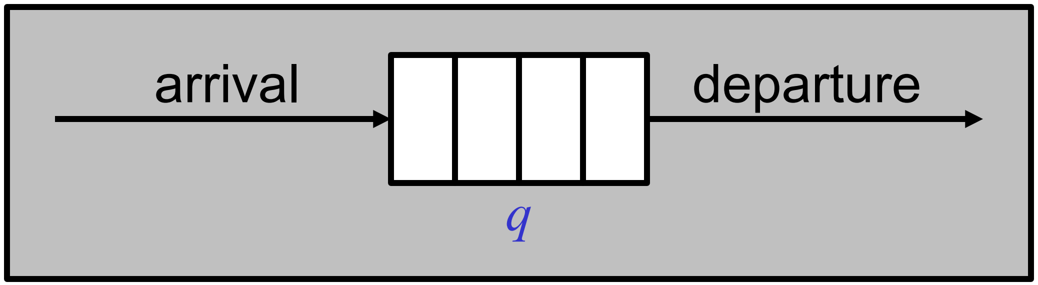
Probabilistic Verification for “Black-Box” Systems

Håkan L. S. Younes
Carnegie Mellon University

Probabilistic Verification



"The probability is at least 0.1 that the queue becomes full within 5 minutes"

Probabilistic Model Checking

- Given a model M, a state s, and a property Φ, does Φ hold in s for M?
  - Model: stochastic discrete event system
  - Property: probabilistic temporal logic formula
- Solution methods:
  - Numerical computation of probabilities
  - Statistical hypothesis testing and simulation (randomized algorithm)

Temporal Stochastic Logic

- Standard logic operators: $\neg\Phi$, $\Phi \wedge \Psi$, …
- Probabilistic operator: $\mathcal{P}_{\geq\theta}[\varphi]$
  - Holds in state s iff probability is at least θ for paths satisfying φ and starting in s
- Until: $\Phi\,\mathcal{U}^{\leq T}\,\Psi$
  - Holds over path σ iff Ψ becomes true along σ within time T, and Φ is true until then

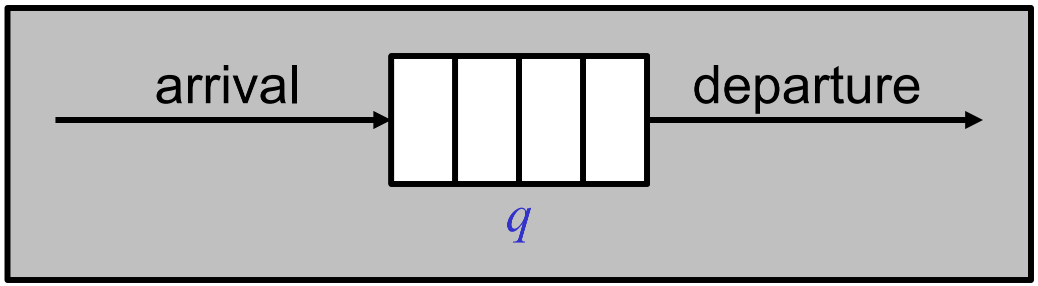
Property Example

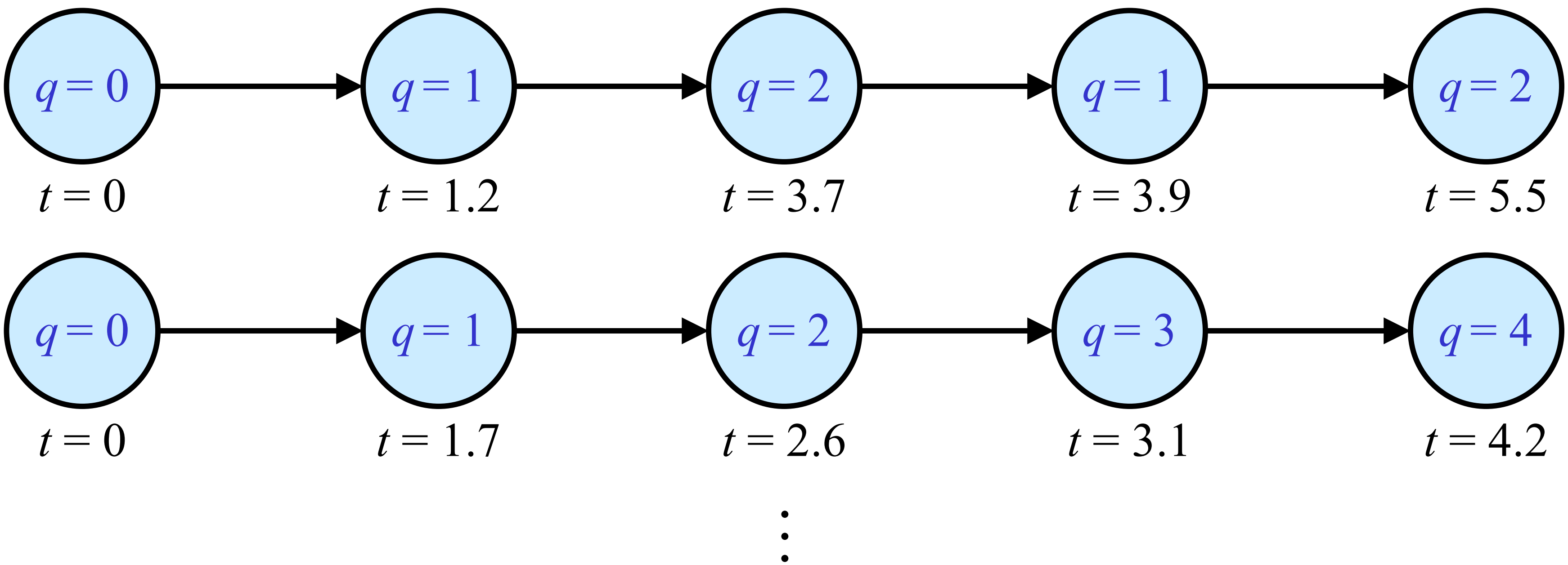
- "The probability is at least 0.1 that the queue becomes full within 5 minutes"
  - $\mathcal{P}_{\geq 0.1}[\top\ \mathcal{U}^{\leq 5}\ \mathit{full}\ ]$

Black-Box Verification

- What if the system is a black box?
  - Unknown system dynamics (no model)
  - Information about system must be obtained through observation during actual execution
  - Numerical computation and discrete-event simulation not possible without model

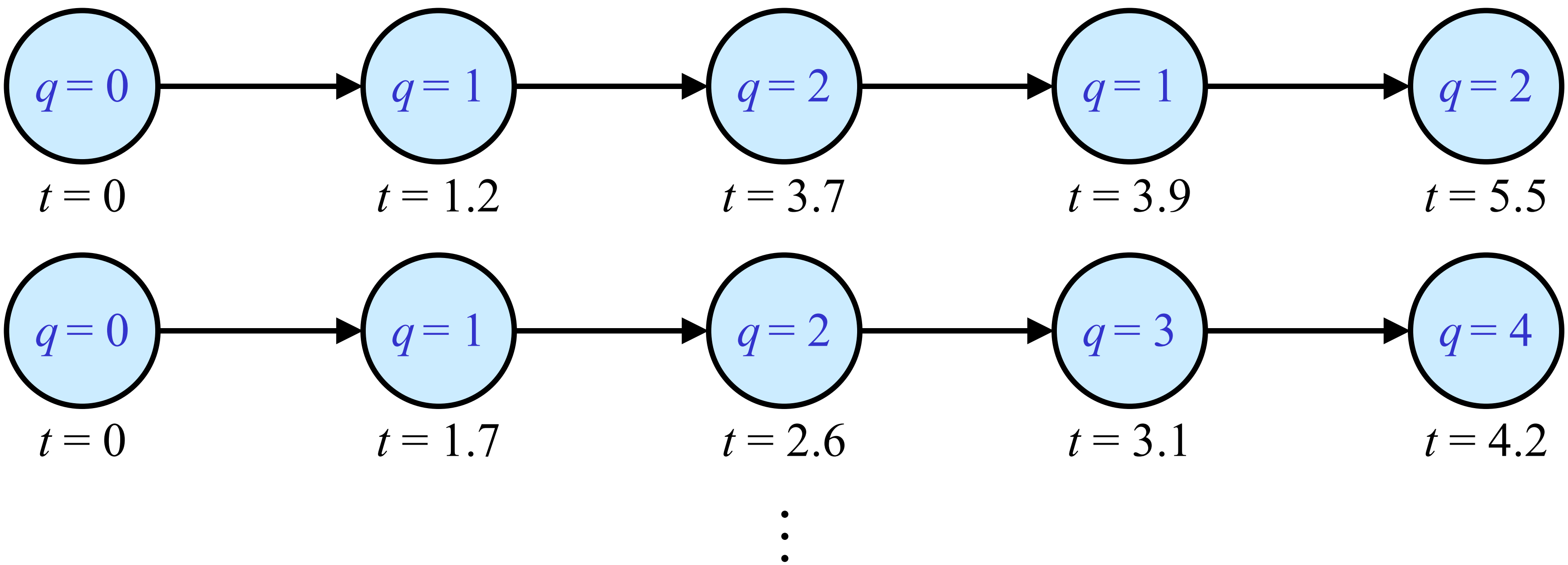
System Execution Traces





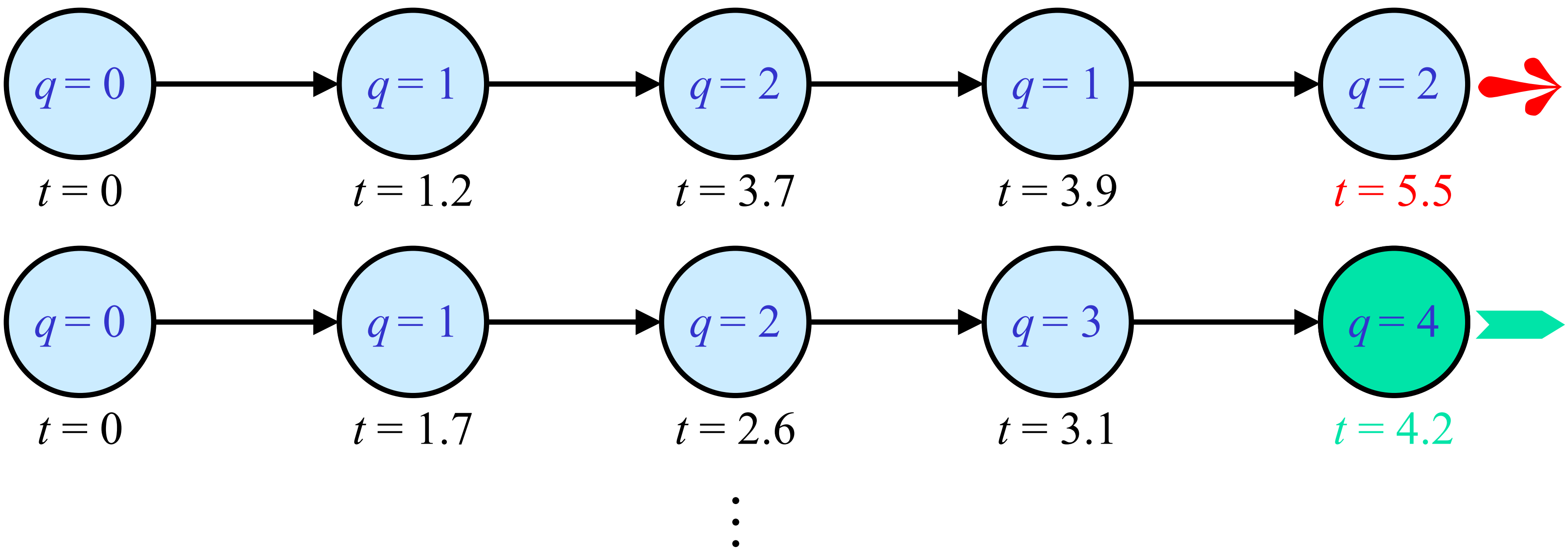
Probabilistic Verification using System Execution Traces

Does $\mathcal{P}_{\geq 0.1}[\top\ \mathcal{U}^{\leq 5}\ \mathit{full}\,]$ hold?



Verifying Path Formulae

Does $\top \, \mathcal{U}^{\leq 5} \, \mathit{full}$ hold?



Verifying Probabilistic Formulae

- Verify $\mathcal{P}_{\geq\theta}[\varphi]$ given n execution traces:
  1. Verify φ over each execution trace
  2. Let d be the number of "positive" traces
  3. Accept $\mathcal{P}_{\geq\theta}[\varphi]$ as true if d is "sufficiently large" and reject $\mathcal{P}_{\geq\theta}[\varphi]$ as false otherwise

Measure of Confidence: p-value

- Low p-value implies high confidence
- Definition of p-value:
  - Probability of the given or a more extreme observation provided that the rejected hypothesis is true

Measure of Confidence: p-value

- Probability of observing at most d positive traces given a p probability measure for the set of positive traces:

$$F(d;n,p) = \sum_{i=0}^{d} \binom{n}{i} p^i (1-p)^{n-i}$$

	Accept $\mathcal{P}_{\geq\theta}[\varphi]$	Reject $\mathcal{P}_{\geq\theta}[\varphi]$
p-value	$1-F(d-1;n,\theta)$	$F(d;n,\theta)$

Choosing the Acceptance Threshold

- When is d sufficiently large?
  - Compute p-value for both answers
  - Choose answer with lowest p-value
  - No need to compute explicit threshold
- Note: Sen et al. (CAV'04) use $\lceil n\theta \rceil - 1$ as threshold, which can lead to an answer with a larger p-value than the alternative

Example

- Should we accept $\mathcal{P}_{\geq 0.1}[\top\ \mathcal{U}^{\leq 5}\ full\,]$ if we have 37 positive and 63 negative traces?
  - Acceptance: $1-F(36;100,0.1) \approx 5.48\cdot 10^{-13}$
  - Rejection: $F(37;100,0.1) \approx 1-10^{-13}$

Computing p-values for Composite Formulae

- Negation $\neg\Phi$:
  - same p-value as for Φ
- Conjunction $\Phi \wedge \Psi$:

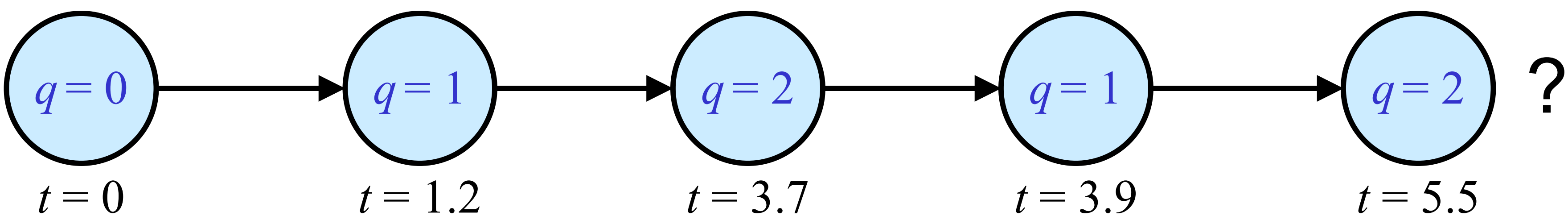
Φ	Ψ	$\Phi \wedge \Psi$	p-value
			$\max(pv_\Phi, pv_\Psi)$
			pv_Ψ
			pv_Φ
			$\min(pv_\Phi, pv_\Psi)$

Sen et al. (CAV'04): $pv_\Phi + pv_\Psi$

Handling Truncated Traces

- Execution traces are finite

Does $\top \, \mathcal{U}^{\leq 10} \, \mathit{full}$ hold?



Handling Truncated Traces

- Computing p-value intervals:
  - n' verifiable traces of n total traces
  - d' positive traces of n' verifiable traces
  - Between d' and $d' + n - n'$ total positive traces

	Accept $\mathcal{P}_{\geq\theta}[\varphi]$	Reject $\mathcal{P}_{\geq\theta}[\varphi]$
p-value	$[1-F(d_{\max}-1;n,\theta), 1-F(d_{\min}-1;n,\theta)]$	$[F(d_{\min};n,\theta), F(d_{\max};n,\theta)]$

Black-Box Verification vs. Statistical Model Checking

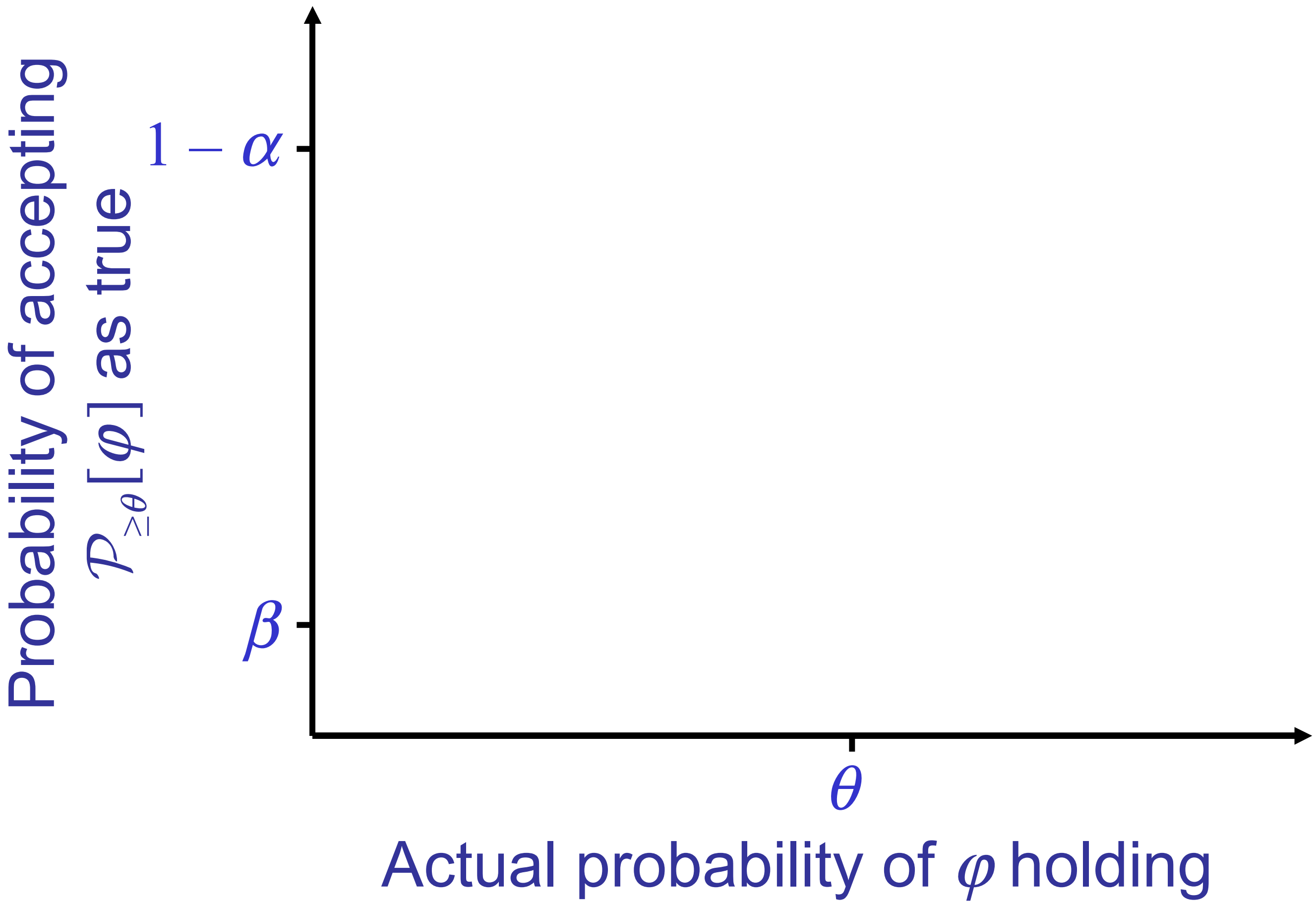
- Black-box verification
  - Fixed set of execution traces
  - Find answer with lowest p-value
- Statistical model checking
  - Traces can be generated from model
  - User determines *a priori* error bounds
  - Number of traces depends on error bounds

Error Bounds for Statistical Model Checking

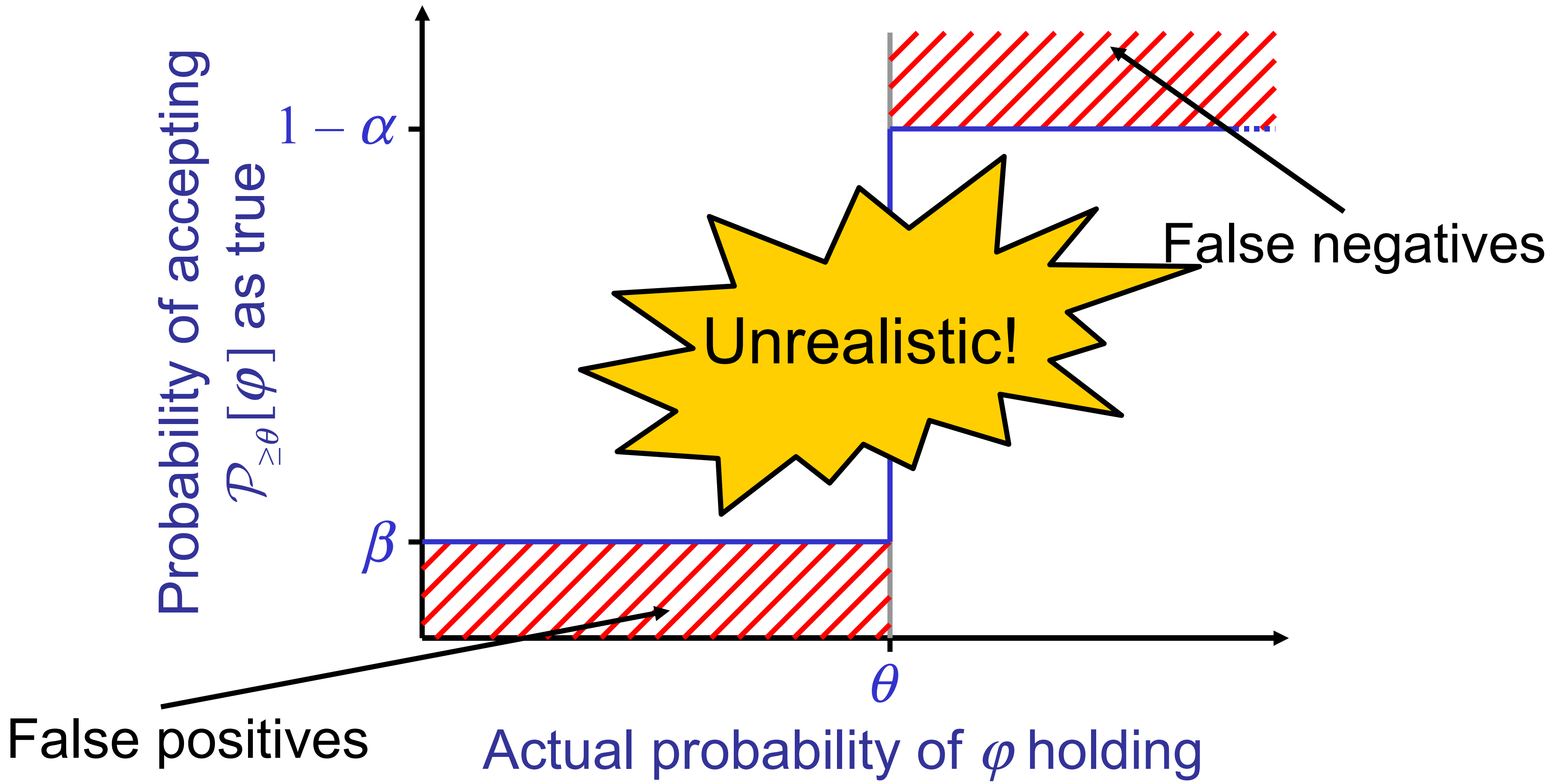
- Probability of false negative: $\leq \alpha$
  - We say that Φ is false when it is true
- Probability of false positive: $\leq \beta$
  - We say that Φ is true when it is false

$(1 - \alpha)$ complete
$(1 - \beta)$ sound

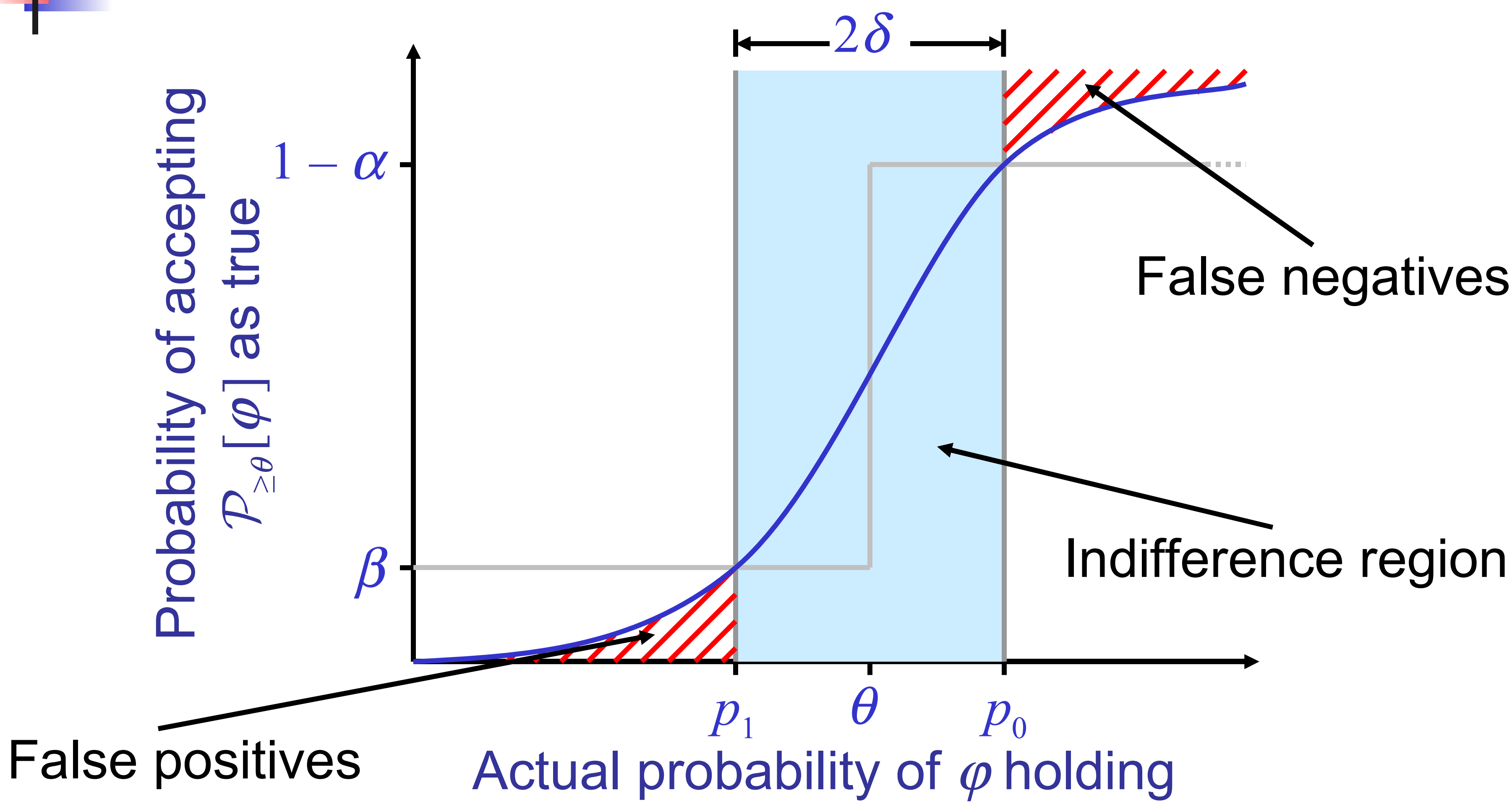
Operational Characteristics of Statistical Model Checking



Ideal Operational Characteristics



Realistic Operational Characteristics



How to Achieve Error Bounds

- Fixed-size sample (single sampling plan)
  - Pick sample size n and acceptance threshold c such that $F(c; n, p_0) \leq \alpha$ and $1 - F(c; n, p_1) \leq \beta$
- Sequential Probability Ratio Test (SPRT)
  - At each stage, compute probability ratio f
  - Accept if $f \leq \beta / (1 - \alpha)$; reject if $f \geq (1 - \beta) / \alpha$; generate additional traces otherwise
  - Sample size is random variable

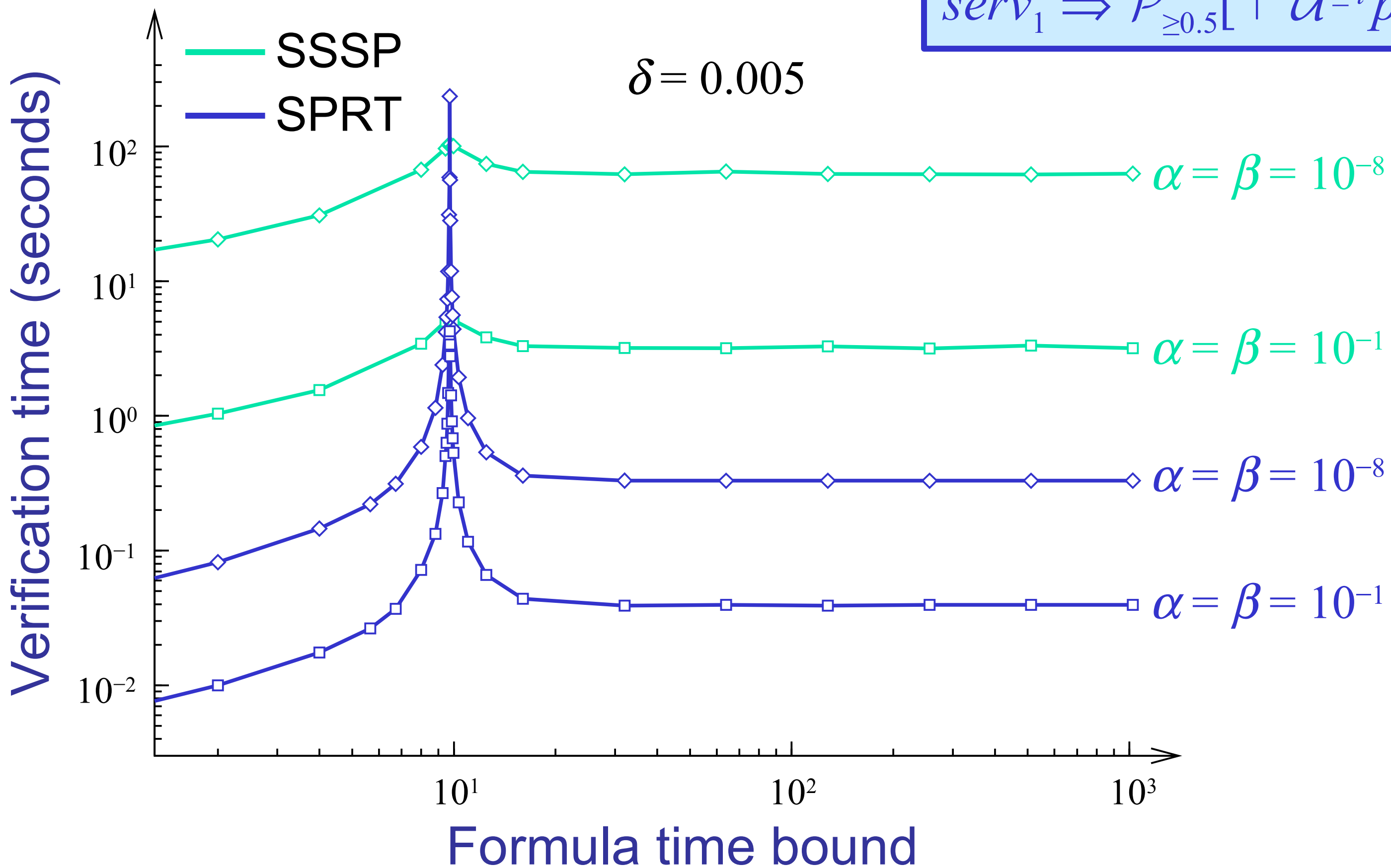
Error Bounds for Composite Formulae

- Negation $\neg\Phi$:
  - $\alpha = \beta_\Phi$
  - $\beta = \alpha_\Phi$
- Conjunction $\Phi \wedge \Psi$:
  - $\alpha = \min(\alpha_\Phi, \alpha_\Psi)$
  - $\beta = \max(\beta_\Phi, \beta_\Psi)$

Younes & Simmons (CAV'02);
Sen et al. (CAV'05): $\beta = \beta_\Phi + \beta_\Psi$

Single Sampling Plan vs. Sequential Probability Ratio Test

$serv_1 \Rightarrow \mathcal{P}_{\geq 0.5}[\top\, \mathcal{U}^{\leq t} poll_1]$



Complexity of Statistical Approach

- Complexity of verifying $\mathcal{P}_{\geq 0.1}[\Phi\, \mathcal{U}^{\leq t}\, \Psi\,]$ is $O(n \cdot e \cdot q \cdot t)$
  - n: sample size
  - e: simulation effort per transition
  - q: expected number of transition per time unit

Statistical Model Checking of Unbounded Until

- Time bound guarantees that finite sample paths suffices
- Sen et al. (CAV'05) use "stopping probability" to ensure finite sample paths
  - In reality, stopping probability must be extremely small to give any correctness guarantees (10^{-8} for $|S|=10$; 10^{-17} for $|S|=20$)

Do not overestimate the power of statistical methods!

Conclusions

- Black-box verification useful to analyze system based on existing execution traces
- Statistical model checking useful when sample paths can be generated at will
- Complementary, not competing, approaches

Ymer:
A Statistical Model Checker

- http://sweden.autonomy.ri.cmu.edu/ymer/
  - Distributed acceptance sampling