tu dortmund university

bci Fakultät Bio- und Chemieingenieurwesen

# Probabilistic validation

**Prof. Sergio Lucia**
**Chair of Process Automation Systems**
**Department of Biochemical and Chemical Engineering**

# It is fun to work with nice smart people

A lot of this work is thanks to:

Benjamin Karg

Moritz Heinlein

Teodoro Alamo

# Motivation

- The robust NMPC design problem is very complex

- We know how to design good approximate NMPC controllers
  - Multi-stage NMPC
  - Ellipsoidal-based NMPC
  - Robust NMPC with approximate reachable sets
  - Tube-based NMPC
  - Approximate robust NMPC via NNs

# Complexity of design vs. test

**Design:** Robust NMPC requires solving minmax optimization problems

- Exponential growth with horizon length and uncertainty dimension.

**Test:** Running simulations once an (approximate) controller exists is simple

- It may not scale exponentially

Can we derive conclusions by testing performance *N* times?

- To obtain some guarantees, does *N* scale exponentially with problem size?

# We will obtain an answer for this question

How many simulations *N* do I have to run with a given controller $\kappa$, so that I can say with a confidence of 99.9999 % that the constraints will not be violated with a probability larger than 98% the next time I apply the controller, provided that I only observe constraint violations in 4 of my simulations?

$$N = 1297$$

$$N \geq \frac{1}{\epsilon}\left(r - 1 + \ln\frac{1}{\delta} + \sqrt{2(r-1)\ln\frac{1}{\delta}}\right).$$

# Thank you for your attention!

# Some basics of probability theory

# Really, just some basics

- Binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

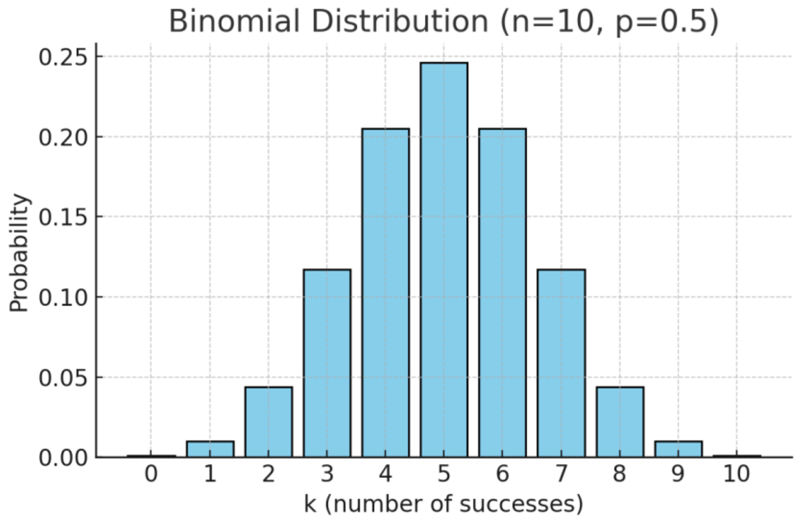
- Describes the number of distinct ways to choose *k* elements from a set of *n* elements ignoring the order

$$\binom{5}{2} = 10 \qquad \binom{n}{0} = \binom{n}{n} = 1$$

# Binomial distribution

- Models the number of successes in $n$ independent trials, each with probability $p$
- Example: Toss a coin $n = 10$ times $p = 0.5$
- Sampling with *replacement*
- Probabiltiy mass function:

$$\Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n$$



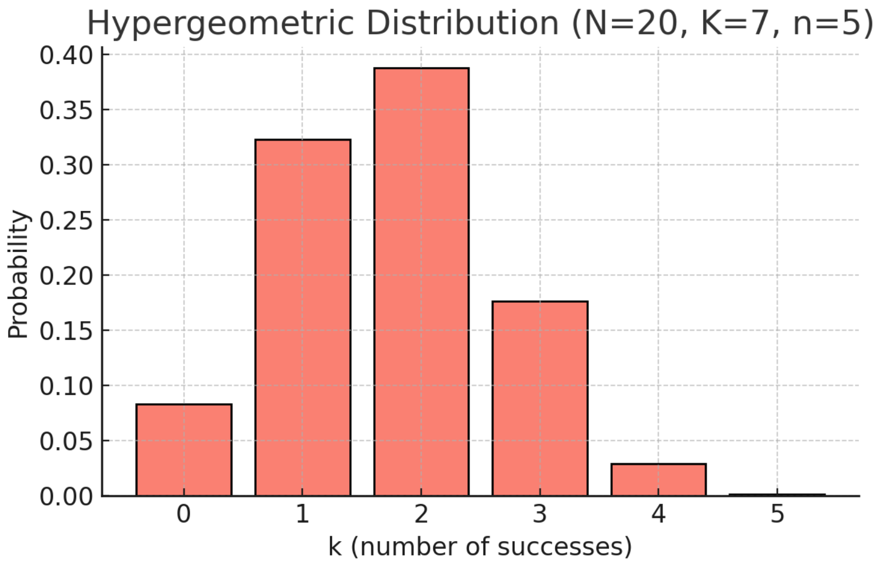

- Mean and variance

$$\mathbb{E}[X] = np, \quad \mathrm{Var}(X) = np(1 - p)$$

# Hypergeometric distribution

- Models the number of successes when drawing $n$ items from a finite population size $N$ with $K$ sucesses
- Example: From an urn with $N = 20$ balls, with $\mathrm{K} = 7$ red, draw $\mathrm{n} = 5$ balls one by one
- Sampling without *replacement*
- Probabiltiy mass function:

$$\Pr(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$



# Markov's inequality

If $X$ is a non-negative random variable, then for any $a > 0$

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

If you have a random variable that can only take **nonnegative values**, then the probability that it takes a **very large value** is at most its average divided by that large value.

**Intuition:** If the average is small, then the variable cannot be very large very often.

# Chebyshev’s inequality

For a random variable $X$ with mean $\mu$ and variance $\sigma$, for all $\varepsilon > 0$:

$$\Pr(|X - \mu| \geq \varepsilon) \ \leq \ \frac{\sigma^2}{\varepsilon^2}.$$

A random variable is **very unlikely to be far from its mean**, provided we know its variance

Intuition: if variance is small, the variable stays close to its mean most of the time.

# Back to probabilistic validation

# Performance function

We consider a performance function a $f(w;\theta) : \mathbb{W} \to \mathbb{R}$

- $\theta$ represents a given choice of design parameters
- $w$ represents an uncertainty that follows a known distribution $\mathcal{P}_{\mathcal{W}}$

**Main question:**

What is the probability that a certain performance level is achieved?

$$\mathrm{Pr}_{\mathcal{W}}\{f(w;\theta) \leq \gamma\}$$

# Examples of performance functions

For a closed-loop system affected by uncertainty : $w$

- Constraints are not violated for a simulation of 1000 steps
- Closed-loop cost is less than a given threshold for 1000 simulation steps

Easy to evaluate for a fixed $w^{(i)}$

▶ Can we draw conclusions from multiple evaluations?

# Probability estimation

**Naive approach:**

- Draw $N$ i.i.d samples of $f(w^{(i)}; \theta)$ and count successes (Bernoulli sample)
- Relative frequency:

$$\hat{P}_N = \frac{\text{count}_{i=1,\ldots,N}(f(w^{(i)}; \theta) \leq \gamma)}{N}$$

**How many samples** are needed to obtain a reliable estimate?

- Two-level probability: accuracy $\epsilon$ with confidence $1 - \delta$

$$\Pr_{\mathcal{W}^N}\{|\Pr_{\mathcal{W}}\{f(w;\theta) \leq \gamma\} - \hat{P}_N| \leq \epsilon\} \geq 1 - \delta$$

# Sample complexity: well known bounds

Chebyshev bound

$$N \geq \frac{1}{4\epsilon^2\delta}$$

Chernoff bound

$$N \geq \frac{\ln(2/\delta)}{2\epsilon^2}$$

- Independent of the number of uncertainties and distribution
- Assumes bounded random variables (e.g. [0,1])

# Chebyshev bound

- The true but unknown probability is $\Pr_{\mathcal{W}}\{f(w;\theta) \leq \gamma\} = p$
- Relative frequency

$$\hat{P}_N = \frac{\text{count}_{i=1,\ldots,N}(f(w^{(i)};\theta) \leq \gamma)}{N}$$

  The numerator is is a binomial dist. with *N* trials and probability *p*
- Therefore $\mathbb{E}\left[\hat{P}_N\right] = p$ and $\text{Var}\left[\hat{P}_N\right] = \frac{p(1-p)}{N}$ (see slide of binomial dist.)

We can now rewrite our desired probability

$$\Pr_{\mathcal{W}^N}\{|\Pr_{\mathcal{W}}\{f(w;\theta) \leq \gamma\} - \hat{P}_N| \leq \epsilon\} \geq 1 - \delta$$

as:

$$\Pr_{\mathcal{W}^N}\{|\mathbb{E}\left[\hat{P}_N\right] - \hat{P}_N| \leq \epsilon\} \geq 1 - \delta$$

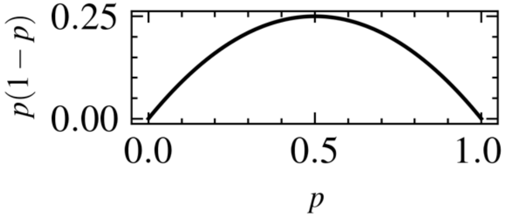
# Chebyshev bound (II)

- Recall Chebyschev's inequality: $\Pr(|X-\mu| \geq \varepsilon) \ \leq \ \frac{\sigma^2}{\varepsilon^2}.$
- Change direction of inequalities in the probabilities

$$\Pr_{\mathcal{W}^N}\{|\Pr_{\mathcal{W}}\{f(w;\theta) \leq \gamma\} - \hat{P}_N| \leq \epsilon\} \geq 1-\delta$$

$$\downarrow$$

$$\Pr_{\mathcal{W}^N}\{|\Pr_{\mathcal{W}}\{f(w;\theta) \leq \gamma\} - \hat{P}_N| \geq \epsilon\} \leq \ \delta$$




- From previous slide we have: $\mathrm{Var}\left[\hat{P}_N\right] = \frac{p(1-p)}{N}$
- Using Chebyschev's inequality, we obtain:

$$p(1-p) \leq \frac{1}{4}$$

$$\Pr_{\mathcal{W}^N}\{|\Pr_{\mathcal{W}}\{f(w;\theta) \leq \gamma\} - \hat{P}_N| \geq \epsilon\} \leq \frac{\mathrm{Var}\left[\hat{P}_N\right]}{\epsilon^2} = \frac{p(1-p)}{N\epsilon^2} \leq \frac{1}{4N\epsilon^2} \leq \delta$$

$$\downarrow$$

$$N \geq \frac{1}{4\epsilon^2\delta}$$

# Sample complexity: well known bounds

Chebyshev bound

$$N \geq \frac{1}{4\epsilon^2\delta}$$

Chernoff bound

$$N \geq \frac{\ln(2/\delta)}{2\epsilon^2}$$

- Independent of the number of uncertainties and distribution
- Assumes bounded random variables (e.g. [0,1])
- But often leads to a large number of samples

$$\epsilon = \delta = 0.005 \quad \longrightarrow \quad N = 119,830$$

- Fortunately, it can be improved! [Tempo, Bai, Dabbene, SCL 1997]

# Well known bounds… and great improvements

Fortunately, Chernoff bound can be improved! [Tempo, Bai, Dabbene, 1997]

- Focus on the worst-case performance / empirical maximum

$$J_1^N\left(\mathbf{w} = w^{[1:N]}; \theta\right) = \max_{i=1,\ldots,N} f\left(w^{(i)}; \theta\right)$$

If the number of samples is such that then the following inequality holds:

$$N \geq \frac{1}{\epsilon} \ln \frac{1}{\delta}$$

$$\Pr_{\mathcal{W}^N}\left\{\Pr_{\mathcal{W}}\{f(w;\theta) > J_1^N(\mathbf{w};\theta)\} \leq \varepsilon\right\} \geq 1-\delta$$

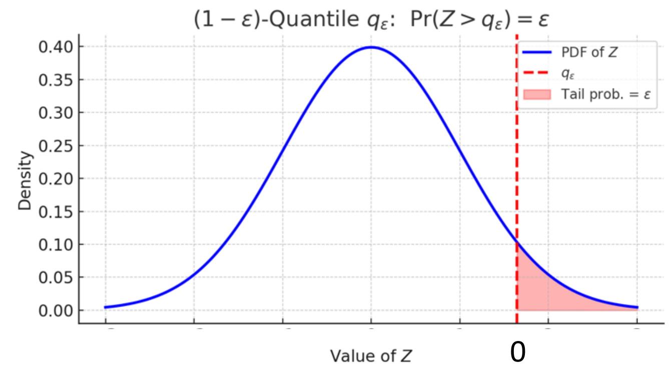
$$\epsilon = \delta = 0.005 \longrightarrow N \geq 1,058$$

Use this method to verify optimality of NN controllers using the dual

[Zhang et al., ACC 2019]

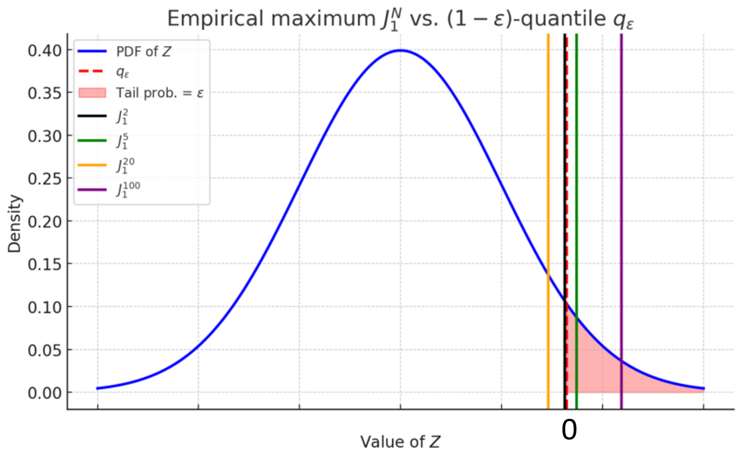
# Sketch of the derivation

- We are interested in the $(1-\varepsilon)$ quantile $q_\varepsilon$




- Consider the empirical maximum:

$$J_1^N\left(\mathbf{w} = w^{[1:N]};\theta\right) = \max_{i=1,\ldots,N} f\left(w^{(i)};\theta\right)$$




- Failure event: what is the probability that none of the N samples lands in the $\varepsilon$-tail?

$$\Pr\left(J_1^N < q_\varepsilon\right) = (1-\varepsilon)^N$$

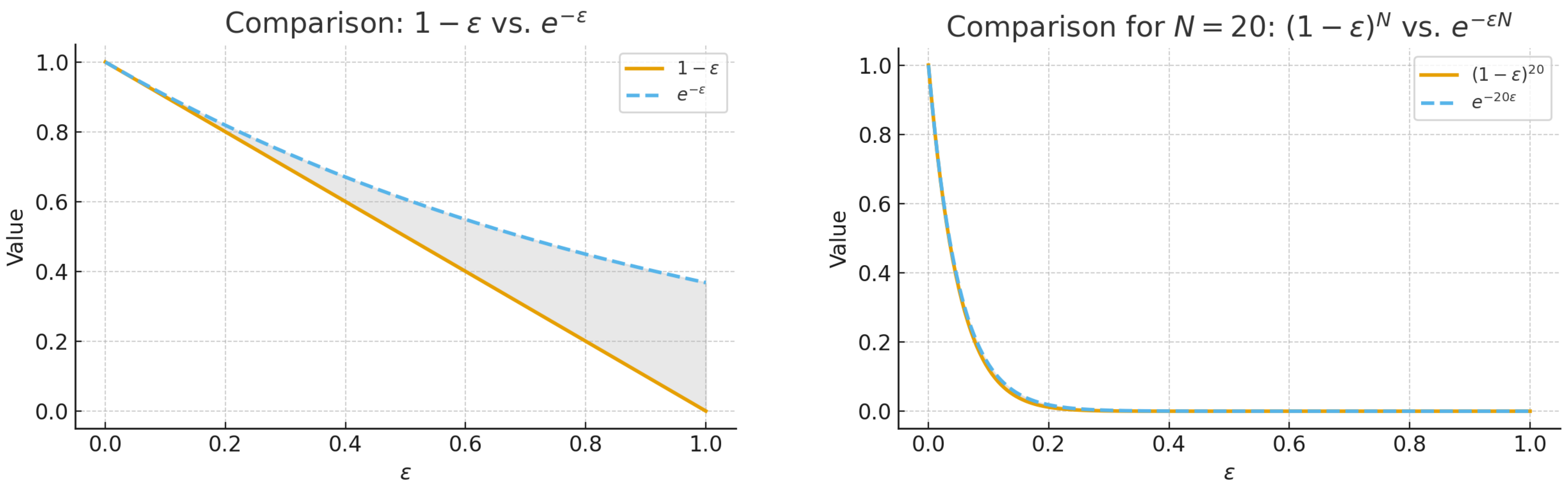
- Which can be bounded by:

$$\Pr\left(J_1^N < q_\varepsilon\right) = (1-\varepsilon)^N \;\leq\; e^{-\varepsilon N}.$$

# Sketch of the derivation (II)

- Why does this bound hold? $\Pr(J_1^N < q_\varepsilon) = (1-\varepsilon)^N \leq e^{-\varepsilon N}.$

0



# Sketch of the derivation (III)

- Given $\Pr\left(J_1^N < q_\varepsilon\right) = (1-\varepsilon)^N \ \leq \ e^{-\varepsilon N}.$

- Choose the number of samples $N$ such that it holds that

$$\Pr\left(J_1^N < q_\varepsilon\right) = (1-\varepsilon)^N \leq \delta$$

By taking logarithm on both sides of the inequality we get:

$$(1-\varepsilon)^N \ \leq \ e^{-\varepsilon N} \ \leq \ \delta \qquad \Rightarrow \qquad N \ \geq \ \frac{1}{\varepsilon} \ln \frac{1}{\delta}.$$

# Additional improvements

The previous bound has a problem for increasing $N$ and general unbounded uncertainty because

$$\lim_{N\to\infty} J_1^N(\boldsymbol{w};\theta) = \infty$$

- [Alamo et al., 2016, 2018] proposes the use of generalized maximum
  - (remove the r largest values and then take the maximum)

$$N \geq \frac{1}{\epsilon}\left(r - 1 + \ln\frac{1}{\delta} + \sqrt{2(r-1)\ln\frac{1}{\delta}}\right).$$

# Quick derivation (no details shown)

- The sum of the Bernoulli variables $Z_i$ follows a binomial distrib.

$$K = \sum_{i=1}^{N} \mathbf{1}\{Z_i > q_\varepsilon\} \sim \mathrm{Bin}(N, \varepsilon)$$

$$\Pr(J_r^N < q_\varepsilon) = \Pr(K \leq r-1) \leq \exp\left(-\frac{(N\varepsilon - (r-1))^2}{2N\varepsilon}\right) \leq \delta$$

- Solve for N:

$$N \geq \frac{1}{\epsilon}\left(r - 1 + \ln\frac{1}{\delta} + \sqrt{2(r-1)\ln\frac{1}{\delta}}\right).$$

# Probabilistic validation for approximate control

# Using prob. validation for approximate control

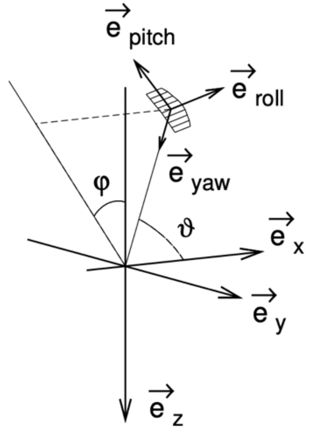
**We can use:**

- General performance functions
  - Closed-loop constraint violations for a finite-time simulation
  - Closed-loop cost for a finite-time simulation
  - Including detailed models for simulation or estimation errors
- Discard the $r$ largest values facilitate validation
- Obtain probabilistic validation statements
- (Adapt controller if the validation is not satisfactory)

# A towing kite

Probabilistically safe, embedded robust output-feedback NMPC

$$\dot{\theta}_{\text{kite}} = \frac{v_a}{L_T}\left(\cos\psi_{\text{kite}} - \frac{\tan\theta_{\text{kite}}}{E}\right),$$
$$\dot{\phi}_{\text{kite}} = -\frac{v_a}{L_T \sin\theta_{\text{kite}}}\sin\psi_{\text{kite}},$$
$$\dot{\psi}_{\text{kite}} = \frac{v_a}{L_T}\tilde{u} + \dot{\phi}_{\text{kite}}\cos\theta_{\text{kite}},$$

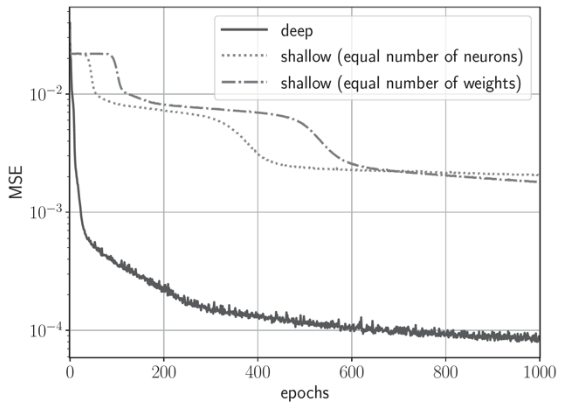


Erhard and Strauch, 2012

- Objective is to maximize thrust
- Two states can be measured, EKF to estimate
- Uncertain aerodynamic coefficients and wind parameters
- Minimum height constraint

# Controller design and training

- Robust NMPC with 8 scenarios for the uncertain parameters
- Add a backoff to the constraint:
  - Estimation errors
  - Approximation errors
  - Scenario errors
- Generate data (closed-loop simulations)
- Train a deep neural network




| | Symbol | Type | Values / Constraints | Units |
|---|---|---|---|---|
| kite model | $\theta_{\text{kite}}$ | State | $[0, \frac{\pi}{2}]$ | rad |
| | $\phi_{\text{kite}}$ | State | $[-\frac{\pi}{2}, \frac{\pi}{2}]$ | rad |
| | $\psi_{\text{kite}}$ | State | $[0, 2\pi]$ | rad |
| | $\tilde{u}$ | Control input | $[-10, 10]$ | N |
| | $\tilde{c}$ | Known parameter | 0.028 | - |
| | $\beta$ | Known parameter | 0 | rad |
| | $\rho$ | Known parameter | 1 | $\text{kg}\,\text{m}^{-3}$ |
| | $h_{\text{min}}$ | Known parameter | 100 | m |
| | $E_0$ | Uncertain parameter | unif(4, 6) | - |
| wind model | $p_v$ | State | - | s |
| | $k_{\sigma_v}$ | Known parameter | 0.14 | - |
| | $L_v$ | Known parameter | 100 | m |
| | $T_v$ | Known parameter | 0.15 | s |
| | $v_{\text{m}}$ | Uncertain parameter | unif(7, 9) | $\text{m}\,\text{s}^{-1}$ |
| | $w_{\text{tb}}$ | Uncertain parameter | normal(0, 0.25) | - |

# Probabilistic validation

Define the performance function (with backoff $\eta$)

$$\phi(w; N_{\text{sim}}, \kappa_{\text{dnn},\eta}) = \max_{j=0,\dots,N_{\text{sim}}} (h_{\min} - h(x(j,w))),$$

The controller is probabilistically safe if with probability $1-\delta$:

$$\Pr_{\mathcal{W}}(\phi(w; N_{\text{sim}}, \kappa_{\text{dnn},\eta}) > 0) \leq \epsilon,$$

We need the following number of samples:

$$\epsilon = 0.02, \delta = 1 \times 10^{-6}, r = 4 \longrightarrow N = 1388$$

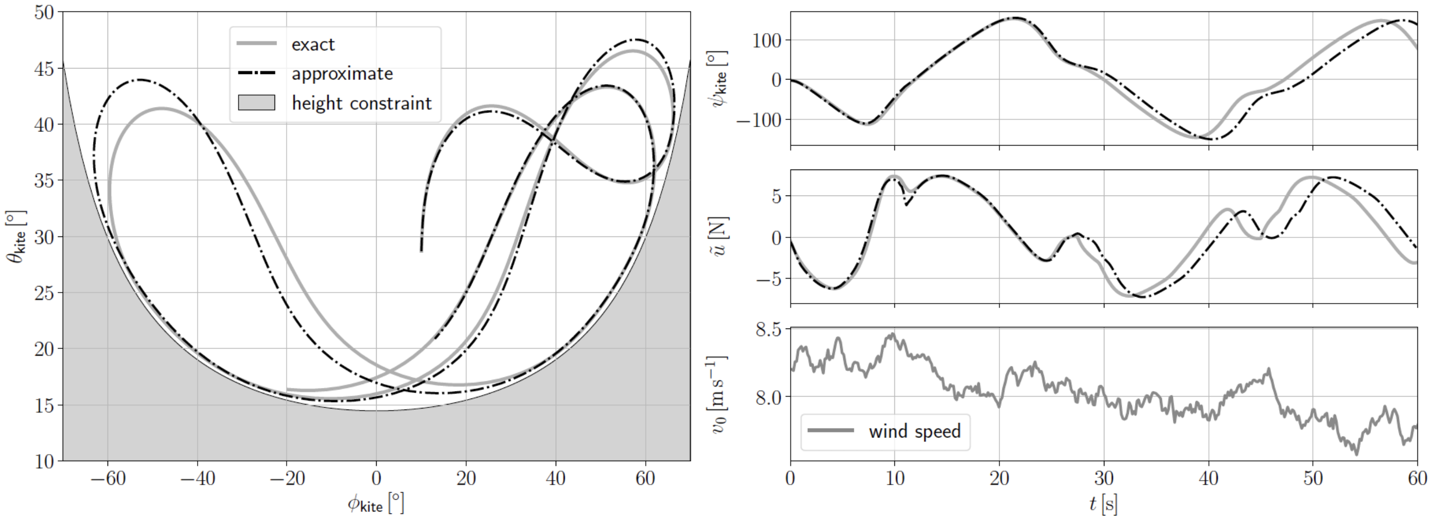
More details in:

B. Karg, T. Alamo, und S. Lucia, „Probabilistic performance validation of deep learning-based robust NMPC controllers“, *Int J Robust Nonlinear Control*, Bd. 31, Nr. 18, S. 8855–8876, 2021, doi: 10.1002/rnc.5696.

# Results

Embedded real-time implementation on an ARM-Cortex M3

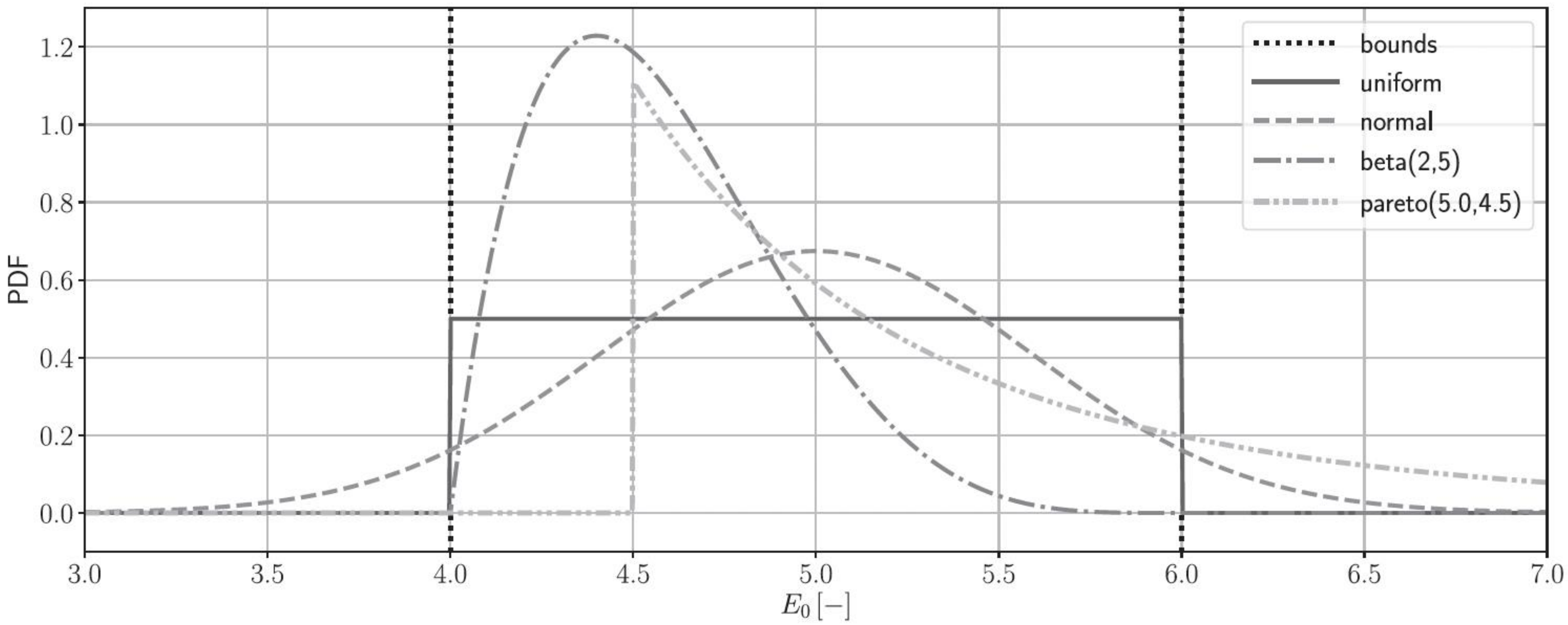
- 96 kB memory footprint, 32 ms running time for DNN and 28 ms for EKF




| controller | $\kappa_{\text{dnn},0}$ | $\kappa_{\text{dnn},2}$ | $\kappa_{\text{dnn},4}$ | $\kappa_{\text{dnn},6}$ |
|---|---|---|---|---|
| feasible trajectories | 660/1388 | 1380/1388 | 1385/1388 | 1387/1388 |
| $\psi(\mathbf{v}, 4)$ [m] | 1.682 | 0.273 | -0.316 | -1.818 |
| $T_{\text{F}}$ (avg.) [kN] | 227.516 | 225.997 | 224.185 | 222.179 |
| probabilistically safe | No | No | Yes | Yes |

# Test with other distributions

- If samples for validation are generated from a different distribution as the final validation, it still works in this case



# Other popular use of the same theory

- The scenario approach

# Motivation for scenario approach

- Probabilistic validation gives a posteriori guarantees for one performance metric

- But the same theoretical basis can be used for other problems

- Scenario approach gives probabilistic guarantees for sampled convex problems

▶ Largely developed by M. Campi and G. Calafiore

[Calafiore and Campi. The scenario approach to robust control design 2006, TAC]

# Application of the scenario approach to MPC

- How to solve a stochastic MPC with chance constraints?

$$
\begin{aligned}
& \min_{\boldsymbol{u}_{[0:N-1]|t}} \sum_{i=0}^{N-1} \mathbb{E}\left[\ell(x_{i|t}, u_{i|t})\right], \\
\text{s.t.} \quad & x_{i+1|t} = A(w_{t+i})x_{i|t} + B(w_{t+i})u_{i|t} + d(w_{t+i}), \\
& x_{0|t} = x_t \quad \forall i = 0, \ldots, N-1, \\
& \Pr_{w_{[t:t+N-1]} \sim \mathcal{W}^N}\left[x_{[i:i+N]|t} \notin \mathbb{X}\right] \leq \epsilon, \\
& u_{i|t} \in \mathbb{U} \quad \forall i = 0, \ldots, N-1.
\end{aligned}
$$

# Many applications in MPC

$$\text{SP}_N : \min_{\theta \in \Theta} c^T \theta$$

$$\text{subject to } \theta \in \cap_{i=1,\ldots,N} \Theta_{w^{(i)}}$$

- Extensions to multiple chance constraints (Schildbach et al., 2014)

$$\begin{aligned}
\min_{\boldsymbol{u}_{[0:N-1]|t}} \quad & \sum_{i=0}^{N-1} \mathbb{E}\big[\ell(x_{i|t}, u_{i|t})\big], \\
\text{s.t.} \quad & x_{i+1|t} = A(w_{t+i}) x_{i|t} + B(w_{t+i}) u_{i|t} + d(w_{t+i}), \\
& x_{0|t} = x_t \quad \forall i = 0, \ldots, N-1, \\
& \Pr_{w_{[t:t+N-1]} \sim \mathcal{W}^N} \big[x_{[i:i+N]|t} \notin \mathbb{X}\big] \leq \epsilon, \\
& u_{i|t} \in \mathbb{U} \quad \forall i = 0, \ldots, N-1.
\end{aligned}$$

$$\longrightarrow$$

$$\begin{aligned}
\min_{\boldsymbol{u}_{[0:N-1]|t}} \quad & \sum_{k=1}^{N_{\text{sample}}} \sum_{i=0}^{N-1} \ell(x_{i|t}, u_{i|t}), \\
\text{s.t.} \quad & x_{i+1|t}^{(k)} = A(w_{t+i}^{(k)}) x_{i|t}^{(k)} + B(w_{t+i}^{(k)}) u_{i|t} + d(w_{t+i}^{(k)}), \\
& x_{0|t}^{(k)} = x_t \quad \forall k = 1, \ldots, N_{\text{sample}}, \\
& x_{i|t}^{(k)} \in \mathbb{X} \quad \forall i = 0, \ldots, N, \forall k = 1, \ldots, N_{\text{sample}}, \\
& u_{i|t} \in \mathbb{U} \quad \forall i = 0, \ldots, N-1.
\end{aligned}$$

- Extensions to include feedback in the predictions

# Generalization in the scenario approach

**Theorem:** If the SP is convex and there exist a unique solution to the SP, then the following bound on the probability of violation holds:

$$\mathrm{Pr}_{\mathcal{W}^N}\{V(\theta^*) > \epsilon\} \leq \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$$

We want that the right-hand side is $\leq \delta$

Then the inequality holds if

$$N \geq \frac{2}{\epsilon}\left(\log\left(\frac{1}{\delta}\right) + d - 1\right)$$

# Main characteristics of the scenario approach

- Applicable to any uncertain convex program
- The number of scenarios increases with the number of optimization variables, but not with the dimension of the uncertainty
- The result is independent of the probability distribution. You just need **independent samples** from it (which can be experimental data)
- Provides a controller **design algorithm** and not just an a posteriori validation

# Some numbers

$$N \geq \frac{2}{\epsilon}\left(\log\left(\frac{1}{\delta}\right) + d - 1\right)$$

- $\delta = 10^{-3}, \epsilon = 10^{-2}, d = 10 \;\rightarrow N \geq 3182$
- $\delta = 10^{-10}, \epsilon = 10^{-2}, d = 10 \;\rightarrow N \geq 6406$
- $\delta = 10^{-10}, \epsilon = 10^{-2}, d = 100 \;\rightarrow N \geq 24405$
- $\delta = 10^{-10}, \epsilon = 10^{-2}, d = 1000 \;\rightarrow N \geq 204406$
- $\delta = 10^{-10}, \epsilon = 5 \cdot 10^{-2}, d = 1000 \;\rightarrow N \geq 40882$

# Summary

# Summary: Probabilistic validation

## Main Assumptions / Limitations

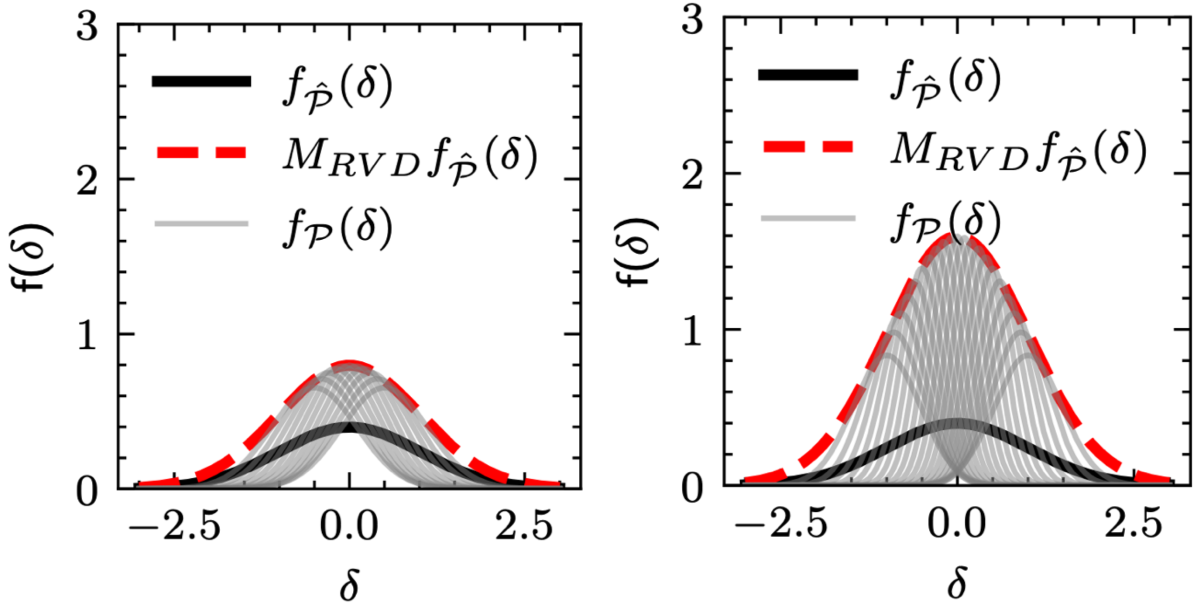
- Your performance indicator is a Bernoulli variable
- You can draw samples from the **real** probability distribution
- Number of samples grows fast for very small $\varepsilon$
- A controller is available

## Main strengths

- Sample complexity does not depend on the dimension of the uncertainty
- Can be used with detailed simulators
- Obtain probabilistic guarantees for very general cases

# Distributionally robustness

- Can we say something if we do not know the probability distribution exactly?




- In some situations, this results simply in an adapted risk $\varepsilon$

[Heinlein, Alamo and Lucia, CDC 2025. Available online: arXiv:2409.01177]

# Final discussion

- Are probabilistic guarantees in control valuable?
- Is all this theory really relevant to achieve robust control of complex systems?

# Complexity of design vs. test

**Design:** Robust NMPC requires solving minmax optimization problems

- Exponential growth with horizon length and uncertainty dimension.

**Test:** Run simulations once a (approximate) controller exists is simple

- It may not scale exponentially

Can we derive conclusions by testing performance *N* times?

- To obtain some guarantees, does N scale exponentially with problem size?

# Extra slides

# Why do these bounds hold?

Sketches of proofs and derivations

Largely based on
[T. Alamo, J. M. Manzano, and E. F. Camacho, „Robust Design Through Probabilistic Maximization“, in *Uncertainty in Complex Networked Systems: In Honor of Roberto Tempo*, T. Başar, Springer 2018]

# A quick dive into order statistics

- Define the generalized max function
  - Given vector $\boldsymbol{v} = \left[v^{(1)}, v^{(2)}, ..., v^{(N)}\right]^T \in \mathbb{R}^N$
  - Rearrange it into non-increasing order $\boldsymbol{v}_+ = \left[v_+^{(1)}, v_+^{(2)}, ..., v_+^{(N)}\right]^T \in \mathbb{R}^N$ with

$$v_+^{(1)} \geq v_+^{(2)} \geq ... \geq v_+^{(N-1)} \geq v_+^{(N)}$$

  - Then for integer $1 \leq r \leq N : \phi(\boldsymbol{v}, r) = v_+^{(r)}$
- For us, $\boldsymbol{v}$ corresponds to sampled observations of the performance metric

$$J_r^N(\boldsymbol{w}; \theta) = \phi\left(\begin{bmatrix} f(w^{(1)}; \theta) \\ ... \\ f(w^{(N)}; \theta) \end{bmatrix}, r\right)$$

- This ordering trick shifts the perspective away from the underlying uncertainty $\Pr_w$ towards the distribution resulting from the ordering

# Defining the probability of failure

- Probability of asymptotic failure
  - How likely does the $r$ worst value out of $N$ samples fail to bound the $\epsilon$-th quantile of the performance metric

$$\Pr_{\mathcal{W}^N}\{\Pr_{\mathcal{W}}\{f(w;\theta) > J_r^N(\boldsymbol{w};\theta)\} > \epsilon\}$$

- Probability of non-asymptotic failure
  - The $r$ worst value out of $N$ samples $\boldsymbol{w}_a \in \mathbb{R}^N$ fails to bound the $s$ worst value out of $M$ other samples $\boldsymbol{w}_b \in \mathbb{R}^M$

$$\Pr_{\mathcal{W}^{N+M}}\{J_r^N(\boldsymbol{w}_a;\theta) < J_s^M(\boldsymbol{w}_b;\theta)\}$$

  - Simple case: $M = s = 1$: How likely will the next sample be worse?

# Bounding the probability of non-asymptotic failure

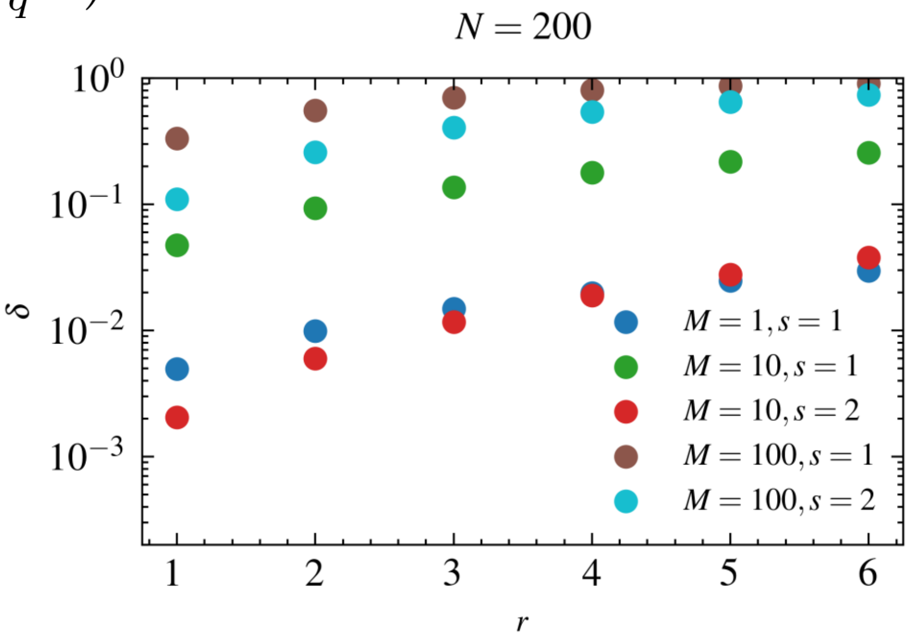

- Given $N, M, r, s$ with $1 \le r \le N$, $1 \le s \le M$, , :
- Follows hypergeometric distribution
- Independent on uncertainty and performance metric
- Proof based on drawing samples from an ordered vector


- Given $N, M, r, s$ with $1 \le r \le N$, $1 \le s \le M$, , :

$$\Pr_{\mathcal{W}^{N+M}}\{J_r^N(\boldsymbol{w}_a;\theta) < J_s^M(\boldsymbol{w}_b;\theta)\} \le \sum_{i=0}^{r-1} \frac{\binom{N}{i}\binom{M}{q-i}}{\binom{N+M}{q}}, \quad \text{with } q = r + s - 1$$

- Follows hypergeometric distribution
- Independent on uncertainty and performance metric
- Proof based on drawing samples from an ordered vector

$$\boldsymbol{v}_+ = \left[v_+^{(1)}, v_+^{(2)}, \ldots, v_+^{(N+M)}\right]^T \in \mathbb{R}^{N+M}$$



# Proof of the bound for non-asymptotic failure

- Proof: Ordered vector with $N+M$ samples: $\boldsymbol{v}_+ = \left[v_+^{(1)}, v_+^{(2)}, ..., v_+^{(N+M)}\right]^T \in \mathbb{R}^{N+M}$
  - Denote its first $q = r + s - 1$ components: $\boldsymbol{z} = \left[v_+^{(1)}, v_+^{(2)}, ..., v_+^{(q)}\right]^T \in \mathbb{R}^q$
  - $\boldsymbol{I} \in \mathcal{C}_N^{N+M}$ is a configuration of $N$ drawn samples of $N+M$ samples without replacement, $\boldsymbol{I}^C$ is the complement
  - Failure occurs, when less than $r$ components of $\boldsymbol{z}$ are from $\boldsymbol{I}$
  - Given $i$ with $0 \le i \le r-1$, the probability that $\boldsymbol{v}(\boldsymbol{I})$ has exactly $i$ components in $\boldsymbol{z}$ is given by: $$\frac{\binom{N}{i}\binom{M}{q-i}}{\binom{N+M}{q}}$$
  - Then sum up until $\mathrm{i} = r - 1$

$$\Pr_{\mathcal{C}_N^{N+M}}\{\phi(\boldsymbol{v}(\boldsymbol{I}), r) < \phi(\boldsymbol{v}(\boldsymbol{I}^C), s)\} \le \sum_{i=0}^{r-1} \frac{\binom{N}{i}\binom{M}{q-i}}{\binom{N+M}{q}}, \text{ with } q = r + s - 1$$

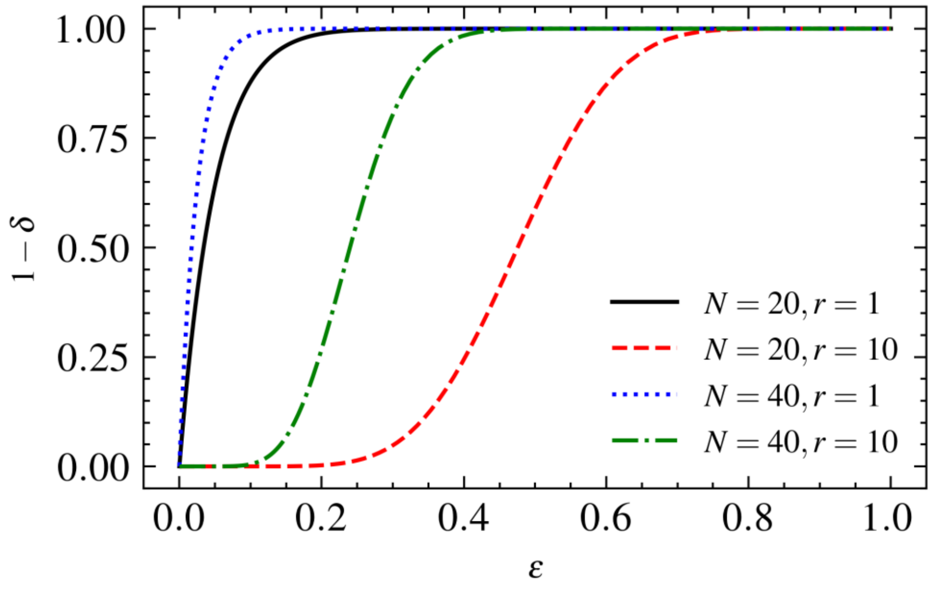
# Bounding the probability of asymptotic failure

- Given $r$ and $N$ with $1 \leq r \leq N$ and $\epsilon \in (0,1)$

$$\Pr_{\mathcal{W}^N}\{\Pr_{\mathcal{W}}\{f(w;\theta) > J_r^N(\boldsymbol{w};\theta)\} > \epsilon\} \leq \sum_{i=0}^{r-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$$

- Follows beta distribution
- Mean probability of failure given by

$$\mathbb{E}_{\mathcal{W}^N}\left[\Pr_{\mathcal{W}}\{f(w;\theta) > J_r^N(\boldsymbol{w};\theta)\}\right] \leq \frac{r}{N+1}$$



# Proof for the bound of asymptotic failure

$$\Pr_{\mathcal{W}^N}\{\Pr_{\mathcal{W}}\{f(w;\theta) > J_r^N(\boldsymbol{w};\theta)\} > \epsilon\} \leq \sum_{i=0}^{r-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$$

- Probability of non-asymptotic failure with $s(M) = \lceil \epsilon M \rceil$

$$\Pr_{\mathcal{W}^{N+M}}\{J_r^N(\boldsymbol{w}_a;\theta) < J_s^M(\boldsymbol{w}_b;\theta)\} \leq \sum_{i=0}^{r-1} \frac{\binom{N}{i}\binom{M}{q-i}}{\binom{N+M}{q}}, \text{ with } q = r+s-1$$

  - Asymptotic convergence of hypergeometric to binomial distribution

$$\Pr_{\mathcal{W}^N}\{\Pr_{\mathcal{W}}\{f(w;\theta) > J_r^N(\boldsymbol{w};\theta)\} > \epsilon\} \leq \lim_{M\to\infty} \Pr_{\mathcal{W}^{N+M}}\{J_r^N(\boldsymbol{w}_a;\theta) < J_{s(M)}^M(\boldsymbol{w}_b;\theta)\}$$

$$\leq \lim_{M\to\infty} \sum_{i=0}^{r-1} \frac{\binom{N}{i}\binom{M}{r+s(M)+1-i}}{\binom{N+M}{r+s(M)+1-i}} = \sum_{i=0}^{r-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$$

# Bounds on the sample size

- The finite sample guarantees are derived by bounds on the respective distributions

$$\Pr_{\mathcal{W}^N}\{\Pr_{\mathcal{W}}\{f(w;\theta) > J_r^N(\boldsymbol{w};\theta)\} > \epsilon\} \leq \sum_{i=0}^{r-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$$

$$N \geq \frac{1}{\epsilon}\left(r - 1 + \ln\frac{1}{\delta} + \sqrt{2(r-1)\ln\frac{1}{\delta}}\right)$$

r=1

$$N \geq \frac{1}{\epsilon}\ln\frac{1}{\delta}$$

- Probability of failure for the next sample:

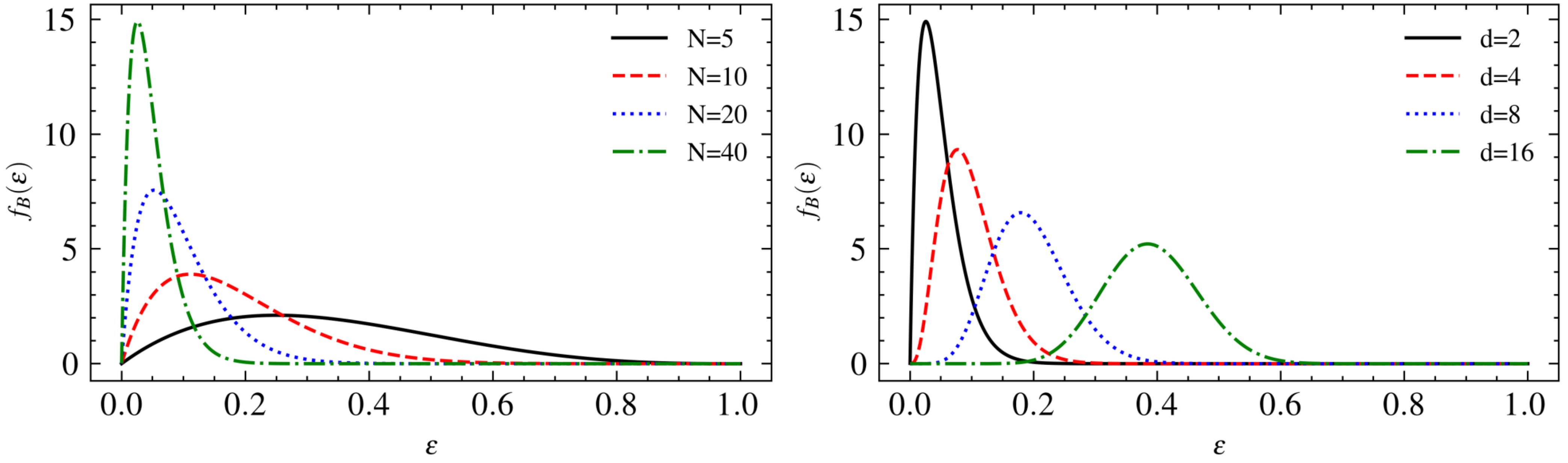
$$\Pr_{\mathcal{W}^N}\{\Pr_{\mathcal{W}}\{f(w;\theta) > J_r^N(\boldsymbol{w};\theta)\}\} \leq \frac{1}{N+1}$$

# Density of the right hand side for several param.

- Area under the curve corresponds to confidence $1-\delta$

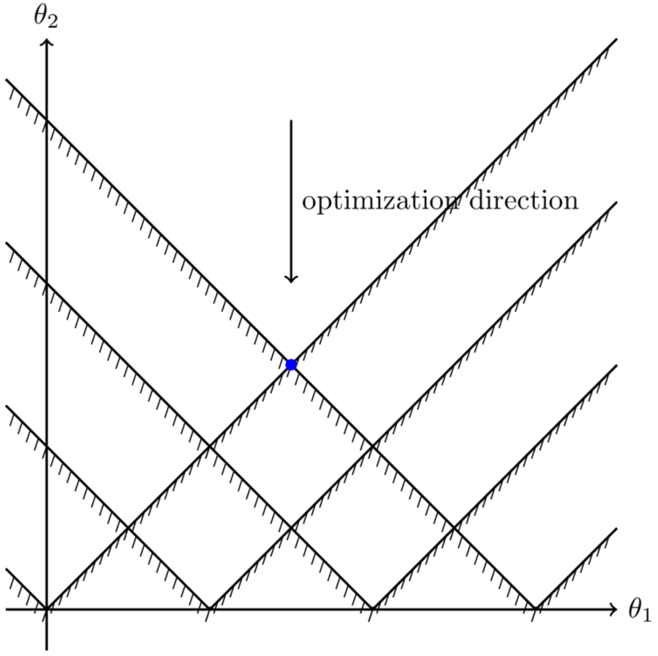
For $d = 2$

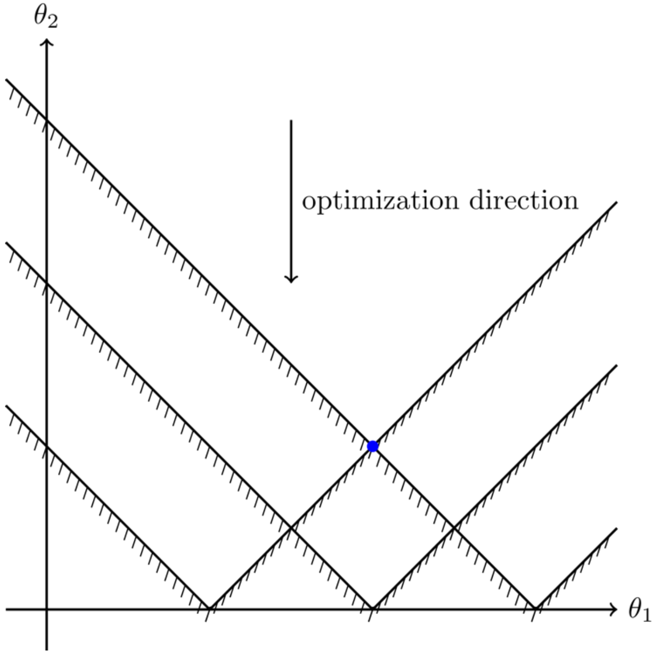
For $N = 40$



# Proof: support constraints

- A constraint is a support constraint if its removal improces the solution of the program
- A convex program with $d$ degrees of freedom has at most $d$ support constraints



# Idea behind the proof

- Assume we have a 2D problem with uniform uncertainty

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \mathbf{1}^\top \boldsymbol{x}$$

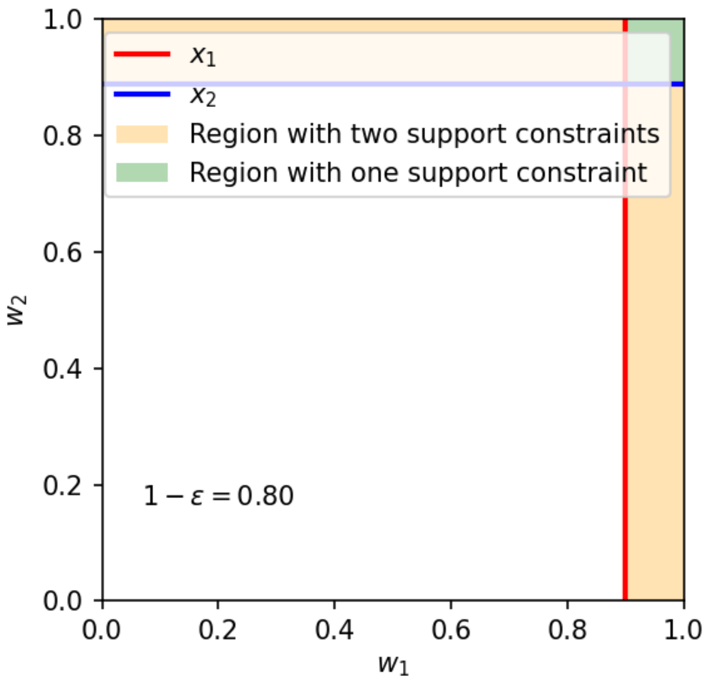
$$\text{subject to:} \quad \mathbb{P}_{\boldsymbol{w}\sim\mathcal{U}_{[0,1]^2}}\{\max(x_1 - \boldsymbol{w}_1, 0) + \max(x_2 - \boldsymbol{w}_2, 0) \geq 0\} \leq \epsilon$$

- Scenario problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \mathbf{1}^\top \boldsymbol{x}$$

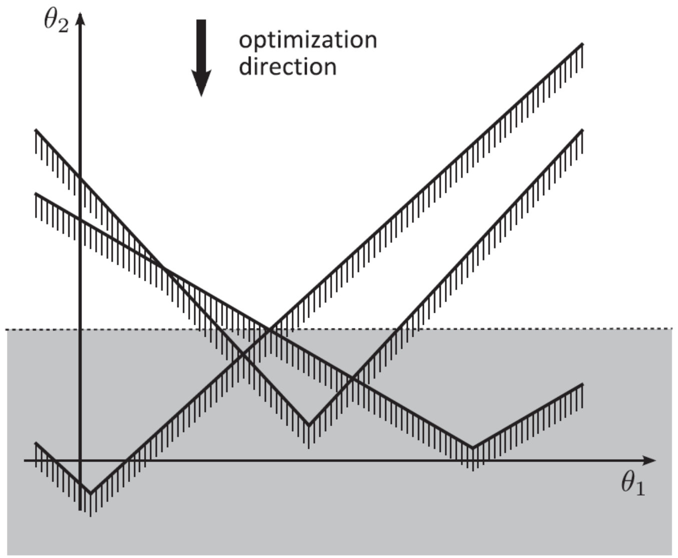
$$\text{subject to:} \quad \{\max(x_1 - \boldsymbol{w}_1^j, 0) + \max(x_2 - \boldsymbol{w}_2^j, 0) \geq 0\} \leq 0, \;\; j = 1, \ldots, N$$

- We need at most two samples in the region with area $\epsilon$ out of $N$ samples
- Binomial distribution: draw at least 2 out of N with success chance $\epsilon$



# Helly‘s lemma

- Proof number of constraints
  - Convex sets defined by each constraints and the saded region with super-optimal points
  - Proof by contradiction: if support constraints larger than 2 then intersection with shaded region non-empty, which contradicts the fact that the original \theta is optimal



# Generalization in the scenario approach

- The cummulative distribution of $V(\theta^*)$ can be bounded as

$$F_V(\epsilon) := \Pr_{\mathcal{W}^N}\{V(\theta) \leq \epsilon\} \leq 1 - \sum_{i=0}^{d-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i}$$

- The right hand side is the cumulative distribution of a Beta distribution with $d$ and $N - d + 1$ degres of freedom. Its density is

$$f_B(\epsilon) = d\binom{N}{d}\epsilon^{d-1}(1-\epsilon)^{N-d}$$

# Removing scenarios to improve guarantees

- It is possible to sample and discard some scenarios
  - Improve bound on cost
  - In case of unbounded disturbance

$$\Pr_{\mathcal{W}^N}\{V(\theta_k^*) > \epsilon\} \leq \binom{k+d-1}{k} \sum_{i=0}^{d+k-1} f_B(N, i, \epsilon) = \delta$$

- Bound more conservative ($\delta > 1$ possible)

# Order statistics in stochastic optimization

- Order statistics shifts the perspective from the underlying distributions to the family of binomial distributions
- Results are independent on the number of uncertainties
- For convex programs: Scenario approach gives a prioiri guarantees by including samples of the uncertain constraint
- Beyond convex programs: a posteriori guarantees
  - Determine the number of support constraints after the solution
  - Validate a guess by testing it multiple times

# Applications and extensions

- Conformal quantile regression/ conformal predictors
  - Use neural networks to fit error model
  - Quantify their error with probabilistic validation
  - Use this error as backoff
- Derive bounds on error norm to use in robust control $\Pr_{\mathcal{W}}\{\|w\|_P > \gamma\} \leq \eta$
- Compare finite families of controllers against each other

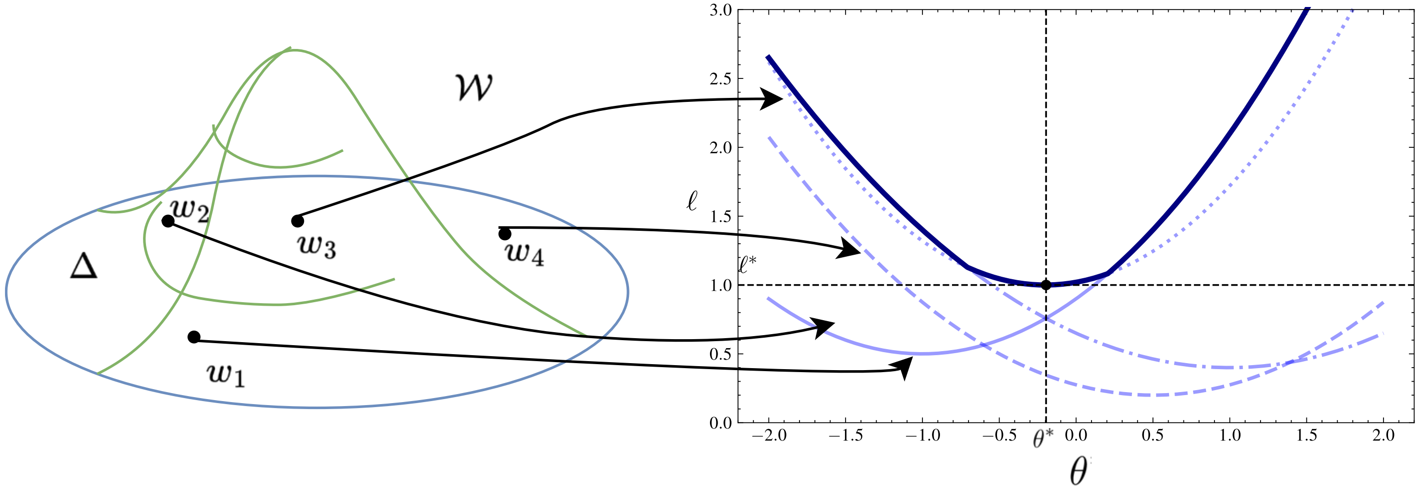
- BUT: always only one performance metric

B. Karg, T. Alamo, und S. Lucia, „Probabilistic performance validation of deep learning-based robust NMPC controllers“, *Int J Robust Nonlinear Control*, Bd. 31, Nr. 18, S. 8855–8876, 2021, doi: 10.1002/rnc.5696.
Y. Romano, E. Patterson, und E. Candes, „Conformalized Quantile Regression“, in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2019. Zugegriffen: 18. Juli 2024. [Online]. Verfügbar unter: https://proceedings.neurips.cc/paper/2019/hash/5103c3584b063c431bd1268e9b5e76fb-Abstract.html
K. Margellos, P. Goulart, und J. Lygeros, „On the Road Between Robust Optimization and the Scenario Approach for Chance Constrained Optimization Problems“, *IEEE Trans. Automat. Contr.*, Bd. 59, Nr. 8, S. 2258–2263, Aug. 2014, doi: 10.1109/TAC.2014.2303232.

# The scenario approach: the simplest setting

- Consider convex cost functions
- Consider only a finite number of samples in a worst-case setting

$$\min_{\theta \in \mathbb{R}^{d-1}} \left[ \max_{i=1,\ldots,N} \ell(\theta, w^{(i)}) \right]$$




- Probabilistic validation can be understood as a convex optimization problem with the performance metric as single decision variable

# Idea behind the improvement

- $\Pr_{\mathcal{W}^N}\{\Pr_{\mathcal{W}}\{f(w;\theta) > J_1^N(\boldsymbol{w};\theta) \leq \epsilon\} \geq 1-\delta$ with $N \geq \frac{1}{\epsilon}\ln\frac{1}{\delta}$

- Let's say $\Pr_{\mathcal{W}}\{f(w;\theta) > J_1^N(\boldsymbol{w};\theta)\} > \epsilon$

- How likely is it, that in N samples, we have not observed a larger value than $J_1^N(\boldsymbol{w};\theta)$?

# Analogy for the setting

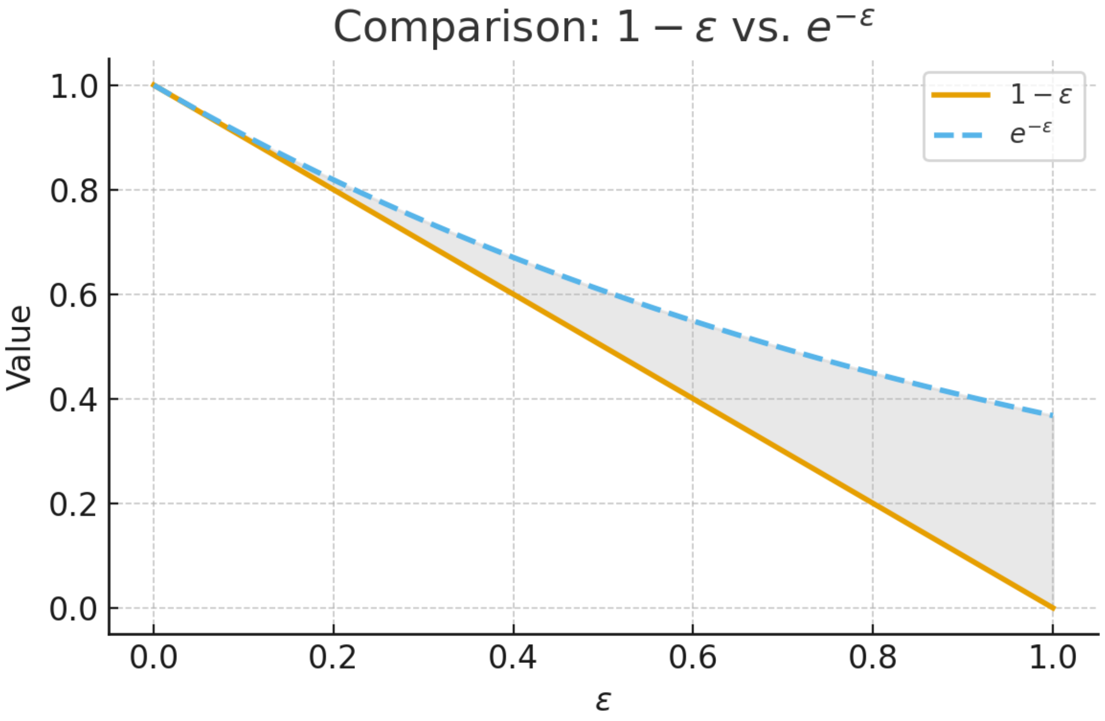
- You have a game with win rate $1 - \epsilon$
- Play N times without loosing
- Chance of success $\Pr = (1-\epsilon)^N$
- Here, a win is equivalent to observing a sample ≤ $J_1^N(\boldsymbol{w};\theta)$
- As $\Pr_{\mathcal{W}}\{f(w;\theta) > J_1^N(\boldsymbol{w};\theta)\} > \epsilon$ confidence at least

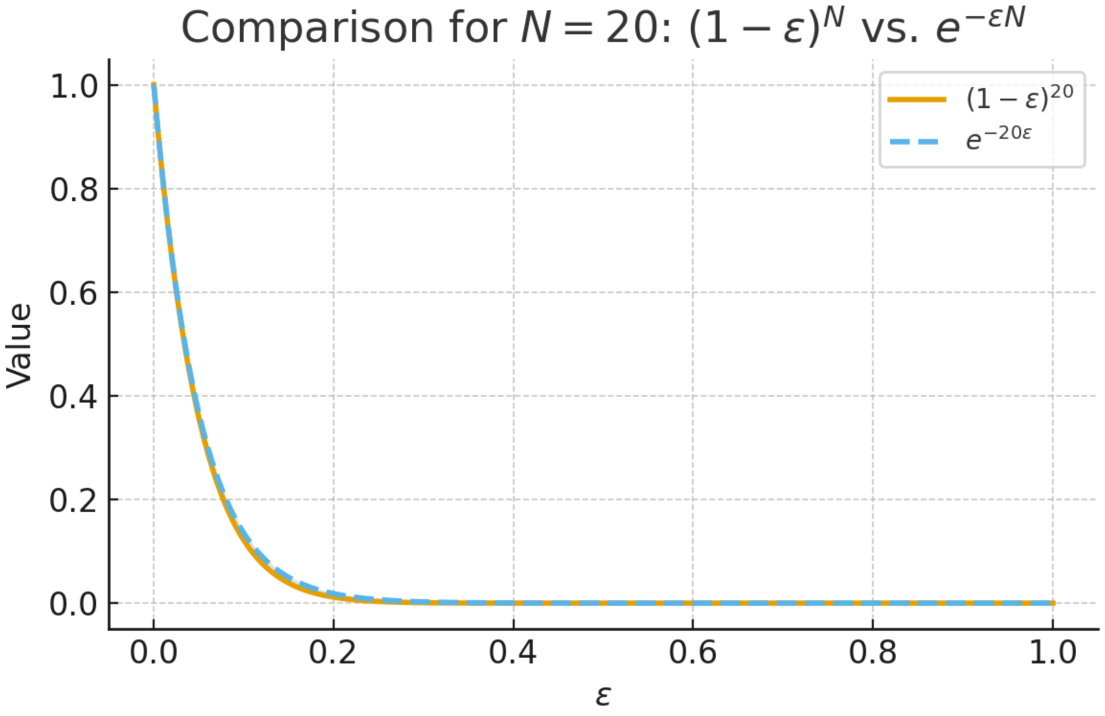
$$(1-\epsilon)^N \le \delta$$

- Use simplification $(1-\epsilon)^N \le \exp(-\epsilon N) \le \delta$
- Take logarithmus $\exp(-\epsilon N) \le \delta$

$$\Leftrightarrow -\epsilon N \le \ln(\delta)$$

$$\Leftrightarrow N \ge \frac{1}{\epsilon} \ln \frac{1}{\delta}$$


Comparison: $1-\varepsilon$ vs. $e^{-\varepsilon}$
$1-\varepsilon$
$e^{-\varepsilon}$
Value
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
$\varepsilon$


Comparison for $N=20$: $(1-\varepsilon)^N$ vs. $e^{-\varepsilon N}$
$(1-\varepsilon)^{20}$
$e^{-20\varepsilon}$
Value
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
$\varepsilon$