Robustness Risk over Discrete-Time Stochastic Processes

Lars Lindemann, Nikolai Matni, and George J. Pappas

***Abstract*— We present a framework to interpret signal temporal logic (STL) formulas over discrete-time stochastic processes in terms of the induced risk. Each realization of a stochastic process either satisfies or violates an STL formula. In fact, we can assign a robustness value to each realization that indicates how robustly this realization satisfies an STL formula. We then define the risk of a stochastic process not satisfying an STL formula robustly, referred to as the "*STL robustness risk*". In our definition, we permit general classes of risk measures such as, but not limited to, the value-at-risk. While in general hard to compute, we propose an approximation of the STL robustness risk. This approximation has the desirable property of being an upper bound of the STL robustness risk when the chosen risk measure is monotone, a property satisfied by most risk measures. Motivated by the interest in data-driven approaches, we present a sampling-based method for calculating an upper bound of the approximate STL robustness risk for the value-at-risk that holds with high probability. While we consider the case of the value-at-risk, we highlight that such sampling-based methods are viable for other risk measures.**

I. Introduction

Consider the scenario in which an autonomous car equipped with noisy sensors navigates through urban traffic. As a consequence of imperfect sensing, the environment is not perfectly known. Instead, we can describe the scenario as a stochastic process that models each possible outcome along with the probability of an outcome. In this paper, we are interested in quantifying the associated risk in such safety-critical systems. In particular, we consider system specifications that are formulated in signal temporal logic (STL) [1] and, for the first time, propose a systematic way to assess the risk associated with such system specifications when evaluated over discrete-time stochastic processes.

Signal temporal logic has been introduced as a formalism to express a large variety of complex system specifications. STL particularly allows to express temporal and spatial system properties, e.g., surveillance ("visit regions A, B, and C every $10-60$ sec"), safety ("always between $5-25$ sec stay at least 1 m away from region D"), and many others. STL specifications are evaluated over deterministic signals and a given signal, for instance the trajectory of a robot, either satisfies or violates the STL specification at hand. Towards quantifying the robustness by which a signal satisfies an STL specification, the authors in [2] proposed the robustness degree as a (maximal) tube around a nominal signal so that all signals in this tube either satisfy or violate the specification. In this way, the size of the tube describes the robustness of the nominal signal with respect to the specification. As the robustness degree is in general hard to calculate, the authors in [2] and many other works proposed approximate yet easier to calculate robust semantics such as the space and the time robustness [3], the arithmetic-geometric mean robustness [4], the smooth cumulative robustness [5], averaged STL [6], or robustness metrics tailored for guiding reinforcement learning algorithms [7]. Further related is the work [8] where metrics for STL formulas are presented as well as [9] in which a connection with linear time-invariant filtering is established allowing to define various types of quantitative semantics.

All of the aforementioned works deal with deterministic signals. For stochastic signals, the authors in [10]–[14] propose notions of probabilistic signal temporal logic in which chance constraints are defined over the atomic elements (called predicates) of STL, while the Boolean and temporal operators of STL are not altered. Similarly, notions of risk signal temporal logic have recently appeared in [15] and [16] by defining risk constraints over the atomic elements only. The work in [17] considers the probability of an STL specification being satisfied instead of applying chance or risk constraints on the atomic level. More with a control synthesis focus and for the less expressive formalism of linear temporal logic, the authors in [18]–[20] consider control over belief spaces, while the authors in [21] consider probabilistic satisfaction over Markov decision processes. In contrast, in this work we quantify the risk of not satisfying an STL specification robustly. Probably closest to our paper is the work in [22] in which the authors present a framework for the robustness of STL under stochastic models. Our work differs from [22] in several directions. Most importantly, we do not limit our attention to probabilities, but allow for general risk measures, including probabilities, towards an axiomatic risk theory for temporal logics. We also argue that the STL robustness risk should conceptually be defined differently than in [22]. We further that the STL semantics are measurable and present an efficient way to calculate the STL robustness risk for the value-at-risk.

The theory of risk has a longstanding history of use in finance [23], [24]. More recently, there has been an interest to also apply such risk measures in robotics and control applications [25]. Risk-aware control and estimation frameworks have recently appeared in [26]–[33] using various forms of risk. We remark that these frameworks are orthogonal to our work as they present design tools while we provide a generic framework for quantifying the risk of complex

The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: larsl@seas.upenn.edu (L. Lindemann), nmatni@seas.upenn.edu (N. Matni), and pappasg@seas.upenn.edu (G.J. Pappas)

This research was generously generously supported by NSF award CPS-2038873 and NSF CAREER award ECCS-2045834.

system specifications expressed in STL. We hope that such quantification will be useful to guide the design and analysis process in the future.

In this paper, we consider signal temporal logic specifications interpreted over discrete-time stochastic processes. Our contributions can be summarized as follows:

1) We first show that the semantics, the robust semantics, and the robustness degree of STL are measurable functions so that probabilities or the risk over these functions is well-defined.
2) We define the risk of a discrete-time stochastic process not satisfying an STL specification robustly and refer to this definition as the "robustness risk".
3) We show that the robustness risk is in general hard to calculate and propose an approximation of the robustness risk that has the desireable property of being an over-approximation of the robustness risk whenever the employed risk measure is monotonic.
4) We present a sampling-based estimate of the approximate robustness risk when the value-at-risk is used. We show that this estimate is an upper bound of the approximate robustness risk with high probability. We thereby establish an interesting connection between data-driven design approaches and the risk of an STL specification. In this way, one can make statements about the risk of machine learning algorithms.

In Section II, we present background on signal temporal logic, stochastic processes, and risk measures. In Section III, we define the robustness risk, while we show in Section IV how the approximate robustness risk can be obtained via a sampling-based method for the case of the value-at-risk. A case study is presented in Section V followed by conclusions in Section VI. The appendix, including all technical proofs, can be found in https://github.com/Lindemann1989/STLRisk.

II. Background

True and false are encoded as $\top := 1$ and $\bot := -1$, respectively, with the set $\mathbb{B} := \{\top, \bot\}$. Let $\mathbb{R}$ and $\mathbb{N}$ be the set of real and natural numbers. Also let $\mathbb{R}_{\geq 0}$ be the set of non-negative real numbers and $\mathbb{R}^n$ be the real n-dimensional vector space. For a metric space (S, d), a point $s \in S$, and a nonempty set $S' \subseteq S$, let $\bar{d}(s, S') := \inf_{s' \in S'} d(s, s')$ be the distance of s to S'. It holds that the function $\bar{d}(s, S')$ is continuous in s [34, Chapter 3]. For $t \in \mathbb{R}$ and $I \subseteq \mathbb{R}$, let $t \oplus I$ and $t \ominus I$ denote the Minkowski sum and the Minkowski difference of t and I, respectively. For $a, b \in \mathbb{R}$, let

$$\mathbb{I}(a \leq b) := \begin{cases} 1 & \text{if } a \leq b \\ 0 & \text{otherwise} \end{cases}$$

be the indicator function indicating whether or not $a \leq b$. Let $\mathfrak{F}(T, S)$ denote the set of all measurable functions mapping from the domain T into the domain S, i.e., $f \in \mathfrak{F}(T, S)$ is a function $f : T \to S$.

A. Signal Temporal Logic

Signal temporal logic [1] is based on deterministic signals $x : T \to \mathbb{R}^n$. The atomic elements of STL are predicates that are functions $\mu : \mathbb{R}^n \to \mathbb{B}$. Let now M be a set of such predicates μ and let us associate an observation map $O^\mu \subseteq \mathbb{R}^n$ with μ. The observation map O^μ indicates regions within the state space where μ is true, i.e.,

$$O^\mu := \mu^{-1}(\top)$$

where $\mu^{-1}(\top)$ denotes the inverse image of $\top$ under μ. We assume throughout the paper that the sets O^μ and $O^{\neg\mu}$ are non-empty and measurable for any $\mu \in M$, i.e., O^μ and $O^{\neg\mu}$ are elements of the Borel σ-algebra $\mathcal{B}^n$.

Remark 1: For convenience, the predicate μ is often defined via a predicate function $h : \mathbb{R}^n \to \mathbb{R}$ so that

$$\mu(\zeta) := \begin{cases} \top & \text{if } h(\zeta) \geq 0 \\ \bot & \text{otherwise} \end{cases}$$

for $\zeta \in \mathbb{R}^n$. In this case, we have $O^\mu = \{\zeta \in \mathbb{R}^n | h(\zeta) \geq 0\}$.

For $\mu \in M$, the syntax of STL, also referred to as the grammar of STL, is defined as

$$\phi \ ::= \ \top \mid \mu \mid \neg\phi \mid \phi' \wedge \phi'' \mid \phi' U_I \phi'' \mid \underline{U}_I \phi'' \qquad (1)$$

where ϕ' and ϕ'' are STL formulas and where U_I is the future until operator with $I \subseteq \mathbb{R}_{\geq 0}$, while $\underline{U}_I$ is the past until-operator. The operators $\neg$ and $\wedge$ encode negations and conjunctions. Also define the set of operators

$$\begin{aligned}
\phi' \vee \phi'' &:= \neg(\neg\phi' \wedge \neg\phi'') && \text{(disjunction operator)},\\
F_I\phi &:= \top U_I \phi && \text{(future eventually operator)},\\
\underline{F}_I\phi &:= \top \underline{U}_I \phi && \text{(past eventually operator)},\\
G_I\phi &:= \neg F_I \neg\phi && \text{(future always operator)},\\
\underline{G}_I\phi &:= \neg \underline{F}_I \neg\phi && \text{(past always operator)}.
\end{aligned}$$

1) Semantics: We can now give an STL formula ϕ as in (1) a meaning by defining the satisfaction function $\beta^\phi : \mathfrak{F}(T, \mathbb{R}^n) \times T \to \mathbb{B}$. In particular, $\beta^\phi(x, t) = \top$ indicates that the signal x satisfies the formula ϕ at time t, while $\beta^\phi(x, t) = \bot$ indicates that x does not satisfy ϕ at time t. For a formal definition of $\beta^\phi(x, t)$, we refer to Definition 3 in Appendix I. An STL formula ϕ is said to be satisfiable if $\exists x \in \mathfrak{F}(T, \mathbb{R}^n)$ such that $\beta^\phi(x, 0) = \top$. The following example is used as a running example throughout the paper.

Example 1: Consider a scenario in which a robot operates in a hospital environment. The robot needs to perform two time-critical sequential delivery tasks in regions A and B while avoiding areas C and D in which humans operate. In particular, we consider the STL formula

$$\phi := G_{[0,3]}\neg(\mu_C \wedge \mu_D) \wedge F_{[1,2]}(\mu_A \wedge F_{[0,1]}\mu_B). \qquad (2)$$

To define μ_A, μ_B, μ_C, and μ_D, let a, b, c, and d denote the midpoints of the regions A, B, C, and D as

$$\begin{aligned}
a &:= \begin{bmatrix} 4 & 5 \end{bmatrix}^T \qquad & b &:= \begin{bmatrix} 7 & 2 \end{bmatrix}^T\\
c &:= \begin{bmatrix} 2 & 3 \end{bmatrix}^T \qquad & d &:= \begin{bmatrix} 6 & 4 \end{bmatrix}.^T
\end{aligned}$$

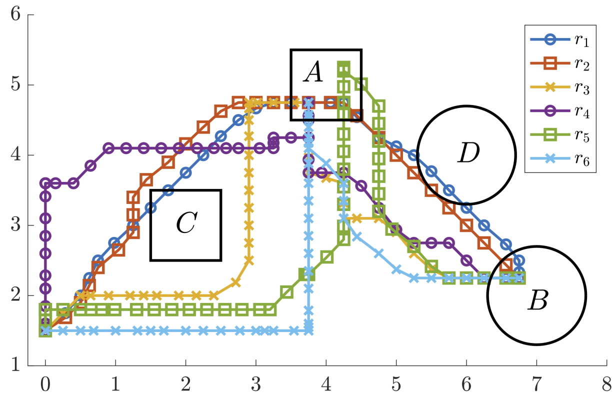

Fig. 1: Shown are six robot trajectories r_1-r_6 along with four regions A, B, C, and D. The specification given in (2) is violated by r_1 and satisfied by r_2-r_6. Furthermore, r_2 only marginally satisfies ϕ, while r_3-r_6 satisfy ϕ robustly.

Also let the state $x(t) \in \mathbb{R}^{10}$ at time t be

$$x(t) := \begin{bmatrix} r(t) & a & b & c & d \end{bmatrix}^T$$

where r is the robot position at time t. The predicates μ_A, μ_B, μ_C, and μ_D are now described by the observation maps

$$\begin{aligned} O^{\mu_A} &:= \{x \in \mathbb{R}^{10} | \|r - a\|_\infty \leq 0.5\}, \\ O^{\mu_B} &:= \{x \in \mathbb{R}^{10} | \|r - b\|_2 \leq 0.7\}, \\ O^{\mu_C} &:= \{x \in \mathbb{R}^{10} | \|r - c\|_\infty \leq 0.5\}, \\ O^{\mu_D} &:= \{x \in \mathbb{R}^{10} | \|r - d\|_2 \leq 0.7\}. \end{aligned}$$

where $\|\cdot\|_2$ is the Euclidean and $\|\cdot\|_\infty$ is the infinity norm. In Fig. 1, six different robot trajectories r_1-r_6 are displayed. It can be seen that the signal x_1 that corresponds to r_1 violates ϕ, while x_2-x_6 satisfy ϕ, i.e., we have $\beta^\phi(x_1, 0) = \top$ and $\beta^\phi(x_i, 0) = \perp$ for all $i \in \{2, \ldots, 6\}$.

2) Robustness: One may now be interested in more information than just whether or not the signal x satisfies the STL formula ϕ at time t and consider the quality of satisfaction. Therefore, one can look at the robustness by which a signal x satisfies the STL formula ϕ at time t. For this purpose, the *robustness degree* has been introduced in [2, Definition 7].

Let us define the set of signals that satisfy ϕ at time t as

$$\mathcal{L}^\phi(t) := \{x \in \mathfrak{F}(T, \mathbb{R}^n) | \beta^\phi(x, t) = \top\}.$$

Intuitively, the robustness degree then tells us how much the signal x can be perturbed by additive noise before changing the Boolean truth value of the specification ϕ. Towards a formal definition, let us define the signal metric

$$\kappa(x, x^*) := \sup_{t \in T} d\big(x(t), x^*(t)\big)$$

where d is a vector metric assigning a distance in $\mathbb{R}^n$. The metric d can, for instance, be the Euclidean norm. Note that $\kappa(x, x^*)$ is the L_∞ norm of the signal $x - x^*$. The distance of x to the set $\mathcal{L}^\phi(t)$ is then defined via the metric κ as

$$\text{dist}^\phi(x, t) = \bar{\kappa}\big(x, \text{cl}(\mathcal{L}^\phi(t))\big) := \inf_{x^* \in \text{cl}(\mathcal{L}^\phi(t))} \kappa(x, x^*),$$

where $\text{cl}(\mathcal{L}^\phi(t))$ denotes the closure of $\mathcal{L}^\phi(t)$. The robustness degree is now given in Definition 1.

Definition 1 (Robustness Degree): Given an STL formula ϕ and a signal $x \in \mathfrak{F}(T, \mathbb{R}^n)$, the robustness degree at time t is defined as [2, Definition 7]:

$$\mathcal{RD}^\phi(x, t) := \begin{cases} \text{dist}^{\neg\phi}(x, t) \text{ if } x \in \mathcal{L}^\phi(t) \\ -\text{dist}^\phi(x, t) \text{ if } x \notin \mathcal{L}^\phi(t). \end{cases}$$

By definition, the following holds. If $|\mathcal{RD}^\phi(x, t)| \neq 0$ and $x \in \mathcal{L}^\phi(t)$, it follows that all signals $x^* \in \mathfrak{F}(T, \mathbb{R}^n)$ that are such that $\kappa(x, x^*) < |\mathcal{RD}^\phi(x, t)|$ satisfy $x^* \in \mathcal{L}^\phi(t)$.

The robustness degree is a *robust neighborhood*. A robust neighborhood of x is a tube of diameter $\epsilon \geq 0$ around x so that for all x^* in this tube we have $\beta^\phi(x, t) = \beta^\phi(x^*, t)$. Specifically, for $\epsilon \geq 0$ and $x : T \to \mathbb{R}^n$ with $x \in \mathcal{L}^\phi(t)$, a set $\{x' \in \mathfrak{F}(T, \mathbb{R}^n) | \kappa(x, x') < \epsilon\}$ is a robust neighborhood if $x^* \in \{x' \in \mathfrak{F}(T, \mathbb{R}^n) | \kappa(x, x') < \epsilon\}$ implies $x^* \in \mathcal{L}^\phi(t)$.

3) Robust Semantics: Note that it is in general difficult to calculate the robustness degree $\mathcal{RD}^\phi(x, t)$ as the set $\mathcal{L}^\phi(t)$ is hard to obtain. The authors in [2] introduce the *robust semantics* $\rho^\phi : \mathfrak{F}(T, \mathbb{R}^n) \times T \to \mathbb{R}$ as an alternative way of finding a robust neighborhood. For a formal definition, we refer to Definition 4 in Appendix I.

Importantly, it was shown in [2, Theorems 28] that

$$-\text{dist}^\phi(x, t) \leq \rho^\phi(x, t) \leq \text{dist}^{\neg\phi}(x, t). \tag{3}$$

In other words, it holds that

$$\begin{cases} 0 \leq \rho^\phi(x, t) \leq \text{dist}^{\neg\phi}(x, t) \text{ if } x \in \mathcal{L}^\phi(t) \\ -\text{dist}^\phi(x, t) \leq \rho^\phi(x, t) \leq 0 \text{ if } x \notin \mathcal{L}^\phi(t) \end{cases}$$

so that $|\rho^\phi(x, t)| \leq |\mathcal{RD}^\phi(x, t)|$. The robust semantics $\rho^\phi(x, t)$ hence provide a more tractable under-approximation of the robustness degree $\mathcal{RD}^\phi(x, t)$. The robust semantics are sound in the following sense [2, Proposition 30]:

$$\begin{aligned} \beta^\phi(x, t) &= \top \text{ if } \rho^\phi(x, t) > 0, \\ \beta^\phi(x, t) &= \perp \text{ if } \rho^\phi(x, t) < 0. \end{aligned}$$

This result allows to use the robust semantics when reasoning over satisfaction of an STL formula ϕ.

Example 2: For Example 1 and the trajectories shown in Fig. 1, we obtain $\rho^\phi(x_1, 0) = -0.15$, $\rho^\phi(x_2, 0) = 0.01$, and $\rho^\phi(x_i, 0) = 0.25$ for all $i \in \{3, \ldots, 6\}$. The reason for x_1 having negative robustness lies in r_1 intersecting with the region D. Marginal robustness of x_2 is explained as r_2 only marginally avoids the region D while all other trajectories avoid the region D robustly.

B. Random Variables and Stochastic Processes

Instead of interpreting an STL specifications ϕ over deterministic signals, we will interpret ϕ over stochastic processes. Consider therefore the *probability space* $(\Omega, \mathcal{F}, P)$ where Ω is the sample space, $\mathcal{F}$ is a σ-algebra of Ω, and $P : \mathcal{F} \to [0, 1]$ is a probability measure. More intuitively, an element in Ω is an *outcome* of an experiment, while an element in $\mathcal{F}$ is an *event* that consists of one or more outcomes whose probabilities can be measured by the probability measure P.

1) Random Variables: Let Z denote a real-valued *random vector*, i.e., a measurable function $Z : \Omega \to \mathbb{R}^n$.[1] When $n = 1$, we say Z is a *random variable*. We refer to $Z(\omega)$ as a realization of the random vector Z where $\omega \in \Omega$. Since Z is a measurable function, a probability space can be defined for Z so that probabilities can be assigned to events related to values of Z. Consequently, a cumulative distribution function (CDF) $F_Z(z)$ can be defined for Z (see Appendix II).

Given a random vector Z, we can derive other random variables that we call *derived random variables*. Assume for instance a measurable function $g : \mathbb{R}^n \to \mathbb{R}$ and notice that $G : \Omega \to \mathbb{R}$ with $G(\omega) := g(Z(\omega))$ becomes yet another random variable since function composition preserves measureability. See [35] for a more detailed discussion.

2) Stochastic Processes: A *stochastic process* is a function of the variables $\omega \in \Omega$ and $t \in T$ where T is the time domain. We restrict the time domain to be discrete, i.e., $T := \mathbb{N}$, and consider discrete-time stochastic processes. This assumption is made for simplicity. The presented results carry over, with some modifications, to the continuous-time case that we defer to another paper. A stochastic process is now a function $X : T \times \Omega \to \mathbb{R}^n$ where $X(t, \cdot)$ is a random vector for each fixed $t \in T$. A stochastic process can be viewed as a collection of random vectors $\{X(t, \cdot)|t \in T\}$ that are defined on a common probability space $(\Omega, \mathcal{F}, P)$ and that are indexed by T. For a fixed $\omega \in \Omega$, the function $X(\cdot, \omega)$ is a *realization* of the stochastic process. Another equivalent definition is that a stochastic process is a collection of deterministic functions of time

$$\{X(\cdot, \omega)|\omega \in \Omega\} \tag{4}$$

that are indexed by Ω. While the former definition is intuitive, the latter allows to define a *random function* mapping from the sample space Ω into the space of functions $\mathfrak{F}(T, \mathbb{R}^n)$.

C. Risk Measures

A *risk measure* is a function $R : \mathfrak{F}(\Omega, \mathbb{R}) \to \mathbb{R}$ that maps from the set of real-valued random variables to the real numbers. In particular, we refer to the input of a risk measure R as the *cost random variable* since typically a cost is associated with the input of R. Risk measures hence allow for a risk assessment in terms of such cost random variables. Commonly used risk measures are the expected value, the variance, or the conditional value-at-risk [23]. A particular property of R that we need in this paper is monotonicity. For two cost random variables $Z, Z' \in \mathfrak{F}(\Omega, \mathbb{R})$, the risk measure R is monotone if $Z(\omega) \leq Z'(\omega)$ for all $\omega \in \Omega$ implies that $R(Z) \leq R(Z')$.

Remark 2: In Appendix III, we summarize other desireable properties of R such as translation invariance, positive homogeneity, subadditivity, commotone additivity, and law invariance. We also provide a summary of existing risk measures. We emphasize that our presented method is compatible with any of these risk measures as long as they are monotone.

[1]More precisely, we have $Z : \Omega \times \mathcal{F} \to \mathbb{R}^n \times \mathcal{B}^n$ where $\mathcal{B}^n$ is the Borel σ-algebra of $\mathbb{R}^n$, i.e., Z maps a *measurable space* to yet another measurable space. For convenience, this more involved notation is, however, omitted.

III. Risk of STL Specifications

While an STL formula ϕ as defined in Section II-A is defined over deterministic signals x, we will in this paper evaluate the STL formula ϕ over a stochastic process X.

Recall from Section II-B that a stochastic process X can be interpreted as a random function mapping from the sample space Ω into the space of functions $\mathfrak{F}(T, \mathbb{R}^n)$. For a particular realization $X(\cdot, \omega)$ of the stochastic process X, note that we can evaluate whether or not $X(\cdot, \omega)$ satisfies ϕ.

For the stochastic process X, however, it is not clear how to interpret the satisfaction of ϕ by X. In fact, some realizations of X may satisfy ϕ while some other realizations of X may violate ϕ. This fact leads us to use risk measures as introduced in Section II-C to argue about the risk of the stochastic process X not satisfying the specification ϕ.

Before going into the main parts of this paper, we remark that all important symbols that have been or will be introduced are summarized in Table I. We also remark that all technical proofs are provided in the appendix.

A. Measurability of STL Semantics and Robustness Degree

Note that the semantics and the robust semantics as well as the robustness degree become stochastic entities when evaluated over a stochastic process X, i.e., the functions $\beta^\phi(X, t)$, $\rho^\phi(X, t)$, and $RD^\phi(X, t)$ become stochastic entities.

We first provide conditions under which $\beta^\phi(X, t)$ and $\rho^\phi(X, t)$ become (derived) random variables, which bows down to showing that $\beta^\phi(X(\cdot, \omega), t)$ and $\rho^\phi(X(\cdot, \omega), t)$ are measurable in ω for a fixed $t \in T$.

Theorem 1: Let X be a discrete-time stochastic process and let ϕ be an STL specification. Then $\beta^\phi(X(\cdot, \omega), t)$ and $\rho^\phi(X(\cdot, \omega), t)$ are measurable in ω for a fixed $t \in T$ so that $\beta^\phi(X, t)$ and $\rho^\phi(X, t)$ are random variables.

By Theorem 1, the probabilities $P(\beta^\phi(X, t) \in B)$ and $P(\rho^\phi(X, t) \in B)$[2] are well defined for measurable sets B from the corresponding measurable space.

We next show measurability of the distance function $\text{dist}^\phi(X(\cdot, \omega), t)$ and the robustness degree $\mathcal{RD}^\phi(X(\cdot, \omega), t)$.

Theorem 2: Let X be a discrete-time stochastic process and let ϕ be an STL specification. Then $\text{dist}^\phi(X(\cdot, \omega), t)$ and $\mathcal{RD}^\phi(X(\cdot, \omega), t)$ are measurable in ω for a fixed $t \in T$ so that $\text{dist}^\phi(X, t)$ and $\mathcal{RD}^\phi(X, t)$ are random variables.

B. The STL Robustness Risk

Towards defining the risk of not satisfying a specification ϕ, note that the expression $R(\beta^\phi(X, t) = \perp)$ is not well defined as opposed to $P(\beta^\phi(X, t) = \perp)$ that indicates the probability of not satisfying ϕ. The reason for this is that the function R takes a real-valued cost random variable as its input. We can instead evaluate $R(-\beta^\phi(X, t))$, but not much information will be gained due to the binary encoding of the STL semantics $\beta^\phi(X, t)$.

[2]We use the shorthand notations $P(\beta^\phi(X, t) \in B)$ and $P(\rho^\phi(X, t) \in B)$ instead of the more complex notations $P(\{\omega \in \Omega|\beta^\phi(X(\cdot, \omega), t) \in B\})$ and $P(\{\omega \in \Omega|\rho^\phi(X(\cdot, \omega), t) \in B\})$, respectively.

Symbol	Meaning
x	Deterministic signal $x: T \to \mathbb{R}^n$
$\mathfrak{F}(T, \mathbb{R}^n)$	Set of all measurable functions mapping from T to $\mathbb{R}^n$
$\beta^\phi(x,t)$	Boolean semantics $\beta^\phi: \mathfrak{F}(T,\mathbb{R}^n) \times T \to \mathbb{B}$ of an STL formula ϕ at time t
$\mathcal{L}^\phi(t)$	Set of deterministic signals x that satisfy ϕ at time t
$\text{dist}^\phi(x,t)$	Distance of the signal x to the set $\mathcal{L}^\phi(t)$
$\mathcal{RD}^\phi(x,t)$	Robustness degree $\mathcal{RD}^\phi: \mathfrak{F}(T,\mathbb{R}^n) \times T \to \mathbb{R}$ of an STL formula ϕ at time t
$\rho^\phi(x,t)$	Robust semantics $\rho^\phi: \mathfrak{F}(T,\mathbb{R}^n) \times T \to \mathbb{R}$ of an STL formula ϕ at time t
X	Stochastic Process $X: T \times \Omega \to \mathbb{R}^n$
$R(-\text{dist}^{\neg\phi}(X,t))$	The risk of the stochastic process X not satisfying the STL formula ϕ robustly at time t
$R(-\rho^\phi(X,t))$	The approximate risk of the stochastic process X not satisfying the STL formula ϕ at time t

TABLE I: Summary of robustness and risk notions for signal temporal logic.

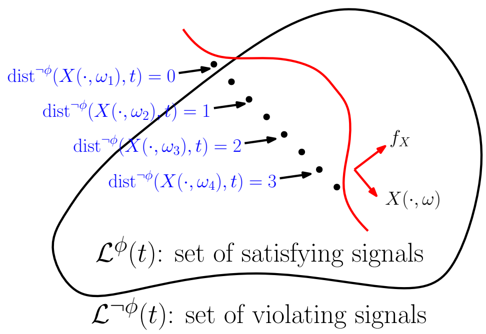

Fig. 2: Illustration of $\text{dist}^{\neg\phi}(X(\cdot,\omega_i),t)$, i.e., the distance between the realization $X(\cdot,\omega_i)$ and the set $\mathcal{L}^{\neg\phi}(t)$.

1) The risk of not satisfying ϕ robustly: Instead, we will define the risk of the stochastic process X not satisfying ϕ robustly by considering $\text{dist}^{\neg\phi}(X,t)$, that is the distance between the stochastic process X and the set of violating signals $\mathcal{L}^{\neg\phi}(t)$. We refer to this definition as the "robustness risk" for brevity.

Definition 2: Given an STL formula ϕ and a stochastic process $X: T \times \Omega \to \mathbb{R}^n$, the risk of X not satisfying ϕ robustly at time t is defined as

$$R(-\text{dist}^{\neg\phi}(X,t)).$$

Fig. 2 illustrates the idea underlying Definition 2 and shows $\text{dist}^{\neg\phi}(X(\cdot,\omega_i),t)$ for realizations $X(\cdot,\omega_i)$ of the stochastic process X where $\omega_i \in \Omega$, i.e., $\text{dist}^{\neg\phi}(X(\cdot,\omega_i),t)$ is the distance between the realization $X(\cdot,\omega_i)$ and the set $\mathcal{L}^{\neg\phi}(t)$. Positive values of $\text{dist}^{\neg\phi}(X(\cdot,\omega_i),t)$ indicate that the realization $X(\cdot,\omega_i)$ satisfies ϕ at time t, while the value zero indicates that the realization $X(\cdot,\omega_i)$ either marginally satisfies ϕ at time t or does not satisfy ϕ at time t. Furthermore, large positive values of $\text{dist}^{\neg\phi}(X(\cdot,\omega_i),t)$ indicate robust satisfaction and are hence desirable. This is the reason why $-\text{dist}^{\neg\phi}(X,t)$ is considered in Definition 2 as the cost random variable. To complement Fig. 2, note that the red curve sketches a possible distribution of X and hence the probability by which a realization occurs. Note that the robustness degree of the corresponding realizations in Fig. 2 would be $\mathcal{RD}^\phi(X(\omega_1),t) < 0$, $\mathcal{RD}^\phi(X(\omega_2),t) = 1$, $\mathcal{RD}^\phi(X(\omega_3),t) = 2$, and $\mathcal{RD}^\phi(X(\omega_4),t) = 3$.

Example 3: Let us consider the value-at-risk at level $\beta := 0.9$ and let $VaR_\beta(-\text{dist}^{\neg\phi}(X,t)) = -0.5$ be the calculated risk for a given stochastic process X and a given STL formula ϕ. The interpretation is now that the probability of a robustness $\text{dist}^{\neg\phi}(X,t)$ smaller than $|VaR_\beta(-\text{dist}^{\neg\phi}(X,t))| = 0.5$ is smaller than $1-\beta$.

Remark 3: An alternative to Definition 2 would be to consider $R(-\mathcal{RD}^\phi(X,t))$ instead of $R(-\text{dist}^{\neg\phi}(X,t))$. We, however, refrain from such a definition since the meaning of $\mathcal{RD}^\phi(X(\cdot,\omega_i),t)$ for realizations $w_i \in \Omega$ with $\mathcal{RD}^\phi(X(\cdot,\omega_i),t) < 0$ is not what we aim for here. In particular, when $\mathcal{RD}^\phi(X(\cdot,\omega_i),t) < 0$ we have that $\mathcal{RD}^\phi(X(\cdot,\omega_i),t) = -\text{dist}^\phi(X,t)$. In this case, $|\mathcal{RD}^\phi(X(\cdot,\omega_i),t)|$ indicates the robustness by which $X(\cdot,\omega_i)$ satisfies $\neg\phi$, while we are interested in the opposite.

Unfortunately, the risk of not satisfying ϕ robustly, i.e., $R(-\text{dist}^{\neg\phi}(X,t))$, can in most of the cases not be computed. Recall that this was similarly the case for the robustness degree in Definition 1. Instead, we will focus on $R(-\rho^\phi(X,t))$ as an approximate risk of not satisfying ϕ robustly. We next show that this approximation has some desirable properties.

2) Approximating the risk of not satisfying ϕ robustly: A desirable property, which we will show to hold, is that $R(-\rho^\phi(X,t))$ over-approximates $R(-\text{dist}^{\neg\phi}(X,t))$ so that $R(-\rho^\phi(X,t))$ is more risk-aware than $R(-\text{dist}^{\neg\phi}(X,t))$, i.e., that it holds that $R(-\text{dist}^{\neg\phi}(X,t)) \leq R(-\rho^\phi(X,t))$.

Theorem 3: Let R be a monotone risk measure. Then it holds that $R(-\text{dist}^{\neg\phi}(X,t)) \leq R(-\rho^\phi(X,t))$.

This indeed allows to use $R(-\rho^\phi(X,t))$ instead of $R(-\text{dist}^{\neg\phi}(X,t))$. In the next section, we elaborate on sampling-based methods to calculate $R(-\rho^\phi(X,t))$. For two stochastic processes X_1 and X_2, note that $R(-\rho^\phi(X_1,t)) \leq R(-\rho^\phi(X_2,t))$ means that X_1 has less risk than X_2 with respect to the specification ϕ.

Oftentimes, one may be interested in associating a monetary cost with $\text{dist}^{\neg\phi}(X,t)$ that reflects the severity of an event with low robustness. One may hence want to assign high costs to low robustness and low costs to high robustness. Let us define an increasing cost function $C: \mathbb{R} \to \mathbb{R}$ that reflects this preference.

Corollary 1: Let R be a monotone risk measure and C be an increasing cost function. Then it holds that $R(C(-\text{dist}^{\neg\phi}(X,t))) \leq R(C(-\rho^\phi(X,t)))$.

IV. Computing the STL Robustness Risk

There are two main challenges in computing the approximate risk of not satisfying ϕ robustly, i.e., in computing $R(-\rho^\phi(X,t))$. First, note that exact calculation of

$R(-\rho^\phi(X,t))$ requires knowledge of the CDF of $\rho^\phi(X,t)$ no matter what the choice of the risk measure R will be. However, the CDF of $\rho^\phi(X,t)$ is not known (only the CDF of X is known) and deriving the CDF of $\rho^\phi(X,t)$ is in many cases not possible. Second, calculating $R(-\rho^\phi(X,t))$ may often involve solving high dimensional integrals for which in most of the cases no closed-form expression exists.

In this paper, we present a sample average approximation $\overline{R}(-\rho^\phi(X,t))$ of $R(-\rho^\phi(X,t))$ for the value-at-risk and show that this approximation has the favorable property of being an upper bound to $R(-\rho^\phi(X,t))$, i.e., that it holds that $R(-\rho^\phi(X,t)) \leq \overline{R}(-\rho^\phi(X,t))$ with high probability.

Remark 4: We remark that our approach is especially compatible in a data-driven setting where realizations X^i of the stochastic process X are observed, e.g., obtained from experiments, and where not even the CDF of X is known.

A. Sample Average Approximation of the Value-at-Risk (VaR)

Let us first obtain a sample approximation of the value-at-risk (VaR) and define for convenience the random variable

$$Z := -\rho^\phi(X,t).$$

For a risk level of $\beta \in (0,1)$, the VaR of Z is given by

$$VaR_\beta(Z) := \inf\{\alpha \in \mathbb{R} | F_Z(\alpha) \geq \beta\}$$

where we recall that F_Z is the CDF of Z.

Let now $X^1,\ldots,X^N$ be N independent copies of X and define $Z^i := -\rho^\phi(X^i,t)$ that are hence N independent copies of Z, i.e., all Z^i are independent and identically distributed. To estimate $F_Z(\alpha)$, define the empirical CDF

$$\overline{F}(\alpha, Z^1,\ldots,Z^N) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(Z^i \leq \alpha)$$

where $\mathbb{I}$ denotes the indicator function. Let us next select a finite set $A \subset \mathbb{R}$ which can, for now, be an arbitrary set. For a risk level β, we calculate an upper bound of $VaR_\beta(Z)$ as

$$\begin{aligned}\overline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta) := \min\Big\{&\alpha \in A| \\ &\overline{F}(\alpha, Z^1,\ldots,Z^N) - \sqrt{\frac{\log(|A|/\delta)}{2N}} \geq \beta\Big\}.\end{aligned}$$

We next show that $\overline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)$ is an upper bound of $VaR_\beta(Z)$ with high probability.

Theorem 4: Let $\delta \in (0,1)$ be a probability threshold and $\beta \in (0,1)$ be a risk level. Let $\overline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)$ be based on a finite set $A \subset \mathbb{R}$ and on $Z^1,\ldots,Z^N$ that are N independent copies of Z. With probability of at least $1-\delta$, it holds that

$$VaR_\beta(Z) \leq \overline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta).$$

Theorem 4 now provides an upper bound for the risk $VaR_\beta(Z)$ as we desired. However, note that the bound $VaR_\beta(Z) \leq \overline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)$ may be conservative. We emphasize that the choice of A can be arbitrary in Theorem 4, while in general it affects the conservatism of the given bound.

To alleviate this issue, we will find a lower bound $\underline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)$ of $VaR_\beta(Z)$ under conditions on A to determine the tightness of the estimate obtained in Theorem 4. Towards obtaining a lower bound, we assume the following. First, assume that $VaR_\beta(Z) \in \mathcal{A}$ where $\mathcal{A} \subset \mathbb{R}$ is a compact set. Let now $A \subset \mathcal{A}$ be a finite set that is a subset of $\mathcal{A}$ and has the property of being a η-net of $\mathcal{A}$, i.e., for each $\alpha' \in \mathcal{A}$ there exists an $\alpha \in A$ such that $|\alpha' - \alpha| \leq \eta$. Second, we assume that $F_Z(\alpha)$ is locally Lipschitz continuous on the set $\mathcal{A}$ with Lipschitz constant L_Z.

Remark 5: It is not restrictive to assume that a compact set $\mathcal{A}$ entails $VaR_\beta(Z)$ as ρ^ϕ will be bounded in most of the applications. An η-net A of $\mathcal{A}$ can easily be constructed using gridding techniques. The Lipschitz constant L_Z is in general not known, but exists except for some pathological cases. We remark that we do not need knowledge of L_z and intend to show in the remainder what effect the choice of A has on the tightness of the bound in Theorem 4.

Let us now calculate a lower bound of $VaR_\beta(Z)$ as

$$\begin{aligned}\underline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta) := \min\{&\alpha \in A| \\ &\overline{F}(\alpha, Z^1,\ldots,Z^N) + \sqrt{\frac{\log(|A|/\delta)}{2N}} + L_Z\eta \geq \beta\}.\end{aligned}$$

Theorem 5: Let $\delta \in (0,1)$ be a probability threshold and $\beta \in (0,1)$ be the risk level. Let $VaR_\beta(Z) \in \mathcal{A}$ where $\mathcal{A}$ is a compact set. Let $\underline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)$ be based on 1) the finite set $A \subset \mathcal{A}$ that is an η-net of $\mathcal{A}$, and 2) the random variables $Z^1,\ldots,Z^N$ that are N independent copies of Z. Let $F_Z(\alpha)$ be locally Lipschitz continuous on $\mathcal{A}$ with Lipschitz constant L_Z. With probability of at least $1-\delta$, it holds that

$$\underline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta) - \eta \leq VaR_\beta(Z).$$

Theorem 5 hence motivates to select the set A in Theorem 4 to be an η-net. Theorem 5 shows further that a decreasing η, which corresponds to an increasing cardinality of A, results in a tighter estimate in Theorem 4.

By Theorems 3, 4, and 5 we have now a procedure to find a tight upper bound $\overline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)$ of $VaR_\beta(Z)$ with high probability. To summarize, we observe N realizations $X^1,\ldots,X^N$ of the stochastic process X. We then select $\delta, \beta \in (0,1)$ and the set A and are guaranteed that with probability of at least $1-\delta$

$$\begin{aligned}VaR_\beta(-\text{dist}^{\neg\phi}(X,t)) &\leq VaR_\beta(-\rho^\phi(X,t)) \\ &\leq \overline{VaR}_\beta(-\rho^\phi(X^1,t),\ldots,-\rho^\phi(X^N,t),A,\delta).\end{aligned}$$

Remark 6: Concentration inequalities for other risk measures than the value-at-risk can often be derived. For the expected value $E(-\rho^\phi(X,t))$, concentration inequalities for the sample average approximation of $E(-\rho^\phi(X,t))$ can be obtained by applying Hoeffding's inequality when $\rho^\phi(X,t)$ is bounded. For the conditional value-at-risk $CVaR(-\rho^\phi(X,t))$, concentration inequalities are presented in [36]–[39]. We plan to address this in future work.

V. Case Study

We continue with the case study presented in Example 1. Now, however, the environment is uncertain as the regions C and D in which humans operate are not exactly known. Let therefore c and d be Gaussian random vectors as

$$c \sim \mathcal{N}\Big(\begin{bmatrix}2\\3\end{bmatrix}, \begin{bmatrix}0.125 & 0\\ 0 & 0.125\end{bmatrix}\Big),$$
$$d \sim \mathcal{N}\Big(\begin{bmatrix}6\\4\end{bmatrix}, \begin{bmatrix}0.125 & 0\\ 0 & 0.125\end{bmatrix}\Big).$$

Consequently, the signals x_1-x_6 become stochastic processes denoted by X_1-X_6. Our goal is now to calculate

$$\overline{VaR}_\beta(-\rho^\phi(X_i^1,t),\ldots,-\rho^\phi(X_i^N,t),A,\delta)$$

for each $i \in \{1,\ldots,6\}$ to compare the risk between the six robot trajectories r_1-r_6. We set $\delta := 0.001$ and $N := 12500$. The set A is chosen to be a uniform grid of the set $\mathcal{A} := [-1,1]$ with step size 0.001. For different β, the resulting $\overline{VaR}_\beta$ are shown in the following table.

i \ β	0.9	0.925	0.95	0.975
1	0.341	0.366	0.394	0.543
2	0.163	0.184	0.226	0.364
3	-0.165	-0.149	-0.117	0.012
4	-0.249	-0.249	-0.249	-0.179
5	-0.25	-0.25	-0.25	-0.153
6	-0.249	-0.249	-0.249	-0.249

Across all β, the table indicates that trajectories r_1 and r_2 are not favorable in terms of the induced robustness risk. Trajectory r_3 is better compared to trajectories r_1 and r_2, but worse than r_4-r_6 in terms of the robustness risk of ϕ. For trajectories r_4-r_6, note that a β of 0.9, 0.925, and 0.95 provides the information that the trajectories have roughly the same robustness risk. However, once the risk level β is increased to 0.975, it becomes clear that r_6 is preferable over r_4 that is again preferable over r_5. This matches with what one would expect by closer inspection of Fig. 1.

VI. Conclusion

We defined the risk of a stochastic process not satisfying a signal temporal logic specification robustly which we referred to as the "robustness risk". We also presented an approximation of the robustness risk that is an upper bound of the robustness risk when the used risk measure is monotonic. For the case of the value-at-risk, we presented a sampling-based method to calculate the approximate robustness risk.

References

[1] O. Maler and D. Nickovic, "Monitoring temporal properties of continuous signals," in *Proc. Int. Conf. FORMATS FTRTFT*, Grenoble, France, September 2004, pp. 152–166.

[2] G. E. Fainekos and G. J. Pappas, "Robustness of temporal logic specifications for continuous-time signals," *Theoret. Comp. Science*, vol. 410, no. 42, pp. 4262–4291, 2009.

[3] A. Donzé and O. Maler, "Robust satisfaction of temporal logic over real-valued signals," in *Proc. Int. Conf. FORMATS*, Klosterneuburg, Austria, September 2010, pp. 92–106.

[4] N. Mehdipour, C. Vasile, and C. Belta, "Arithmetic-geometric mean robustness for control from signal temporal logic specifications," in *Proc. Am. Control Conf.*, Philadelphia, PA, July 2019, pp. 1690–1695.

[5] I. Haghighi, N. Mehdipour, E. Bartocci, and C. Belta, "Control from signal temporal logic specifications with smooth cumulative quantitative semantics," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 4361–4366.

[6] T. Akazaki and I. Hasuo, "Time robustness in mtl and expressivity in hybrid system falsification," in *International Conference on Computer Aided Verification*. Springer, 2015, pp. 356–374.

[7] P. Varnai and D. V. Dimarogonas, "On robustness metrics for learning stl tasks," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 5394–5399.

[8] C. Madsen, P. Vaidyanathan, S. Sadraddini, C.-I. Vasile, N. A. DeLateur, R. Weiss, D. Densmore, and C. Belta, "Metrics for signal temporal logic formulae," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 1542–1547.

[9] A. Rodionova, E. Bartocci, D. Nickovic, and R. Grosu, "Temporal logic as filtering," in *Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control*, 2016, pp. 11–20.

[10] M. Tiger and F. Heintz, "Incremental reasoning in probabilistic signal temporal logic," *International Journal of Approximate Reasoning*, vol. 119, pp. 325–352, 2020.

[11] J. Li, P. Nuzzo, A. Sangiovanni-Vincentelli, Y. Xi, and D. Li, "Stochastic contracts for cyber-physical system design under probabilistic requirements," in *Proceedings of the 15th ACM-IEEE International Conference on Formal Methods and Models for System Design*, 2017, pp. 5–14.

[12] P. Kyriakis, J. V. Deshmukh, and P. Bogdan, "Specification mining and robust design under uncertainty: A stochastic temporal logic approach," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 18, no. 5s, pp. 1–21, 2019.

[13] D. Sadigh and A. Kapoor, "Safe control under uncertainty with probabilistic signal temporal logic," in *Proceedings of Robotics: Science and Systems*, AnnArbor, Michigan, June 2016.

[14] S. Jha, V. Raman, D. Sadigh, and S. A. Seshia, "Safe autonomy under perception uncertainty using chance-constrained temporal logic," *Journal of Automated Reasoning*, vol. 60, no. 1, pp. 43–62, 2018.

[15] L. Lindemann, G. J. Pappas, and D. V. Dimarogonas, "Control barrier functions for nonholonomic systems under risk signal temporal logic specifications," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 1422–1428.

[16] S. Safaoui, L. Lindemann, D. V. Dimarogonas, I. Shames, and T. H. Summers, "Control design for risk-based signal temporal logic specifications," *IEEE Control Systems Letters*, vol. 4, no. 4, pp. 1000–1005, 2020.

[17] S. S. Farahani, R. Majumdar, V. S. Prabhu, and S. Soudjani, "Shrinking horizon model predictive control with signal temporal logic constraints under stochastic disturbances," *IEEE Transactions on Automatic Control*, 2018.

[18] S. Bharadwaj, R. Dimitrova, and U. Topcu, "Synthesis of surveillance strategies via belief abstraction," in *Proc. Conf. Decis. Control*, Miami, FL, Dec. 2018, pp. 4159–4166.

[19] C.-I. Vasile, K. Leahy, E. Cristofalo, A. Jones, M. Schwager, and C. Belta, "Control in belief space with temporal logic specifications," in *Proc. Conf. Decis. Control*, Las Vegas, NV, Dec. 2016, pp. 7419–7424.

[20] M. Lahijanian, S. B. Andersson, and C. Belta, "Formal verification and synthesis for discrete-time stochastic systems," *IEEE Transactions on Automatic Control*, vol. 60, no. 8, pp. 2031–2045, 2015.

[21] M. Guo and M. M. Zavlanos, "Probabilistic motion planning under temporal tasks and soft constraints," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4051–4066, 2018.

[22] E. Bartocci, L. Bortolussi, L. Nenzi, and G. Sanguinetti, "On the robustness of temporal properties for stochastic models," *arXiv preprint arXiv:1309.0866*, 2013.

[23] R. T. Rockafellar, S. Uryasev *et al.*, "Optimization of conditional value-at-risk," *Journal of risk*, vol. 2, pp. 21–42, 2000.

[24] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *Journal of banking & finance*, vol. 26, no. 7, pp. 1443–1471, 2002.

[25] A. Majumdar and M. Pavone, "How should a robot assess risk? towards an axiomatic theory of risk in robotics," in *Robotics Research*. Springer, 2020, pp. 75–84.

[26] S. Samuelson and I. Yang, "Safety-aware optimal control of stochastic systems using conditional value-at-risk," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 6285–6290.

[27] E. Hyeon, Y. Kim, and A. G. Stefanopoulou, "Fast risk-sensitive model predictive control for systems with time-series forecasting uncertainties," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 2515–2520.

[28] S. G. McGill, G. Rosman, T. Ort, A. Pierson, I. Gilitschenski, B. Araki, L. Fletcher, S. Karaman, D. Rus, and J. J. Leonard, "Probabilistic risk metrics for navigating occluded intersections," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4322–4329, 2019.

[29] S. Singh, Y. Chow, A. Majumdar, and M. Pavone, "A framework for time-consistent, risk-sensitive model predictive control: Theory and algorithms," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2905–2912, 2018.

[30] D. S. Kalogerias, L. F. Chamon, G. J. Pappas, and A. Ribeiro, "Better safe than sorry: Risk-aware nonlinear bayesian estimation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5480–5484.

[31] M. P. Chapman, J. Lacotte, A. Tamar, D. Lee, K. M. Smith, V. Cheng, J. F. Fisac, S. Jha, M. Pavone, and C. J. Tomlin, "A risk-sensitive finite-time reachability approach for safety of stochastic dynamic systems," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 2958–2963.

[32] M. Ahmadi, M. Ono, M. D. Ingham, R. M. Murray, and A. D. Ames, "Risk-averse planning under uncertainty," in *2020 American Control Conference (ACC)*, 2020, pp. 3305–3312.

[33] M. Schuurmans and P. Patrinos, "Learning-based distributionally robust model predictive control of markovian switching systems with guaranteed stability and recursive feasibility," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 4287–4292.

[34] J. R. Munkres and T.-A. F. Course, "Prentice hall," *Eaglewood Cliffs, New Jersey*, 1975.

[35] R. Durrett, *Probability: theory and examples*. Cambridge university press, 2019, vol. 49.

[36] S. P. Bhat and P. LA, "Concentration of risk measures: A wasserstein distance approach," *Advances in Neural Information Processing Systems*, vol. 32, pp. 11 762–11 771, 2019.

[37] D. B. Brown, "Large deviations bounds for estimating conditional value-at-risk," *Operations Research Letters*, vol. 35, no. 6, pp. 722–730, 2007.

[38] P. Thomas and E. Learned-Miller, "Concentration inequalities for conditional value at risk," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6225–6233.

[39] Z. Mhammedi, B. Guedj, and R. C. Williamson, "Pac-bayesian bound for the conditional value at risk," *arXiv preprint arXiv:2006.14763*, 2020.

[40] A. H. Guide, *Infinite dimensional analysis*. Springer, 2006.

[41] M. Hertneck, J. Köhler, S. Trimpe, and F. Allgöwer, "Learning an approximate model predictive controller with guarantees," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 543–548, 2018.

APPENDIX I
SEMANTICS AND ROBUST SEMANTICS OF STL

The semantics that are associated with an STL formula ϕ as defined in Section II-A are defined next.

Definition 3 (STL Semantics): For a signal $x : T \to \mathbb{R}^n$, the semantics $\beta^\phi(x,t)$ of an STL formula ϕ are inductively defined as

$$\begin{aligned}
\beta^\top(x,t) &:= \top,\\
\beta^\mu(x,t) &:= \begin{cases}\top & \text{if } x(t)\in O^\mu\\ \bot & \text{otherwise,}\end{cases}\\
\beta^{\neg\phi}(x,t) &:= \neg\beta^\phi(x,t),\\
\beta^{\phi'\wedge\phi''}(x,t) &:= \min(\beta^{\phi'}(x,t),\beta^{\phi''}(x,t)),\\
\beta^{\phi' U_I \phi''}(x,t) &:= \sup_{t''\in t\oplus(I\cap T)}\Big(\min\big(\beta^{\phi''}(x,t''),\\
&\qquad \inf_{t'\in(t,t'')\cap T}\beta^{\phi'}(x,t')\big)\Big),\\
\beta^{\phi' \underline{U}_I \phi''}(x,t) &:= \sup_{t''\in t\ominus(I\cap T)}\Big(\min\big(\beta^{\phi''}(x,t''),\\
&\qquad \inf_{t'\in(t,t'')\cap T}\beta^{\phi'}(x,t')\big)\Big).
\end{aligned}$$

The robust semantics that are associated with an STL formula ϕ as defined in Section II-A are defined next.

Definition 4 (STL Robust Semantics): For a signal $x : T \to \mathbb{R}^n$, the robust semantics $\rho^\phi(x,t)$ of an STL formula ϕ are inductively defined as

$$\begin{aligned}
\rho^\top(x,t) &:= \infty,\\
\rho^\mu(x,t) &:= \begin{cases}\mathrm{dist}^{\neg\mu}(x,t) & \text{if } x\in\mathcal{L}^\mu(t)\\ -\mathrm{dist}^{\mu}(x,t) & \text{otherwise,}\end{cases}\\
\rho^{\neg\phi}(x,t) &:= -\rho^\phi(x,t),\\
\rho^{\phi'\wedge\phi''}(x,t) &:= \min(\rho^{\phi'}(x,t),\rho^{\phi''}(x,t)),\\
\rho^{\phi' U_I \phi''}(x,t) &:= \sup_{t''\in t\oplus(I\cap T)}\Big(\min\big(\rho^{\phi''}(x,t''),\\
&\qquad \inf_{t'\in(t,t'')\cap T}\rho^{\phi'}(x,t')\big)\Big),\\
\rho^{\phi' \underline{U}_I \phi''}(x,t) &:= \sup_{t''\in t\ominus(I\cap T)}\Big(\min\big(\rho^{\phi''}(x,t''),\\
&\qquad \inf_{t'\in(t,t'')\cap T}\rho^{\phi'}(x,t')\big)\Big).
\end{aligned}$$

Finally, we provide the following remark concerning the definition of the robust semantics of predicates $\rho^\mu(x,t)$.

Remark 7: As shown in [2, Lemma 57], it holds that

$$\mathrm{dist}^\mu(x,t)=\bar{d}\big(x(t),\mathrm{cl}(O^\mu)\big):=\inf_{x'\in\mathrm{cl}(O^\mu)} d(x(t),x').$$

Consequently, $\rho^\mu(x,t)$ encodes the signed distance from the signal $x(t)$ to the set O^μ.

APPENDIX II
RANDOM VARIABLES

We can associate a probability space $(\mathbb{R}^n,\mathcal{B}^n,P_Z)$ with the random vector Z where, for Borel sets $B\in\mathcal{B}^n$, the probability measure $P_Z:\mathcal{B}^n\to[0,1]$ is defined as

$$P_Z(B):=P(Z^{-1}(B))$$

where $Z^{-1}(B):=\{\omega\in\Omega|Z(w)\in B\}$ is the inverse image of B under Z.[3] In particular P_Z now describes the *distribution* of Z. For vectors $z\in\mathbb{R}^n$, the *cumulative distribution function (CDF)* of Z is defined as

$$F_Z(z)=P_Z((-\infty,z_1]\times\ldots\times(-\infty,z_n])$$

where z_i is the ith element of z. When the CDF $F_Z(z)$ is absolutely continuous, i.e., when $F_Z(z)$ can be written as

$$F_Z(z)=\int_{-\infty}^{z} f_Z(z')\mathrm{d}z'$$

for some non-negative and Lebesgue measurable function $f_Z(z)$, then Z is a *continuous random vector* and $f_Z(z)$ is called the *probability density function (PDF)* of Z. When the CDF $F_Z(z)$, on the other hand, is discontinuous, then Z is a

[3] Measurability of Z ensures that, for $B\in\mathcal{B}$, $Z^{-1}(B)\in\mathcal{F}$ so that the probability measure P can be pushed through to obtain P_X.

discrete random vector[4] and $f_Z(z)$ is called the *probability mass function (PMF)* satisfying

$$F_Z(z) = \sum_{z' \leq z} f_Z(z').$$

The results that we present in this paper apply to both continuous and discrete random variables unless stated otherwise.[5]

Appendix III
Risk Measures

We next present some desireable properties that a risk measure may have. Let therefore $Z, Z' \in \mathfrak{F}(\Omega, \mathbb{R})$ be cost random variables. A risk measure is *coherent* if the following four properties are satisfied.

1. Monotonicity: If $Z(\omega) \leq Z'(\omega)$ for all $\omega \in \Omega$, it holds that $R(Z) \leq R(Z')$.

2. Translation Invariance: Let $c \in \mathbb{R}$. It holds that $R(Z + c) = R(Z) + c$.

3. Positive Homogeneity: Let $c \in \mathbb{R}_{\geq 0}$. It holds that $R(cZ) = R(Z)$.

4. Subadditivity: It holds that $R(Z + Z') \leq R(Z) + R(Z')$.

If the risk measure additionally satisfies the following two properties, then it is called a distortion risk measure.

5. Comonotone Additivity: If $(Z(\omega) - Z(\omega'))(Z'(\omega) - Z'(\omega')) \geq 0$ for all $\omega, \omega' \in \Omega$ (namely, Z and Z' are commotone), it holds that $R(Z + Z') = R(Z) + R(Z')$.

6. Law Invariance: If Z and Z' are identically distributed, then $R(Z) = R(Z')$.

Common examples of popular risk measures are the expected value $\text{E}(Z)$ (risk neutral) and the worst-case $\text{ess}\sup_{\omega \in \Omega} Z(\omega)$ as well as:

- Mean-Variance: $\text{E}(Z) + \lambda \text{Var}(Z)$ where $c > 0$.
- Value at Risk (VaR) at level $\beta \in (0, 1)$: $VaR_\beta(Z) := \inf\{\alpha \in \mathbb{R} | F_Z(\alpha) \geq \beta\}$.
- Conditional Value at Risk (CVaR) at level $\beta \in (0, 1)$: $CVaR_\beta(Z) := E(Z | Z > VaR_\beta(Z))$.

Many risk measures are not coherent and can lead to a misjudgement of risk, e.g., the mean-variance is not monotone and the value at risk (which is closely related to chance constraints as often used in optimization) does not satisfy the subadditivity property.

Appendix IV
Proof of Theorem 1

Let us define the power set of $\mathbb{B}$ as $2^{\mathbb{B}} := \{\emptyset, \top, \bot, \{\bot, \top\}\}$. Note that $2^{\mathbb{B}}$ is a σ-algebra of $\mathbb{B}$. To prove measurability of $\beta^\phi(X(\cdot, \omega), t)$ in ω for a fixed $t \in T$, we need to show that, for each $B \in 2^{\mathbb{B}}$, it holds that the inverse image of B under $\beta^\phi(X(\cdot, \omega), t)$ for a fixed $t \in T$ is contained within $\mathcal{F}$, i.e., that it holds that $\{\omega \in \Omega | \beta^\phi(X(\cdot, \omega), t) \in B\} \subseteq \mathcal{F}$. We show measurability of $\beta^\phi(X(\cdot, \omega), t)$ in ω for a fixed $t \in T$ inductively on the structure of ϕ.

$\top$: For $B \in 2^{\mathbb{B}}$, it trivially holds that $\{\omega \in \Omega | \beta^\top(X(\cdot, \omega), t) \in B\} \subseteq \mathcal{F}$ since $\beta^\top(X(\cdot, \omega), t) = \top$ for all $\omega \in \Omega$ according to Definition 3 so that, depending on B, either $\{\omega \in \Omega | \beta^\top(X(\cdot, \omega), t) \in B\} = \emptyset$ or $\{\omega \in \Omega | \beta^\top(X(\cdot, \omega), t) \in B\} = \mathcal{F}$.

μ: Note first that O^μ is measurable and that the indicator function of a measurable set is measurable again (see e.g., [35, Chapter 1.2]). In particular, let $1_{O^\mu} : \mathbb{R}^n \to \mathbb{B}$ be the indicator function of O^μ with $1_{O^\mu}(\zeta) := \top$ if $\zeta \in O^\mu$ and $1_{O^\mu}(\zeta) := \bot$ otherwise. Since $X(t, \omega)$ is measurable in ω for a fixed $t \in T$ by definition, it follows that $\beta^\mu(X(\cdot, \omega), t) = 1_{O^\mu}(X(t, \omega))$ is measurable in ω for a fixed $t \in T$ according to Definition 3. For $B \in 2^{\mathbb{B}}$, it then follows that $\{\omega \in \Omega | \beta^\mu(X(\cdot, \omega), t) \in B\} = \{\omega \in \Omega | 1_{O^\mu}(X(t, \omega)) \in B\} \subseteq \mathcal{F}$.

$\neg\phi$: By assumption, $\beta^\phi(X(\cdot, \omega), t)$ is measurable in ω for a fixed $t \in T$. Recall that $\mathcal{F}$ is a σ-algebra that is, by definition, closed under its complement so that, for $B \in 2^{\mathbb{B}}$, it holds that $\{\omega \in \Omega | \beta^{\neg\phi}(X(\cdot, \omega), t) \in B\} = \Omega \setminus \{\omega \in \Omega | \beta^\phi(X(\cdot, \omega), t) \in B\} \subseteq \mathcal{F}$.

$\phi' \wedge \phi''$: By assumption, $\beta^{\phi'}(X(\cdot, \omega), t)$ and $\beta^{\phi''}(X(\cdot, \omega), t)$ are measurable in ω for a fixed $t \in T$. Hence $\beta^{\phi' \wedge \phi''}(X(\cdot, \omega), t) = \min(\beta^{\phi'}(X(\cdot, \omega), t), \beta^{\phi''}(X(\cdot, \omega), t))$ is measurable in ω for a fixed $t \in T$ since the min operator of measurable functions is again a measurable function.

$\phi' U_I \phi''$ and $\phi' \underline{U}_I \phi''$: Recall that $\beta^{\phi' U_I \phi''}(X(\cdot, \omega), t) := \sup_{t'' \in t \oplus (I \cap T)} \big(\min(\beta^{\phi''}(X(\cdot, \omega), t''), \inf_{t' \in (t, t'') \cap T} \beta^{\phi'}(X(\cdot, \omega), t'))\big)$. By assumption, $\beta^{\phi'}(X(\cdot, \omega), t)$ and $\beta^{\phi''}(X(\cdot, \omega), t)$ are measurable in ω for a fixed $t \in T$. First note that $(t, t'') \cap T$ and $I \cap T$ are countable sets since $T = \mathbb{N}$. According to [40, Theorem 4.27], the supremum and infimum operators over a countable number of measurable functions is again measurable. Consequently, the function $\beta^{\phi' U_I \phi''}(X(\cdot, \omega), t)$ is measurable in ω for a fixed $t \in T$. The same reasoning applies to $\beta^{\phi' \underline{U}_I \phi''}(X(\cdot, \omega), t)$.

The proof for $\rho^\phi(X(\cdot, \omega), t)$ follows again inductively on the structure of ϕ and the goal is to show that $\{\omega \in \Omega | \rho^\phi(X(\cdot, \omega), t) \in B\} \subseteq \mathcal{F}$ for each Borel set $B \in \mathcal{B}^n$. The difference to the proof of $\beta^\phi(X(\cdot, \omega), t)$ above is only in the way predicates μ are handled. Note first, as stated in Remark 7, that $\text{dist}^\mu(X(\cdot, \omega), t) = \bar{d}(X(t, \omega), \text{cl}(O^\mu))$ and $\text{dist}^{\neg\mu}(X(\cdot, \omega), t) = \bar{d}(X(t, \omega), \text{cl}(O^{\neg\mu}))$ so that

$$\begin{aligned}\rho^\mu(X(\cdot, \omega), t) = 0.5(1_{O^\mu}(X(t, \omega)) + 1)\bar{d}(X(t, \omega), \text{cl}(O^{\neg\mu})) \\ + 0.5(1_{O^\mu}(X(t, \omega)) - 1)\bar{d}(X(t, \omega), \text{cl}(O^\mu)).\end{aligned} \tag{5}$$

Since the indicator function $1_{O^\mu}(X(t, \omega))$ is measurable in ω for a fixed $t \in T$ and since function composition preserves measurability, we only need to show that $\bar{d}(X(t, \omega), \text{cl}(O^\mu))$ and $\bar{d}(X(t, \omega), \text{cl}(O^{\neg\mu}))$ are measurable in ω for a fixed $t \in T$. This immediately follows since $X(t, \omega)$ is measurable in ω for a fixed $t \in T$ by definition and since the function $\bar{d}$ is continuous in its first argument, and hence measurable, due to d being a metric defined on the set $\mathbb{R}^n$ (see e.g., [34, Chapter 3]).

[4] We disregard mixed random variables, i.e., random variables with continuous and discrete parts, for simplicity and without loss of generality.

[5] Note that there exists CDFs that are neither absolutely continuous nor discontinuous, e.g., the Cantor distribution, and have no PDF and no PMF. Our results similarly apply to these distributions unless stated otherwise.

APPENDIX V
PROOF OF THEOREM 2

For $\mathrm{dist}^\phi(X(\cdot,\omega),t)$, recall first that $\mathrm{dist}^\phi(X(\cdot,\omega),t) = \bar{\kappa}\big(X(\cdot,\omega),\mathrm{cl}(\mathcal{L}^\phi(t))\big) := \inf_{x^*\in \mathrm{cl}(\mathcal{L}^\phi(t))} \kappa(X(\cdot,\omega),x^*)$ and that κ is a metric defined on the set $\mathfrak{F}(T,\mathbb{R}^n)$ as argued in [2]. Therefore, it follows that the function $\bar{\kappa}$ is continuous in its first argument, and hence measurable (see e.g., [34, Chapter 3]). Consequently, the function $\mathrm{dist}^\phi(X(\cdot,\omega),t)$, which is a mapping $\mathrm{dist}^\phi : \mathfrak{F}(T,\mathbb{R}^n)\times T \to \mathbb{R}^n$, is continuous and hence measurable in its first argument $X(\cdot,\omega)$ for a fixed $t\in T$.[6] Note next that the stochastic process $X(\cdot,\omega):\omega\to \mathfrak{F}(T,\mathbb{R}^n)$ is measurable in ω. Since function composition preserves measurability, it holds that $\mathrm{dist}^\phi(X(\cdot,\omega),t)$ is measurable in ω for a fixed $t\in T$.

For $\mathcal{RD}^\phi(X(\cdot,\omega),t)$, note that we can write

$$\begin{aligned}\mathcal{RD}^\phi(X(\cdot,\omega),t) = 0.5(1_{\mathcal{L}^\phi(t)}(X(\cdot,\omega))+1)\mathrm{dist}^{\neg\phi}(X(\cdot,\omega),t)\\ + 0.5(1_{\mathcal{L}^\phi(t)}(X(\cdot,\omega))-1)\mathrm{dist}^{\phi}(X(\cdot,\omega),t)\end{aligned}$$

similarly to (5). Note next that $\mathcal{L}^\phi(t)$ is a measurable set since $\beta^\phi(x,t)$ is measurable in x for a fixed $t\in T$ as argued before. Due to the functions $1_{\mathcal{L}^\phi(t)}(X(\cdot,\omega))$ and $\mathrm{dist}^\phi(X(\cdot,\omega),t)$ being measurable in ω for a fixed $t\in T$, it follows using similar arguments as in the proof of Theorem 1 that $\mathcal{RD}^\phi(X(\cdot,\omega),t)$ is measurable in ω for a fixed $t\in T$.

APPENDIX VI
PROOF OF THEOREM 3

First note that $\rho^\phi(X(\cdot,\omega),t)\le \mathrm{dist}^{\neg\phi}(X(\cdot,\omega),t)$ for each realization $X(\cdot,\omega)$ of the stochastic process X with $\omega\in\Omega$ due to (3). Consequently, we have that $-\mathrm{dist}^{\neg\phi}(X(\cdot,\omega),t)\le -\rho^\phi(X(\cdot,\omega),t)$ for all $\omega\in\Omega$. If R is now monotonic, it directly follows that $R(-\mathrm{dist}^{\neg\phi}(X,t))\le R(-\rho^\phi(X,t))$.

APPENDIX VII
PROOF OF THEOREM 4:

We first state the next Lemma, which is inspired by [41] and proven in Appendix VIII, to then prove Theorem 4.

Lemma 1: Let $\delta\in(0,1)$ be a probability threshold and let $A\subset\mathbb{R}$ be a finite set. Let $\overline{F}_Z(\alpha,Z^1,\ldots,Z^N)$ be based on $Z^1,\ldots,Z^N$ that are N independent copies of Z. With probability of at least $1-\delta/|A|$, it holds that

$$|\overline{F}(\alpha,Z^1,\ldots,Z^N)-F_Z(\alpha)|\le\sqrt{\frac{\log(|A|/\delta)}{2N}} \tag{6}$$

From Lemma 1, we can observe that the empirical CDF $\overline{F}(\alpha,Z^1,\ldots,Z^N)$ approaches $F_Z(\alpha)$ as N increases or as $|A|/\delta$ decreases.

We next derive a lower bound for the probability that (6) holds for all $\alpha\in A$, i.e., we find a lower bound for

$$P\Big(\bigcap_{\alpha\in A}|\overline{F}(\alpha,Z^1,\ldots,Z^N)-F_Z(\alpha)|\le\sqrt{\frac{\log(|A|/\delta)}{2N}}\Big)$$

Towards this goal, first note that

$$\begin{aligned}&P\Big(\bigcup_{\alpha\in A}|\overline{F}(\alpha,Z^1,\ldots,Z^N)-F_Z(\alpha)|>\sqrt{\frac{\log(|A|/\delta)}{2N}}\Big)\\ &\overset{(a)}{\le}\sum_{\alpha\in A}P\Big(|\overline{F}(\alpha,Z^1,\ldots,Z^N)-F_Z(\alpha)|>\sqrt{\frac{\log(|A|/\delta)}{2N}}\Big)\\ &\overset{(b)}{\le}\sum_{\alpha\in A}\delta/|A|\le\delta\end{aligned}$$

where (a) follows by Boole's inequality and (b) follows by Lemma 1 and the negation property of probabilities. From this chain of inequalities, we get

$$\begin{aligned}P\Big(\bigcap_{\alpha\in A}|\overline{F}(\alpha,Z^1,\ldots,Z^N)-F_Z(\alpha)|\\ \le\sqrt{\frac{\log(|A|/\delta)}{2N}}\Big)\ge 1-\delta\end{aligned} \tag{7}$$

by again applying the negation property of probabilities.[7]

With a probability of at least $1-\delta$, it now holds that

$$\begin{aligned}\{\alpha\in A|\overline{F}(\alpha,Z^1,\ldots,Z^N)-\sqrt{\frac{\log(|A|/\delta)}{2N}}\ge\beta\}\\ \subseteq\{\alpha\in A|F_Z(\alpha)\ge\beta\}.\end{aligned}$$

Consequently, it holds with a probability of at least $1-\delta$

$$\begin{aligned}\overline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)=\inf\{\alpha\in A|\overline{F}(\alpha,Z^1,\ldots,Z^N)\\ -\sqrt{\frac{\log(|A|/\delta)}{2N}}\ge\beta\}\ge\inf\{\alpha\in A|F_Z(\alpha)\ge\beta\}.\end{aligned}$$

Note finally that $\inf\{\alpha\in A|F_Z(\alpha)\ge\beta\}\ge\inf\{\alpha\in\mathbb{R}|F_Z(\alpha)\ge\beta\}=VaR_\beta(Z)$ since $\{\alpha\in A|F_Z(\alpha)\ge\beta\}\subseteq\{\alpha\in\mathbb{R}|F_Z(\alpha)\ge\beta\}$ so that $VaR_\beta(Z)\le\overline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)$.

APPENDIX VIII
PROOF OF LEMMA 1

Note first that $\overline{F}_Z(\alpha,Z^1,\ldots,Z^N)$ is a random variable with the expected value according to

$$\begin{aligned}E(\overline{F}(\alpha,Z^1,\ldots,Z^N))&=\frac{1}{N}\sum_{i=1}^N E(\mathbb{I}(Z^i\le\alpha))\\ &=E(\mathbb{I}(Z\le\alpha))=P(Z\le\alpha)=F_Z(\alpha).\end{aligned}$$

We can now apply Hoeffding's inequality and obtain the following two concentration inequalities

$$\begin{aligned}P(\overline{F}(\alpha,Z^1,\ldots,Z^N)-F_Z(\alpha)\ge t)\le\exp(-2Nt^2)\\ P(\overline{F}(\alpha,Z^1,\ldots,Z^N)-F_Z(\alpha)\le -t)\le\exp(-2Nt^2).\end{aligned}$$

These two inequalities can compactly be rewritten as

$$P(|\overline{F}(\alpha,Z^1,\ldots,Z^N)-F_Z(\alpha)|\ge t)\le\exp(-2Nt^2)$$

and which consequently imply that

$$P(|\overline{F}(\alpha,Z^1,\ldots,Z^N)-F_Z(\alpha)|\le t)\ge 1-\exp(-2Nt^2)$$

By setting $\delta/|A|:=\exp(-2Nt^2)$, it follows that (6) holds with a probability of at least $1-\delta/|A|$.

[6] A σ-algebra of $\mathfrak{F}(T,\mathbb{R}^n)$ can be defined as the product σ-algebra of $\mathcal{B}^n$.

[7] Note in particular that $P(A\cup B)\le\kappa$ is equivalent to $1-P(\neg(A\cup B))\le\kappa$ which can be written as $1-\kappa\le P(\neg(A\cup B))=P(\neg A\cap\neg B)$.

Appendix IX
Proof of Theorem 5:

Note that $\underline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)$ is equivalent to the leftmost point $\alpha \in A$ such that $\overline{F}(\alpha,Z^1,\ldots,Z^N)+\sqrt{\frac{\log(|A|/\delta)}{2N}}+L_Z\eta \geq \beta$. All $\alpha' \in A$ with $\alpha' < \underline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)$ are consequently such that $\overline{F}(\alpha,Z^1,\ldots,Z^N)+\sqrt{\frac{\log(|A|/\delta)}{2N}}+L_Z\eta < \beta$. Let us hence define the set of such α' as

$$A' := \{\alpha' \in A | \alpha' < \underline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)\}.$$

Note also that (7) again holds. For each $\alpha' \in A'$, it now holds with a probability of at least $1-\delta$ that

$$\begin{aligned}
\beta &> \overline{F}(\alpha',Z^1,\ldots,Z^N)+\sqrt{\frac{\log(|A|/\delta)}{2N}}+L_Z\eta \\
&\overset{(a)}{\geq} \overline{F}(\alpha',Z^1,\ldots,Z^N)+\sqrt{\frac{\log(|A|/\delta)}{2N}}+F_Z(\alpha'')-F_Z(\alpha') \\
&\overset{(b)}{\geq} F_Z(\alpha'')
\end{aligned}$$

for each α'' that is such that $|\alpha'-\alpha''| \leq \eta$. Note in particular that (a) follows due to F_Z being locally Lipschitz continuous on $\mathcal{A}$, while (b) follows from (7). Since A is an η-net of $\mathcal{A}$, it follows that $\beta > F_Z(\alpha'')$ for all $\alpha'' \in \cup_{\alpha' \in A'}\{\alpha'' \in \mathbb{R} | \alpha'-\alpha''| \leq \eta\}$. Note next that A splits $\mathcal{A}$ into intervals that are at most of length 2η since A is an η-net of $\mathcal{A}$. Consequently, it follows by the definition of $VaR_\beta(Z)$ that $\underline{VaR}_\beta(Z^1,\ldots,Z^N,A,\delta)-\eta \leq VaR_\beta$ with probability of at least $1-\delta$.