Statistical Model Checking

Madhavan Mukund

Chennai Mathematical Institute
http://www.cmi.ac.in/~madhavan

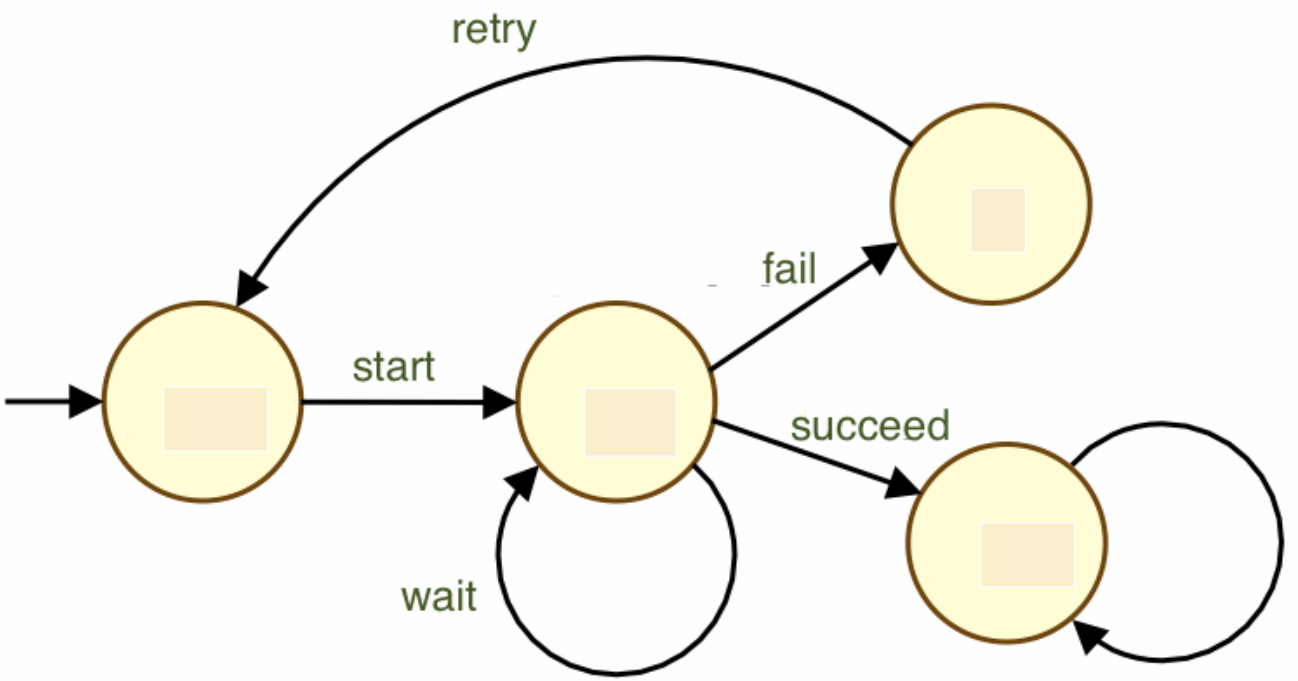
CSI NCFM, 16 October 2014

Stochastic systems

- Next states depends probabilistically on current state and past history
  - Markov property: history prior to current state is ignored
  - Probabilistic transition function
- Useful for modelling
  - Randomization Breaking symmetry in distributed algorithms
  - Uncertainty Environmental interference, imprecise sensors, ...
  - Quantitative properties Performance, quality of service
- Two major domains of application
  - Cyber Physical Systems Auto pilot, anti-lock braking, ...
  - Biological systems Signalling pathways, cell interactions, ...

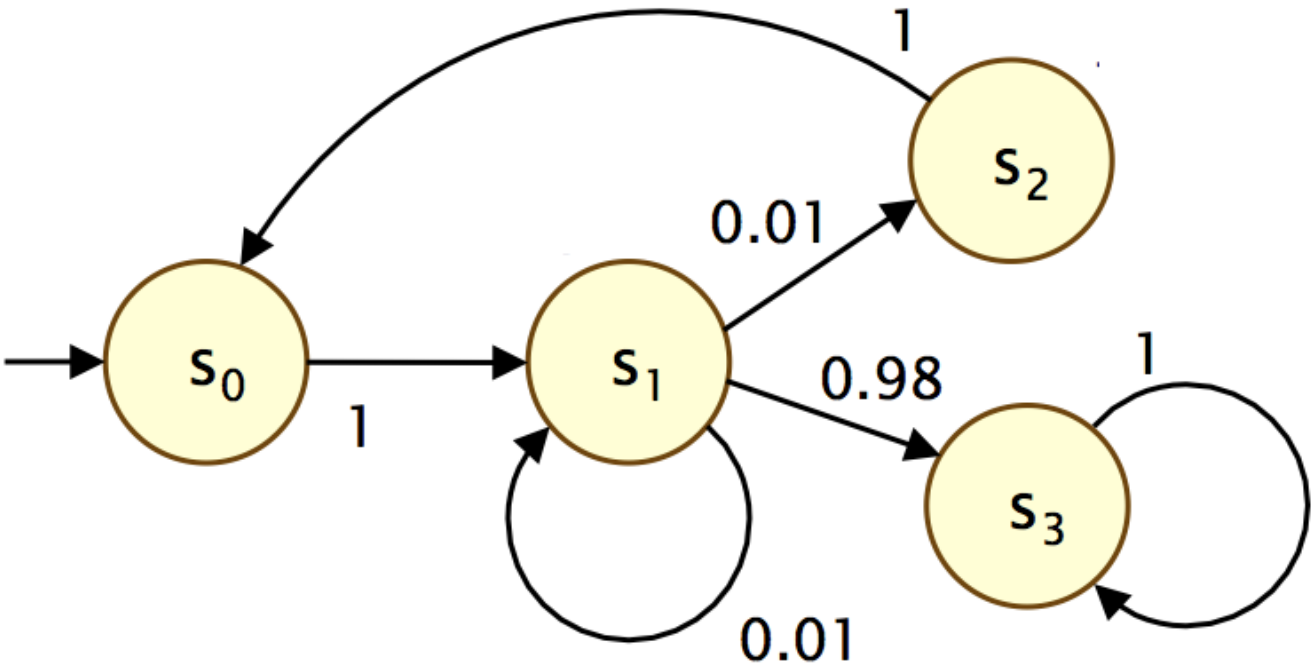
A simple communication protocol

- Sending a message on a channel



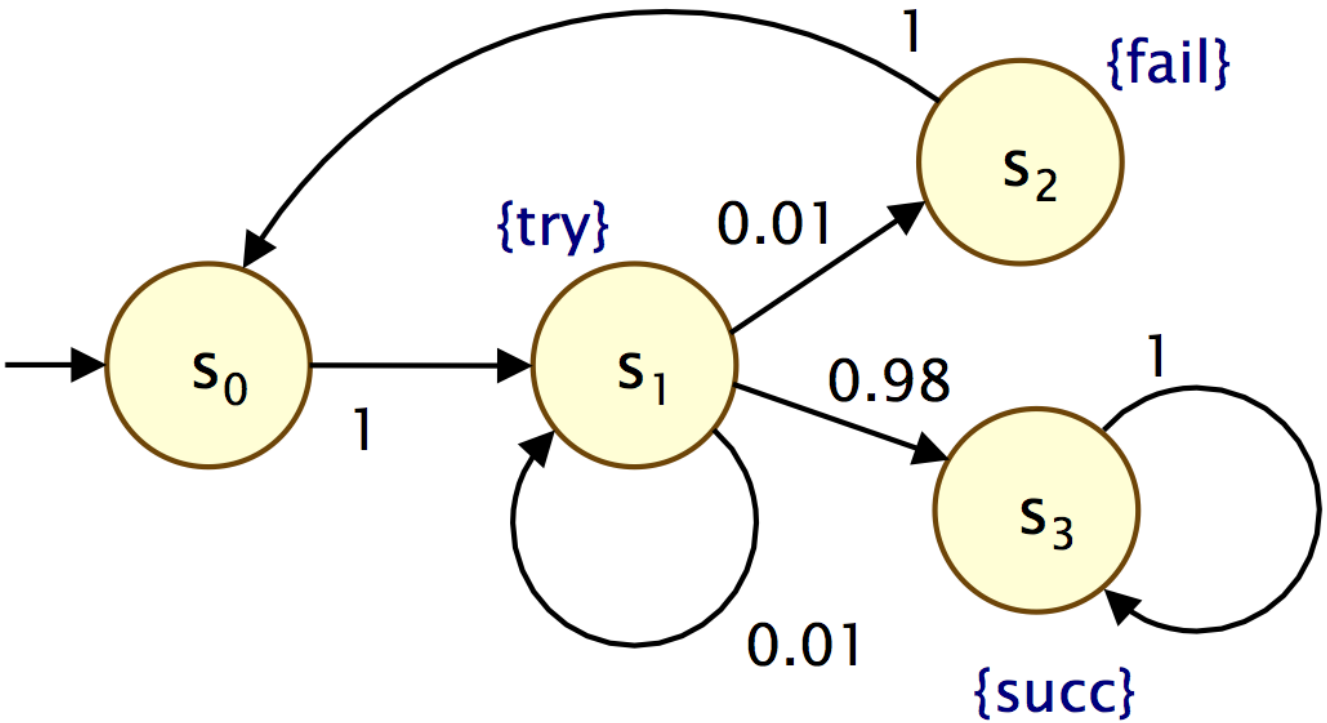
A simple communication protocol

- Associate probabilities with the events



A simple communication protocol

- Label states with atomic propositions



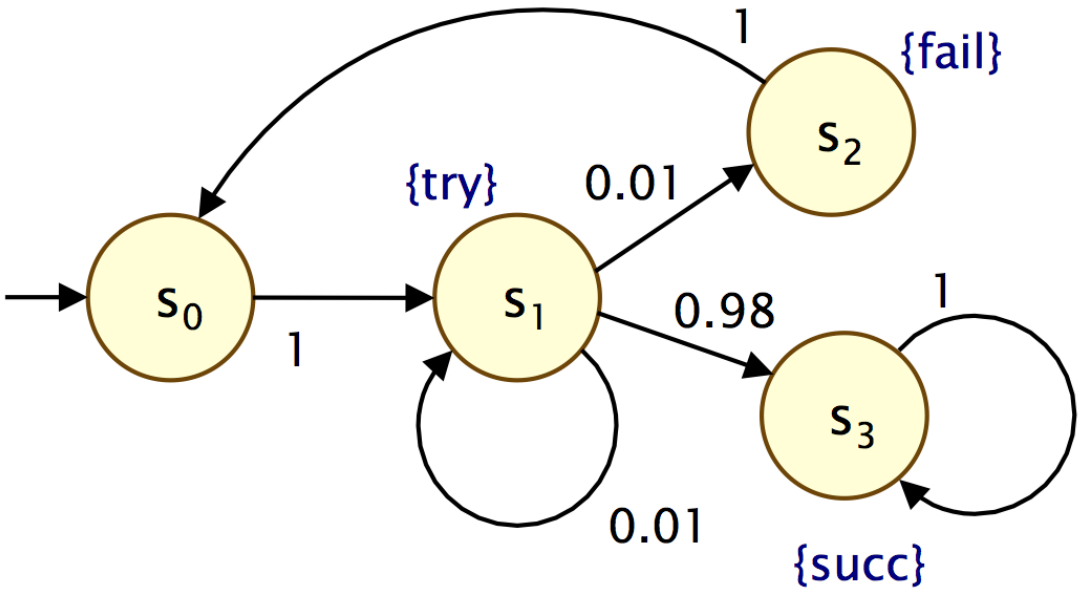
Discrete Time Markov Chain (DTMC)

$D = (S, s_{init}, P, L)$

$S = \{s_0, s_1, s_2, s_3\}$
$s_{init} = s_0$

$AP = \{try, fail, succ\}$
$L(s_0) = \emptyset$,
$L(s_1) = \{try\}$,
$L(s_2) = \{fail\}$,
$L(s_3) = \{succ\}$

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0.01 & 0.01 & 0.98 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



Typical properties of interest

- Path based properties
  What is the probability of requiring more than 10 retries?
- Transient properties
  What is the probability of being in state s_1 after 16 steps?
- Expectation
  What is the average number of retries required?

This talk

- Focus on path based properties

Measuring set of paths

- Properties refer to sets of paths
  - What is the probability of requiring more than 10 retries?
  - Ratio of runs requiring more than 10 retries to set of all runs
  - Runs are infinite paths
- How do we count or measure sets of infinite paths?

Measuring set of paths

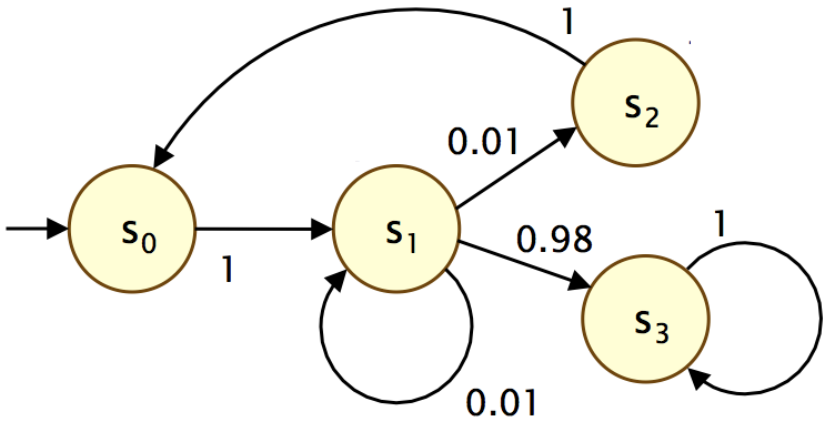
- Probability of a finite path: multiply the probabilities
  - $s_0 \xrightarrow{1} s_1 \xrightarrow{0.01} s_2 \xrightarrow{1} s_0 \xrightarrow{1} s_1 \xrightarrow{0.98} s_3$
  - Probability is $1 \cdot 0.01 \cdot 1 \cdot 1 \cdot 0.98 = 0.0098$
- A single infinite path has probability 0
  - Infinite product of values below 1
- How do we identify sets of infinite paths with non-zero probability?

Measuring set of paths

- A cylinder is a set of paths that share a common prefix
  - $Cyl(s_0 s_1 s_2 s_0 s_1 s_3) = \{\rho \mid \rho = s_0 s_1 s_2 s_0 s_1 s_3 \rho'\}$
  - Collectively, $Cyl(s_0 s_1 s_2 s_0 s_1 s_3)$ has same probability, 0.0098 as the common prefix $s_0 s_1 s_2 s_0 s_1 s_3$
- A set of paths can be measured if:
  - it is a cylinder, or the complement of one
  - it is a countable union of measurable subsets
- The empty set and the set of all runs can be measured (Why?)

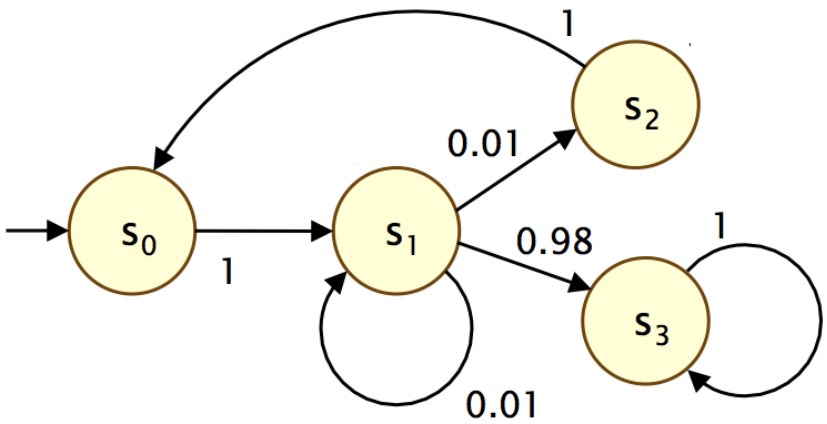
Measuring set of paths

- Paths that fail immediately $Cyl(s_0 s_1 s_2)$



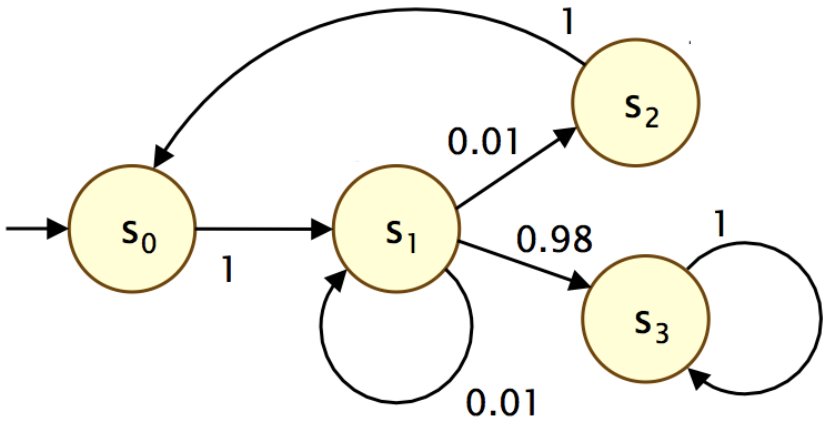
Measuring set of paths

- Paths that fail immediately
  $Cyl(s_0 s_1 s_2)$
- Paths with at least one failure
  $Cyl(s_0 s_1 s_2) \cup$
  $Cyl(s_0 s_1 s_1 s_2) \cup$
  $Cyl(s_0 s_1 s_1 s_1 s_2) \cup \cdots$



Measuring set of paths

- Paths that fail immediately
  $Cyl(s_0 s_1 s_2)$
- Paths with at least one failure
  $Cyl(s_0 s_1 s_2) \cup$
  $Cyl(s_0 s_1 s_1 s_2) \cup$
  $Cyl(s_0 s_1 s_1 s_1 s_2) \cup \cdots$
- Paths with no waiting and at most two failures
  $Cyl(s_0 s_1 s_3) \cup$
  $Cyl(s_0 s_1 s_2 s_0 s_1 s_3) \cup$
  $Cyl(s_0 s_1 s_2 s_0 s_1 s_2 s_0 s_1 s_3)$



Model checking properties

- Focus on probabilistic reachability
  - "Something good happens"
  - $R_s(s')$: probability of reaching s' from s
  - Measure of all paths ρ starting with s that contain s'
- Dual is invariance
  - Behaviour stays within a subset of states $G \subseteq S$
  - "Nothing good happens"
  - Behaviour never reaches $S \setminus G$
- All reachability probabilities are measurable
  - $R_s(s')$: union of all cylinders $Cyl(ss_0s_1 \ldots s_k s')$ where s' does not occur in $ss_0 \ldots s_k$
  - Disjoint cylinders, so add their measures, no double counting

Model checking properties

- Express $R_s(s')$ inductively
  - $R_s(s') = 1$, if $s = s'$
  - $R_s(s') = \sum_{s''} P(s, s'') R_{s''}(s')$, otherwise
- Similar equations for each $R_{s''}(s')$
- Solution we want is a fixed point for this system of equations
- Can be solved iteratively
  - Initially assign 0 to each $R_{s''}(s')$
  - Update using the inductive definition

Drawbacks

- Explicit computation of probabilities needs to examine all states
- Need to manipulate entire reachable state space
- Infeasible for practical applications
  - Cyber physical systems
  - Biological models

Statistical approach

- Simulate the system repeatedly using the underlying probabilities
- For each run determine if it satisfies the property
- Suppose c out N runs are successful
- Estimate the probability of the property holding as $\frac{c}{N}$

Law of large numbers

As N tends to ∞, $\frac{c}{N}$ converges to the true value

Statistical approach

- Why does this help?
- Simulation is easier than exhaustively exploring state space
  - Only need to remember states along the path
  - How much of the path you keep depends on the property
- Easy to parallelize: simulations are independent

Constraints

- Properties must be bounded
  - Each simulation succeeds or fails in a finite amount of time
- Number of simulations may be large
- Guarantees are probabilistic
  - Explicit computations "solve" probabilistic systems "exactly"

Bounded properties

- Is reachability a bounded property?
  - If the simulation reaches s' we can stop the simulation
  - What if the simulation does not visit s' for a long time?
- Bounded reachability: reachable in k steps or less
- Generalize linear-time temporal logic (LTL) to Bounded LTL (BLTL)
  - Atomic propositions
  - Boolean connectives $\neg$, $\wedge$, ...
  - $X\varphi$: φ holds at the next state
  - $\varphi U^k \psi$: within k steps, ψ will hold and until then φ holds
- Interpret along a run $\rho = s_0 s_1 s_2 \ldots$

Monte Carlo model checking

- Input is a BLTL formula φ and a probabilistic system D
- Inductively compute a bound t from φ
  - Non trivial bounds come from subformulas $\psi U^k \psi'$
  - Bound for $\neg\psi$ is same as for ψ
  - For $\psi \wedge \psi'$, use max of the two bounds
  - For $X\psi$, add 1 to the bound for ψ
- Simulate the system N times, each simulation bounded by t steps
- Report $\frac{c}{N}$, where c is the number of good runs

Statistical estimation

Coin tossing

- Toss a coin 100 times, observe 70 heads — estimate $P(h) = 0.7$
  - Maximum likelihood estimate
  - Of all possible values of $P(h)$, 0.7 maximizes probability of the given observation, 70 heads out of 100
- Observe 7 heads out of 10, 70 out of 100, 700 out of 1000
  - All give the same estimate
  - Are all the experiments equivalent?
- Intuitively, more trials give us more confidence in the estimate
- How do we quantify this?

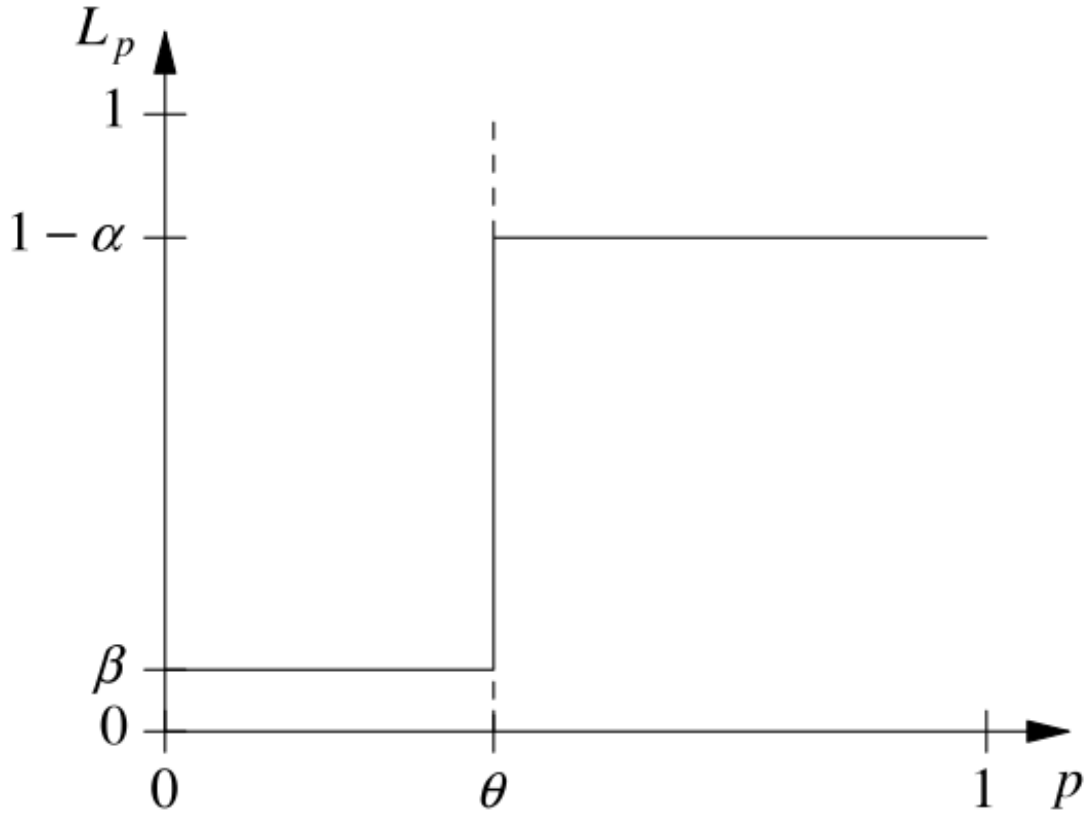
Hypothesis testing

- Rephrase the problem: Is $P(h) \geq \theta$
  - Call this our hypothesis H
  - The converse hypothesis is $K : P(h) < \theta$
- Fix the number of simulations, N, and a threshold, c
- After our simulation
  - If more than c of N simulations succeed, accept H
  - If c or fewer simulations succeed, reject H, accept K
- Errors
  - False negative (Type-I) Accept K when H holds
  - False positive (Type-II) Accept H when K holds
- Want to bound the error of our estimate
  - Probability of a Type-I error is bounded by α
  - Probability of a Type-II error is bounded by β

Hypothesis testing

Probability L_p of accepting hypothesis $H : p \geq \theta$ as a function of p

- Step function requires exhaustive sampling

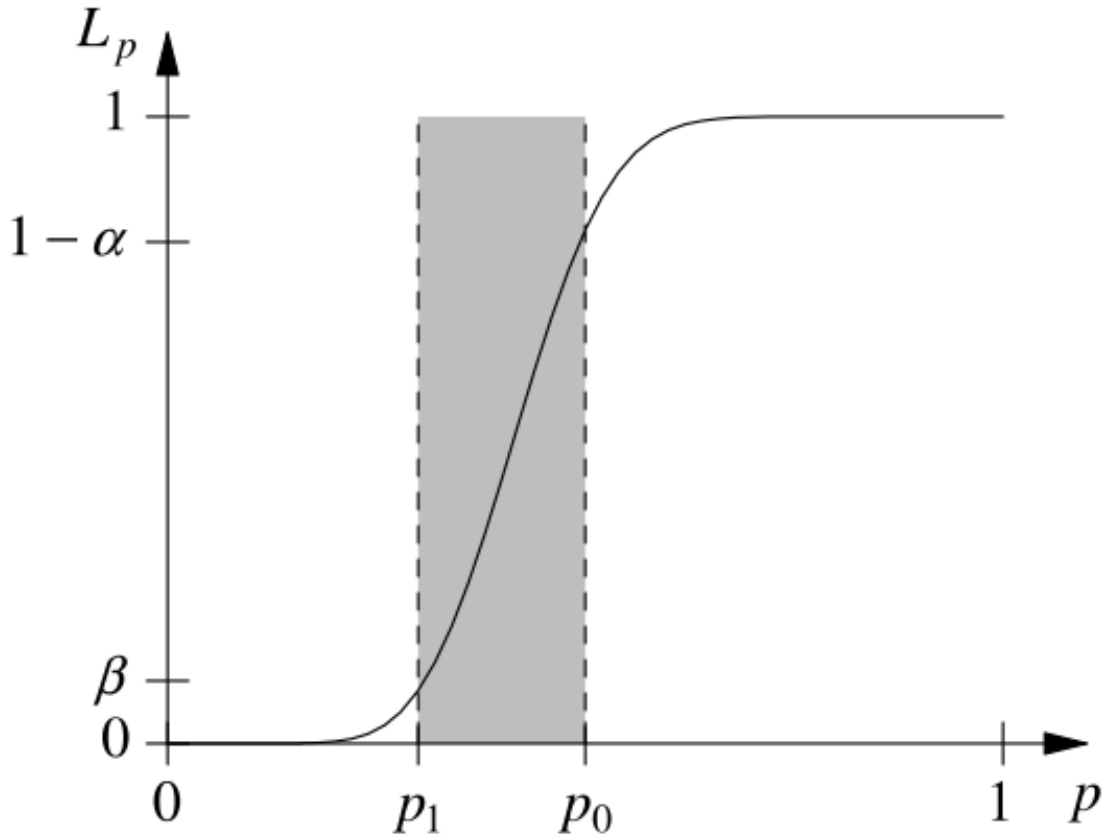


Hypothesis testing

- Instead, introduce an indifference region, $\theta \pm \delta$
  - Hypthesis H: $P(h) \geq \theta + \delta = p_0$
  - Hypthesis K: $P(h) \leq \theta - \delta = p_1$
- Within the indifference region, we are neutral to the answer
  - Too close to call!

Hypothesis testing

Probability L_p of accepting hypothesis $H : p \geq p_0$ as a function of p, with an indifference region



Single sampling plan

- Given $H : p \geq p_0$, $K : p \leq p_1$
- Fix a number of trials N and constant c so that
  - If we see more than c successes, we accept H
  - If we see c or fewer successes, we accept K
- Since we have fixed c, we may make mistakes
  - Type-I: accept K when H holds (false negative)
  - Type-II: accept H when K holds (false positive)
- How do we choose N and c so that we achieve desired error bounds?
  - Probability of Type-I errors bounded by α
  - Probability of Type-II errors bounded by β

Single sampling plan

- Let X be a Bernoulli variable (i.e., a biased coin) with probability p
- Let Y be the number of successes (i.e., heads) after N trials

$$F(c; N, p) \triangleq P(Y \leq c) = \sum_{i=0}^{c} \binom{N}{i} p^i (1-p)^{N-i}$$

- We have fixed a threshold c, so
  - We accept K with probability $F(c; N, p)$
  - We accept H with probability $1 - F(c; N, p)$
- The sampling plan $\langle N, c \rangle$ has strength $\langle \alpha, \beta \rangle$ if
  - $F(c; N, p_0) \leq \alpha$ ..... (accept K when $p \geq p_0$)
  - $1 - F(c; N, p_1) \leq \beta$ ..... (accept H when $p \leq p_1$)

Single sampling plan

- The sampling plan $\langle N, c \rangle$ has strength $\langle \alpha, \beta \rangle$ if
  - $F(c; N, p_0) \leq \alpha$
  - $1 - F(c; N, p_1) \leq \beta$
- Can compute N and c that satisfy these constraints
- Unfortunately, no closed form solution
- Can numerically solve using binary search [Younes]

Adaptive sampling

- Suppose our single sampling plan is $\langle 1000, 700 \rangle$
  - 1000 samples, accept H if 701 or more successes, K otherwise
- We have completed 600 samples and already observed 300 failures
- No point in continuing the test!
- Can we do adaptive sampling?

Sequential probability ratio test (SPRT) [Wald 1945]

- After m samples, suppose we have seen d_m successful samples

$$f_m = \prod_{i=1}^{m} \frac{Pr[X_i = x_i \mid p = p_1]}{Pr[X_i = x_i \mid p = p_0]} = \frac{p_1^{d_m}(1-p_1)^{m-d_m}}{p_0^{d_m}(1-p_0)^{m-d_m}}$$

  - Numerator captures likelihood of current sample with hypothesis K
  - Denominator captures likelihood of current sample with hypothesis H

- Fix two thresholds A, B
  - If ratio f_m is above A, accept K and stop
  - If ratio f_m is below B, accept H and stop
  - Otherwise, continue drawing samples

Sequential probability ratio test (SPRT)

- After m samples, suppose we have seen d_m successful samples

$$f_m = \prod_{i=1}^{m} \frac{Pr[X_i = x_i \mid p = p_1]}{Pr[X_i = x_i \mid p = p_0]} = \frac{p_1^{d_m}(1-p_1)^{m-d_m}}{p_0^{d_m}(1-p_0)^{m-d_m}}$$

- Fix two thresholds A, B
  - If ratio f_m is above A, accept K and stop
  - If ratio f_m is below B, accept H and stop
  - Otherwise, continue drawing samples
- Fixing A and B to give overall strength $\langle \alpha, \beta \rangle$ is nontrivial
  - In practice, choose $A = \dfrac{1-\beta}{\alpha}$ and $B = \dfrac{1-\alpha}{\beta}$
  - This yields a test with strength $\langle \alpha', \beta' \rangle$ where $\alpha' + \beta' \leq \alpha + \beta$
  - At least one of the new bounds is smaller than the original, usually both

SPRT based statistical model checking

- Fix a threshold θ and an indifference region $\theta \pm \delta$ along with error bounds α, β
- Let $p_0 = \theta + \delta$, $p_1 = \theta - \delta$
- Set $A = \dfrac{1-\beta}{\alpha}$ and $B = \dfrac{1-\alpha}{\beta}$
- Draw samples and evaluate the ratio f_i after sample i
  - If $f_i > A$, accept K
  - If $f_i < B$, accept H
  - Otherwise, draw another sample

Boolean combinations

- To verify $\neg\psi$ with Type-I error α and Type-II error β, sufficient to verify ψ with Type-I error β and Type-II error α.
- Let $\varphi = \psi_1 \wedge \psi_2$.
  - Assume that each ψ_i can be decided with Type-I error α_i and Type-II error β_i .
  - Then φ can be decided with Type-I error $\min(\alpha_1, \alpha_2)$ and Type-II error $\max(\beta_1, \beta_2)$.

Nested probabilities

What if we have nested probabilistic operators?

- Compute $Pr_{\geq\theta}(\varphi)$, where φ itself is of the form $Pr_{\geq\theta'}(\psi)$.
- A single sample is not enough:
  - Need to nest sampling for ψ with each sample of φ
- Can be done
  - Can derive errors bounds for φ given bounds for ψ
  - But expensive: exponential blow up in samples

Other challengess

- Rare events
  - If the probability is very low, need more samples for meaningful estimate
  - Alternative notion called importance sampling [Clarke,Zuliani]
- Incorporating nondeterminism
  - Markov Decision Processes: each action determines a separate probability distribution
  - Cannot directly apply statistical techniques
  - Resolve nondeterminism using a scheduler

Tools

- Plasma
  - `https://project.inria.fr/plasma-lab/statistical-model-checking`
  - Highly optimized, with parallel threads
- UPPAAL
  - Extension to UPPAAL for statistical model checking of timed automata
  - `http://people.cs.aau.dk/~adavid/smc`
  - Application of statistical model checking to quantitative properties

Reading

- Håkan L. S. Younes, Reid G. Simmons:
  Statistical probabilistic model checking with a focus on time-bounded properties
  Information and Computation 204(9), (2006) 1368–1409
- Axel Legay, Benoît Delahaye, Saddek Bensalem:
  Statistical Model Checking: An Overview
  Proc. Runtime Verification (RV) 2010,
  Springer LNCS 6418, (2010) 122–135