Advanced Topics in Software Engineering: Statistical transient analysis

Prof. Michele Loreti

Advanced Topics in Software Engineering
Corso di Laurea in Informatica (L31)
Scuola di Scienze e Tecnologie

Statistical inference & Simulation...

Statistical Inference can be used in combination with simulation to perform many kind of study:

- performance estimation;
- (system) parameter estimation and optimisation;
- transient analysis;
- ...

Statistical inference & Simulation...

Statistical Inference can be used in combination with simulation to perform many kind of study:

- performance estimation;
- (system) parameter estimation and optimisation;
- transient analysis;
- ...

Basic ingredients:

Statistical inference & Simulation...

Statistical Inference can be used in combination with simulation to perform many kind of study:

- performance estimation;
- (system) parameter estimation and optimisation;
- transient analysis;
- ...

Basic ingredients:

- A model $\mathcal{M}$ that describes our system as a random variable $X(t)$ associating each time $t \in \mathbb{R}_{\geq 0}$ a vector of observations in $\mathbb{R}$;

Statistical inference & Simulation...

Statistical Inference can be used in combination with simulation to perform many kind of study:

- performance estimation;
- (system) parameter estimation and optimisation;
- transient analysis;
- ...

Basic ingredients:

- A model $\mathcal{M}$ that describes our system as a random variable $X(t)$ associating each time $t \in \mathbb{R}_{\geq 0}$ a vector of observations in $\mathbb{R}$;
- A simulator function $\text{simulate}_{\mathcal{M}}(t)$ that, given a time $t \in \mathbb{R}_{\geq 0}$ sample a path fragment/realisation/computation from $\mathcal{M}$.

The SEIR model is a compartmental models that can be used to study infection diseases.

Statistical inference & Simulation. . .

Example

The SEIR model is a compartmental models that can be used to study infection diseases.

The model consists of four groups of agents:

- **S**uscettible;
- **E**xposed;
- **I**nfected;
- **R**ecovered.

The SEIR model is a compartmental models that can be used to study infection diseases.

The model consists of four groups of agents:

- **S**uscettible;
- **E**xposed;
- **I**nfected;
- **R**ecovered.

Events:

- One **S**uscettible becomes **E**xposed;
- One **E**xposed becomes **I**nfected;
- One **I**nfected becomes **R**ecovered,

Vector Variables: (S, E, I, R)

Vector Variables: $(\mathsf{S}, \mathsf{E}, \mathsf{I}, \mathsf{R})$

Counting Domain: $[0, N] \times [0, N] \times [0, N] \times [0, N]$

Statistical inference & Simulation...

Example: SEIR Population Model

Vector Variables: $(\mathsf{S}, \mathsf{E}, \mathsf{I}, \mathsf{R})$

Counting Domain: $[0, N] \times [0, N] \times [0, N] \times [0, N]$

Initial state: $(N - N_I, 0, N_I, 0)$

Vector Variables: $(\mathsf{S}, \mathsf{E}, \mathsf{I}, \mathsf{R})$

Counting Domain: $[0, N] \times [0, N] \times [0, N] \times [0, N]$

Initial state: $(N - N_I, 0, N_I, 0)$

Transitions:

- $\left(\mathsf{S\_E}, \mathbf{1}_{\mathsf{S}} + \mathbf{1}_{\mathsf{I}}, \mathbf{1}_{\mathsf{E}} + \mathbf{1}_{\mathsf{S}}, \lambda_e \cdot \frac{\mathsf{X}_I}{N} \cdot \mathsf{X}_S\right)$
- $\left(\mathsf{E\_I}, \mathbf{1}_{\mathsf{E}}, \mathbf{1}_{\mathsf{E}}, \lambda_i \cdot \mathsf{X}_I\right)$
- $\left(\mathsf{I\_R}, \mathbf{1}_{\mathsf{I}}, \mathbf{1}_{\mathsf{R}}, \lambda_r \cdot \mathsf{X}_I\right)$

Statistical inference & Simulation...

Example: SEIR Population Model

We can use a simulator to study the fraction of citizens that are *infected* in a given time period:

We can use a simulator to study the fraction of citizens that are *infected* in a given time period:

1. Chose a deadline T;
2. Chose a number of replications n;
3. Chose the sampling time (time points where we register data):

$$D_{t_0}, \ldots, D_{t_k}$$

   where $t_i \in [0, T]$.
4. For each $i \in [0, n]$:
   - $\sigma = \text{simulate}(T)$;
   - $\forall t_i \in \{t_0, \ldots, t_k\} : D_{t_i} = D_{t_i} \oplus \frac{\sigma_S(t_i)}{N}$.
5. Return $\overline{D_{t_0}}, \ldots, \overline{D_{t_k}}$.

From theory to Practice. . .

Example: Leader Election

We want to test a protocol that allow to us to elect a leader among a set of agents.

Example: Leader Election

We want to test a protocol that allow to us to elect a leader among a set of agents.

Naive algorithm:

1. Each active agent randomly select a value in $\{0, 1\}$;
2. An agent survives and start another round if:
   - it has selected 0 and it sees another 0;
   - it has selected 1 and it sees either a 0 or a 1.
3. An agent that has selected 0 that sees a 1 becomes inactive.
4. If an agent does not see any other active agent it becomes the leader.
5. If a leader sees another active agent, it re-start its computation.

Example: Leader Election

Question. . .

Question: How we can build a population model to describe this algorithm?

Example: Leader Election

Population: A, S0, S1, I, L.

Example: Leader Election

Population: A, S0, S1, I, L.

Rules:

- $\mathsf{A} \xrightarrow{\lambda_s \cdot 0.5} \mathsf{S0}$
- $\mathsf{A} \xrightarrow{\lambda_s \cdot 0.5} \mathsf{S1}$
- $\mathsf{S}_1, \mathsf{S}_0 \xrightarrow{\lambda_s \cdot 0.5} \mathsf{C}, \mathsf{S}_0$
- $\mathsf{S}_1, \mathsf{S}_1 \xrightarrow{\lambda_s \cdot 0.5} \mathsf{C}, \mathsf{S}_1$
- $\mathsf{S}_0, \mathsf{S}_0 \xrightarrow{\lambda_s \cdot 0.5} \mathsf{C}, \mathsf{S}_0$
- $\mathsf{S}_0, \mathsf{S}_1 \xrightarrow{\lambda_s \cdot 0.5} \mathsf{F}, \mathsf{S}_1$
- $\mathsf{C} \xrightarrow{\lambda_s \cdot 0.5} \mathsf{L}$
- $\mathsf{L} \xrightarrow{\lambda} \mathsf{C}$

Example: Leader Election

We can use simulation to estimate the number of actives, followers and leaders in the system.

Example: Leader Election

We can use simulation to estimate the number of actives, followers and leaders in the system.

We can also estimate the probability to elect a single leader within T time units.

Example: Leader Election

We can use simulation to estimate the number of actives, followers and leaders in the system.

We can also estimate the probability to elect a single leader within T time units.

This is equivalent to estimate the probability of a Bernulli's distribution (like tossing a coin).

Recall. . .

Let $X_1, \ldots, X_n$ be independent random variables distributed according the same Bernulli's distribution with parameter p, we have that:

$$Pr(|\overline{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$$

Recall. . .

Let $X_1, \dots, X_n$ be independent random variables distributed according the same Bernulli's distribution with parameter p, we have that:

$$Pr(|\overline{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$$

Given a threshold ϵ, and a probability δ, we can use this formula to compute the required number of iterations n to guarantee the probability.

Recall...

Let $X_1, \ldots, X_n$ be independent random variables distributed according the same Bernulli's distribution with parameter p, we have that:

$$Pr(|\overline{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$$

Given a threshold ϵ, and a probability δ, we can use this formula to compute the required number of iterations n to guarantee the probability.

$$Pr(|\overline{X}_n - p| > \epsilon) \leq \delta$$

Recall. . .

Let $X_1, \ldots, X_n$ be independent random variables distributed according the same Bernulli's distribution with parameter p, we have that:

$$Pr(|\overline{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$$

Given a threshold ϵ, and a probability δ, we can use this formula to compute the required number of iterations n to guarantee the probability.

$$Pr(|\overline{X}_n - p| > \epsilon) \leq \delta \quad \Rightarrow \quad 2e^{-2n\epsilon^2} \leq \delta$$

Recall. . .

Let $X_1,\ldots,X_n$ be independent random variables distributed according the same Bernulli's distribution with parameter p, we have that:

$$Pr(|\overline{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$$

Given a threshold ϵ, and a probability δ, we can use this formula to compute the required number of iterations n to guarantee the probability.

$$\begin{aligned} Pr(|\overline{X}_n - p| > \epsilon) \leq \delta \quad &\Rightarrow \quad 2e^{-2n\epsilon^2} \leq \delta \\ &\Leftrightarrow \quad -2n\epsilon^2 \leq \log\left(\frac{\delta}{2}\right) \end{aligned}$$

Recall...

Let $X_1, \ldots, X_n$ be independent random variables distributed according the same Bernulli's distribution with parameter p, we have that:

$$Pr(|\overline{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$$

Given a threshold ϵ, and a probability δ, we can use this formula to compute the required number of iterations n to guarantee the probability.

$$\begin{aligned}
Pr(|\overline{X}_n - p| > \epsilon) \leq \delta & \Rightarrow & 2e^{-2n\epsilon^2} \leq \delta \\
& \Leftrightarrow & -2n\epsilon^2 \leq \log\left(\frac{\delta}{2}\right) \\
& \Leftrightarrow & n \geq \frac{-\log\left(\frac{\delta}{2}\right)}{2\varepsilon^2}
\end{aligned}$$

Recall...

Let $X_1, \ldots, X_n$ be independent random variables distributed according the same Bernulli's distribution with parameter p, we have that:

$$Pr(|\overline{X}_n - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$$

Given a threshold ϵ, and a probability δ, we can use this formula to compute the required number of iterations n to guarantee the probability.

$$\begin{aligned}
Pr(|\overline{X}_n - p| > \epsilon) \leq \delta \quad & \Rightarrow \quad 2e^{-2n\epsilon^2} \leq \delta \\
& \Leftrightarrow \quad -2n\epsilon^2 \leq \log\left(\frac{\delta}{2}\right) \\
& \Leftrightarrow \quad n \geq \frac{-\log\left(\frac{\delta}{2}\right)}{2\varepsilon^2} \\
& \Leftrightarrow \quad n = \left\lceil \frac{\log\left(\frac{2}{\delta}\right)}{2\varepsilon^2} \right\rceil
\end{aligned}$$

Number of iterations. . .

$$n = \left\lceil \frac{\log\left(\frac{2}{\delta}\right)}{2\varepsilon^2} \right\rceil$$

	$\delta = 0.1$	$\delta = 0.01$	$\delta = 0.001$
$\varepsilon = 0.1$	149	265	381
$\varepsilon = 0.01$	14979	26492	38005
$\varepsilon = 0.01$	1497867	2649159	3800452

A path (or *realisation*) π of a stochastic process is a sequence of the form:

$$s_0 t_0 s_1 t_1 \ldots s_n t_n \ldots$$

where each $s_i \in \mathbb{R}^k$.

Transient analysis...

Reachability

A path (or *realisation*) π of a stochastic process is a sequence of the form:

$$s_0 t_0 s_1 t_1 \ldots s_n t_n \ldots$$

where each $s_i \in \mathbb{R}^k$.

Let $\phi, \psi : \mathbb{R}^k \to \{\top, \bot\}$ be two predicates and $t \in \mathbb{R}_{\geq 0}$, we let:

$$\mathcal{R}(\phi, t, \psi) = \{\pi | \exists t' \leq t : \psi(\pi[t']) = \top \land \forall t'' < t' : \phi(\pi[t'']) = \top\}$$

Transient analysis. . .

Reachability

A path (or *realisation*) π of a stochastic process is a sequence of the form:

$$s_0 t_0 s_1 t_1 \ldots s_n t_n \ldots$$

where each $s_i \in \mathbb{R}^k$.

Let $\phi, \psi : \mathbb{R}^k \to \{\top, \bot\}$ be two predicates and $t \in \mathbb{R}_{\geq 0}$, we let:

$$\mathcal{R}(\phi, t, \psi) = \{\pi | \exists t' \leq t : \psi(\pi[t']) = \top \wedge \forall t'' < t' : \phi(\pi[t'']) = \top\}$$

N.B. For any ϕ, t and ψ, $\mathcal{R}(\phi, t, \psi)$ is measurable!

A path (or *realisation*) π of a stochastic process is a sequence of the form:

$$s_0 t_0 s_1 t_1 \ldots s_n t_n \ldots$$

where each $s_i \in \mathbb{R}^k$.

Let $\phi, \psi : \mathbb{R}^k \rightarrow \{\top, \bot\}$ be two predicates and $t \in \mathbb{R}_{\geq 0}$, we let:

$$\mathcal{R}(\phi, t, \psi) = \{\pi | \exists t' \leq t : \psi(\pi[t']) = \top \wedge \forall t'' < t' : \phi(\pi[t'']) = \top\}$$

N.B. For any ϕ, t and ψ, $\mathcal{R}(\phi, t, \psi)$ is measurable!

$Pr(\mathcal{R}(\phi, t, \psi))$ is the probability to reach within t time units a configuration satisfying ψ while only configurations satisfying ϕ are traversed.

Transient analysis...

Reachability

A randomised algorithm can be used to estimate $Pr_{\mathcal{M}}(\phi, t, \psi)$. We are guaranteed that the results differs from the correct one more than δ with a probability that is less or equal to ε:

Transient analysis...

Reachability

A randomised algorithm can be used to estimate $Pr_{\mathcal{M}}(\phi, t, \psi)$. We are guaranteed that the results differs from the correct one more than δ with a probability that is less or equal to ε:

1. $n = \left\lceil \frac{\log\left(\frac{2}{\delta}\right)}{2\varepsilon^2} \right\rceil$
2. $sum = 0$
3. for n times:
   - $\pi = simulate(\mathcal{M})$
   - if $\pi \in \mathcal{R}(\phi, t, \psi)$ then $sum = sum + 1$
4. return $\frac{sum}{n}$

From theory to practice...

Transient analysis. . .
Reachability

We can be also interested in the approximation of the function:

$$\rho_{\psi,\psi}(t) = Pr(\mathcal{R}(\phi, t, \psi))$$

Transient analysis...
Reachability

We can be also interested in the approximation of the function:

$$\rho_{\psi,\psi}(t) = Pr(\mathcal{R}(\phi, t, \psi))$$

Function $\rho_{\psi,\psi}$ describes how the probability to reach ψ (it is the CDF of the random variable $Y \leq t$):

Transient analysis...

Reachability

We can be also interested in the approximation of the function:

$$\rho_{\psi,\psi}(t) = Pr(\mathcal{R}(\phi, t, \psi))$$

Function $\rho_{\psi,\psi}$ describes how the probability to reach ψ (it is the CDF of the random variable $Y \leq t$):

1. $n = \left\lceil \frac{\log\left(\frac{2}{\delta}\right)}{2\varepsilon^2} \right\rceil$
2. $\mathcal{D} = \emptyset$
3. for n times:
   - $\pi = simulate(\mathcal{M}, t)$
   - if $min_{t'} \pi \in \mathcal{R}(\phi, t', \psi) \neq \perp$ then $\mathcal{D} \uplus \{t'\}$
4. return $\lambda x. \frac{|\{t' \leq x | x \in \mathcal{D}|\}}{n}$

Hypothesis testing and reachability...

Hypothesis testing can be used to check if $Pr(\mathcal{R}(\phi, t, \psi)) = p > p_0$. We can consider two relevant claims:

- $p > p_0$
- $p \leq p_0$

Hypothesis testing and reachability. . .

Hypothesis testing can be used to check if $Pr(\mathcal{R}(\phi, t, \psi)) = p > p_0$. We can consider two relevant claims:

- $p > p_0$
- $p \leq p_0$

The null hypothesis is:

- $H_0 : p = p_0$

Hypothesis testing and reachability...

Hypothesis testing can be used to check if $Pr(\mathcal{R}(\phi, t, \psi)) = p > p_0$. We can consider two relevant claims:

- $p > p_0$
- $p \leq p_0$

The null hypothesis is:

- $H_0 : p = p_0$

We specify two alternative hypothesis:

- $H_{+1} : p > p_0$
- $H_{-1} : p < p_0$

Hypothesis testing and reachability...

Warning: H_0 cannot be proved correct since we cannot prove statistically that $p \neq p_0$.

Hypothesis testing and reachability. . .

Warning: H_0 cannot be proved correct since we cannot prove statistically that $p \neq p_0$.

However, we can show that H_0 is incorrect. We can also define a procedure to test which of the alternative hypothesis is true.

Hypothesis testing and reachability...

Warning: H_0 cannot be proved correct since we cannot prove statistically that $p \neq p_0$.

However, we can show that H_0 is incorrect. We can also define a procedure to test which of the alternative hypothesis is true.

We sample N paths $\pi_1, \ldots, \pi_n$ from $\mathcal{M}$.

Hypothesis testing and reachability...

Warning: H_0 cannot be proved correct since we cannot prove statistically that $p \neq p_0$.

However, we can show that H_0 is incorrect. We can also define a procedure to test which of the alternative hypothesis is true.

We sample N paths $\pi_1,\ldots,\pi_n$ from $\mathcal{M}$.

We consider random variables $\mathbf{X} = X_1, \ldots, X_N$ where:

$$X_i = \begin{cases} 1 & \pi_i \in \mathcal{R}(\phi, t, \psi) \\ 0 & \text{otherwise} \end{cases}$$

Hypothesis testing and reachability...

Warning: H_0 cannot be proved correct since we cannot prove statistically that $p \neq p_0$.

However, we can show that H_0 is incorrect. We can also define a procedure to test which of the alternative hypothesis is true.

We sample N paths $\pi_1,\ldots,\pi_n$ from $\mathcal{M}$.

We consider random variables $\mathbf{X} = X_1, \ldots, X_N$ where:

$$X_i = \begin{cases} 1 & \pi_i \in \mathcal{R}(\phi, t, \psi) \\ 0 & \text{otherwise} \end{cases}$$

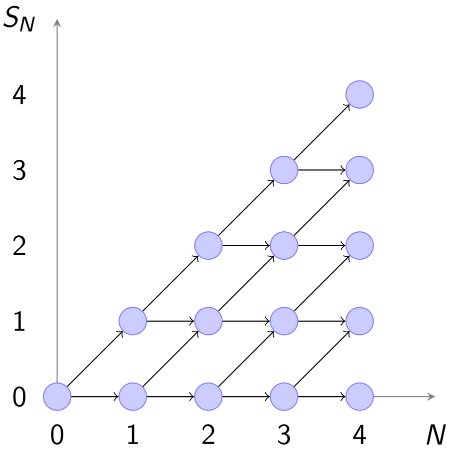
Note that $Pr[X = 1] = p$!

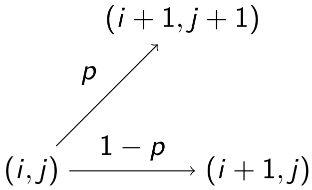
Hypothesis testing and reachability...

We let $S_N(\mathbf{X}) = \sum_{i=1}^{N} X_i$. We can view the evolution of S_N as a DTMC on the space $\mathbb{N} \times \mathbb{N}$:

Hypothesis testing and reachability...

We let $S_N(\mathbf{X}) = \sum_{i=1}^{N} X_i$. We can view the evolution of S_N as a DTMC on the space $\mathbb{N} \times \mathbb{N}$:



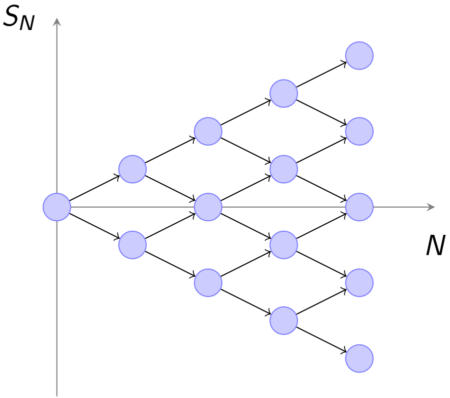


Hypothesis testing and reachability...

We can consider the variable $Z_N = S_N - Np_0$:

Hypothesis testing and reachability...

We can consider the variable $Z_N = S_N - Np_0$:



Hypothesis testing and reachability...

We can observe that...

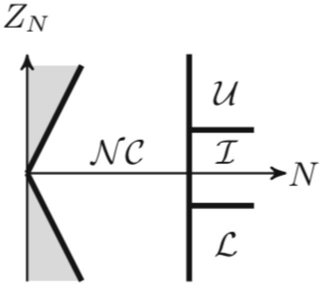
- when $Z_N >> 0$, we have a strong evidence for H_{+1};
- when $Z_N << 0$, we have a strong evidence for H_{-1}.

Hypothesis testing and reachability...

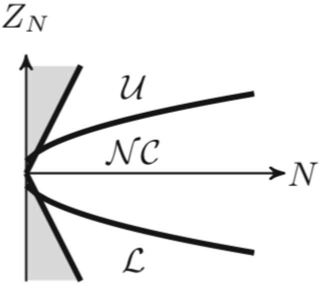
We can observe that...

- when $Z_N >> 0$, we have a strong evidence for H_{+1};
- when $Z_N << 0$, we have a strong evidence for H_{-1}.

More in general we can identify in $\mathbb{R} \times \mathbb{R}$:



Fixed sample size tests



Sequential tests

Hypothesis testing and reachability...

Two kind of tests can be considered:

- Fixed sample size tests: the decision is taken after an a-priori determined number of samples;
- Sequential tests: sampling potentially continue until a decision is reached.

Hypothesis testing and reachability. . .

Two kind of tests can be considered:

- Fixed sample size tests: the decision is taken after an a-priori determined number of samples;
- Sequential tests: sampling potentially continue until a decision is reached.

For sequential tests the goal will be to identify two functions $l(N)$ and $u(N)$ that denote the borders $\mathcal{L} - \mathcal{NC}$ and $\mathcal{U} - \mathcal{NC}$.

Hypothesis testing and reachability...

To provide a measure (in terms of probability) of the error, the following events is considered:

$$
\begin{aligned}
A_{+1} &= \{\text{reach } \mathcal{U} \text{ before } \mathcal{L} \text{ or } \mathcal{I}\} \\
A_{-1} &= \{\text{reach } \mathcal{L} \text{ before } \mathcal{U} \text{ or } \mathcal{I}\} \\
A_{0} &= \{\text{reach } \mathcal{I} \text{ or stay in } \mathcal{NC}\} \\
\neg A_{+1} &= A_{-1} \cup A_{0} \\
\neg A_{-1} &= A_{+1} \cup A_{0}
\end{aligned}
$$

Hypothesis testing and reachability. . .

To provide a measure (in terms of probability) of the error, the following events is considered:

$$
\begin{array}{rcl}
A_{+1} & = & \{\text{reach } \mathcal{U} \text{ before } \mathcal{L} \text{ or } \mathcal{I}\} \\
A_{-1} & = & \{\text{reach } \mathcal{L} \text{ before } \mathcal{U} \text{ or } \mathcal{I}\} \\
A_{0} & = & \{\text{reach } \mathcal{I} \text{ or stay in } \mathcal{NC}\} \\
\neg A_{+1} & = & A_{-1} \cup A_0 \\
\neg A_{-1} & = & A_{+1} \cup A_0
\end{array}
$$

We can impose the following conditions on false positive:

- $Pr(A_{+1} | \neg H_{+1}) \leq \alpha_1$
- $Pr(A_{-1} | \neg H_{-1}) \leq \alpha_2$

Hypothesis testing and reachability. . .

To provide a measure (in terms of probability) of the error, the following events is considered:

$$
\begin{array}{rcl}
A_{+1} & = & \{\text{reach } \mathcal{U} \text{ before } \mathcal{L} \text{ or } \mathcal{I}\} \\
A_{-1} & = & \{\text{reach } \mathcal{L} \text{ before } \mathcal{U} \text{ or } \mathcal{I}\} \\
A_{0} & = & \{\text{reach } \mathcal{I} \text{ or stay in } \mathcal{NC}\} \\
\neg A_{+1} & = & A_{-1} \cup A_0 \\
\neg A_{-1} & = & A_{+1} \cup A_0
\end{array}
$$

We can impose the following conditions on false positive:

- $Pr(A_{+1} | \neg H_{+1}) \leq \alpha_1$
- $Pr(A_{-1} | \neg H_{-1}) \leq \alpha_2$

and false negative:

- $Pr(\neg A_{+1} | H_{+1}) \leq \beta_1$
- $Pr(\neg A_{-1} | H_{-1}) \leq \beta_2$

Hypothesis testing and reachability. . .

To provide a measure (in terms of probability) of the error, the following events is considered:

$$
\begin{array}{rcl}
A_{+1} & = & \{\text{reach } \mathcal{U} \text{ before } \mathcal{L} \text{ or } \mathcal{I}\} \\
A_{-1} & = & \{\text{reach } \mathcal{L} \text{ before } \mathcal{U} \text{ or } \mathcal{I}\} \\
A_{0} & = & \{\text{reach } \mathcal{I} \text{ or stay in } \mathcal{NC}\} \\
\neg A_{+1} & = & A_{-1} \cup A_{0} \\
\neg A_{-1} & = & A_{+1} \cup A_{0}
\end{array}
$$

We can impose the following conditions on false positive:

- $Pr(A_{+1}|\neg H_{+1}) \leq \alpha_1 = \alpha$
- $Pr(A_{-1}|\neg H_{-1}) \leq \alpha_2 = \alpha$

and false negative:

- $Pr(\neg A_{+1}|H_{+1}) \leq \beta_1 = \beta$
- $Pr(\neg A_{-1}|H_{-1}) \leq \beta_2 = \beta$

Criteria...

We will consider three criteria to judge tests:

- the correctness: we call a test correct if its probability of not drawing the correct conclusion is guaranteed to be smaller than α, where $1 - \alpha$ is the confidence level;

Criteria...

We will consider three criteria to judge tests:

- the correctness: we call a test correct if its probability of not drawing the correct conclusion is guaranteed to be smaller than α, where $1 - \alpha$ is the confidence level;

- the power: as the probability that the test will eventually draw a conclusion, that is $1 - Pr(A_0)$;

Criteria...

We will consider three criteria to judge tests:

- the correctness: we call a test correct if its probability of not drawing the correct conclusion is guaranteed to be smaller than α, where $1 - \alpha$ is the confidence level;

- the power: as the probability that the test will eventually draw a conclusion, that is $1 - Pr(A_0)$;

- the efficiency: the number of samples needed (in expectation) before a conclusion can be drawn.

Test classification. . .

Test classification...

- Tests whose probability of drawing a wrong conclusion exceeds α when $|p - p_0|$ is small

Test classification...

I Tests whose probability of drawing a wrong conclusion exceeds α when $|p - p_0|$ is small

II Tests whose probability of drawing no conclusion (or a wrong conclusion) exceeds β when $|p - p_0|$ is small.

Test classification...

I Tests whose probability of drawing a wrong conclusion exceeds α when $|p - p_0|$ is small

II Tests whose probability of drawing no conclusion (or a wrong conclusion) exceeds β when $|p - p_0|$ is small.

III Tests that are always correct and always draw a conclusion, at the cost of drawing an extremely large number of samples before reaching a conclusion when $|p - p_0|$ is small.

Test classification

	Class I	Class II	Class III
Risk when $p \approx p_0$	Correctness: wrong conclusion, i.e., error of first kind (& efficiency)	Power: no conclusion, i.e., error of second kind (& efficiency)	Efficiency: large running time
Parameter	Correctness-indifference level δ	Power-indifference level ζ	Guess γ
Fixed sample size tests	Gauss-SSP	Gauss-CI Chernoff-CI	
Mixed tests		Chow–Robbins	
Sequential tests	SPRT		Azuma Darling

Binomial and Gaussian confidence intervals

The idea behind this test is the confidence interval based on an a priori fixed sample size N.

Binomial and Gaussian confidence intervals

The idea behind this test is the confidence interval based on an a priori fixed sample size N.

We estimate probability p with a confidence interval and check where p_0 is located with respect to this interval.

Binomial and Gaussian confidence intervals

The idea behind this test is the confidence interval based on an a priori fixed sample size N.

We estimate probability p with a confidence interval and check where p_0 is located with respect to this interval.

The Binomial and Gaussian confidence interval (l^*, u^*) is:

- $l^* = \Phi^{-1}(\alpha)\sqrt{N * p_0(1 - p_0)}$
- $u^* = \Phi^{-1}(1 - \alpha)\sqrt{N * p_0(1 - p_0)} = -l^*$

Binomial and Gaussian confidence intervals

binomialCaussianCI($\mathcal{M}$,α,β,p_0,ϕ,t,ψ):

1. $N = f(\beta)$
2. $l^* = \Phi^{-1}(\alpha)\sqrt{N * p_0(1 - p_0)}$
3. $u^* = -l^*$
4. $s = 0$
5. for N times:
    - $\pi = simulate(\mathcal{M})$
    - if $\pi \in \mathcal{R}(\phi, t, \psi)$ then $z = z + 1$
6. if $s - N * p_0 < l^*$ return $\mathcal{L}$
7. if $s - N * p_0 > u^*$ return $\mathcal{U}$
8. return $\mathcal{I}$

Binomial and Gaussian confidence intervals

Choice of N

The choice of $N = f(\beta)$ plays a crucial role in the efficiency of the test.

Binomial and Gaussian confidence intervals

Choice of N

The choice of $N = f(\beta)$ plays a crucial role in the efficiency of the test.

If p is very different from p_0, a small value for N is suffices, and a large value for N is source of inefficiency.

Binomial and Gaussian confidence intervals

Choice of N

The choice of $N = f(\beta)$ plays a crucial role in the efficiency of the test.

If p is very different from p_0, a small value for N is suffices, and a large value for N is source of inefficiency.

If p is close to p_0 large values for N are needed.

Binomial and Gaussian confidence intervals

Choice of N

The choice of $N = f(\beta)$ plays a crucial role in the efficiency of the test.

If p is very different from p_0, a small value for N is suffices, and a large value for N is source of inefficiency.

If p is close to p_0 large values for N are needed.

binomialCaussianCI is a Type II test and we want to guarantee that:

- $Pr(\neg A_{+1} | H_{+1}) \leq \beta$
- $Pr(\neg A_{-1} | H_{-1}) \leq \beta$

Let us assume that $p = p_0 + \zeta$ ($\zeta > 0$).

Binomial and Gaussian confidence intervals
Choice of parameters

Let us assume that $p = p_0 + \zeta$ ($\zeta > 0$).

The probability of not being able to reject H_0 in favour of H_{+1} after drawing N samples is:

$$Pr[\widehat{p_N} \leq p_0 + \Phi^{-1}(1 - \alpha)]$$

Binomial and Gaussian confidence intervals

Choice of parameters

Binomial and Gaussian confidence intervals

Choice of parameters

For large values of N we can assume that $\widehat{p_N} = S_N/N$ is well approximated by a normal distributed random variable with:

- mean $p_0 + \zeta$
- and variance $\sigma^2 = (p_0 + \zeta)(1 - p_0 - \zeta)/N$.

Binomial and Gaussian confidence intervals
Choice of parameters

For large values of N we can assume that $\widehat{p_N} = S_N/N$ is well approximated by a normal distributed random variable with:

- mean $p_0 + \zeta$
- and variance $\sigma^2 = (p_0 + \zeta)(1 - p_0 - \zeta)/N$.

Let $\xi = \Phi^{-1}(1 - \alpha)$ and $\sigma_{H_0}^2 = p_0(1 - p_0)/N$:

Binomial and Gaussian confidence intervals
Choice of parameters

For large values of N we can assume that $\widehat{p_N} = S_N/N$ is well approximated by a normal distributed random variable with:

- mean $p_0 + \zeta$
- and variance $\sigma^2 = (p_0 + \zeta)(1 - p_0 - \zeta)/N$.

Let $\xi = \Phi^{-1}(1 - \alpha)$ and $\sigma^2_{H_0} = p_0(1 - p_0)/N$:

$$\begin{aligned}
Pr[\widehat{p_N} \leq p_0 + \xi\sigma_{H_0}] &= Pr\left[\frac{\widehat{p_N} - p_0 - \zeta}{\sigma} \leq \frac{\xi\sigma_{H_0} - \zeta}{\sigma}\right] \\
&= \Phi\left(\frac{\xi\sigma_{H_0} - \zeta}{\sigma}\right) \\
&= \Phi\left(\frac{\xi\sqrt{p_0(1-p_0)} - \zeta\sqrt{N}}{\sqrt{(p_0+\zeta)(1-p_0-\zeta)}}\right)
\end{aligned}$$

To guarantee that $Pr(\neg A_{+1} | H_{+1}) \leq \beta$ we have:

$$\beta = \Phi\left(\frac{\xi\sqrt{p_0(1-p_0)} - \zeta\sqrt{N}}{\sqrt{(p_0+\zeta)(1-p_0-\zeta)}}\right)$$

$$N_G^+ = \left(\frac{\xi\sqrt{p_0(1-p_0)} - \Phi^{-1}(\beta)\sqrt{(p_0+\zeta)(1-p_0-\zeta)}}{\zeta}\right)^2$$

Analogously, we can compute N_G^- by assuming $p = p_0 - \zeta$ and computing the probability $Pr(\neg A_{-1} | H_{-1})$.

Binomial and Gaussian confidence intervals

Choice of parameters

To guarantee that $Pr(\neg A_{+1} | H_{+1}) \leq \beta$ we have:

$$\beta = \Phi\left(\frac{\xi\sqrt{p_0(1-p_0)} - \zeta\sqrt{N}}{\sqrt{(p_0+\zeta)(1-p_0-\zeta)}}\right)$$

$$N_G^+ = \left(\frac{\xi\sqrt{p_0(1-p_0)} - \Phi^{-1}(\beta)\sqrt{(p_0+\zeta)(1-p_0-\zeta)}}{\zeta}\right)^2$$

Analogously, we can compute N_G^- by assuming $p = p_0 - \zeta$ and computing the probability $Pr(\neg A_{-1} | H_{-1})$.

Finally, we let $N = \max\{N_G^+, N_G^-\}$.

CI using Chernoff-Hoeffding bound

In this test the Chernoff-Hoeffding bound is used to compute the confidence interval:

$$Pr[|\overline{X} - E[\overline{X}]| > \epsilon] \leq 2e^{-2Nt^2}$$

chTest($\mathcal{M}$,α,ϵ,p_0,ϕ,t,ψ)

1. $N = \frac{1}{2\epsilon^2} \log\left(\frac{2}{\alpha}\right)$
2. $s = 0$
3. for N times:
   - $\pi = simulate(\mathcal{M})$
   - if $\pi \in \mathcal{R}(\phi, t, \psi)$ then $z = z + 1$
4. if $s - N * p_0 < -\epsilon$ return $\mathcal{L}$
5. if $s - N * p_0 > \epsilon$ return $\mathcal{U}$
6. return $\mathcal{I}$

CI using Chernoff-Hoeffding bound

Error probability

Let $p = p_0 + \zeta$ ($\zeta > 0$). We have that:

$$
\begin{array}{rcl}
Pr[\overline{X} - p_0 < \varepsilon_N] & = & Pr[p_0 - \overline{X} > -\varepsilon_N] \\
 & = & Pr[p_0 - \overline{X} + \zeta > \zeta - \varepsilon_N] \\
 & \leq & e^{-2N(\zeta - \varepsilon)^2}
\end{array}
$$

CI using Chernoff-Hoeffding bound

Error probability

Let $p = p_0 + \zeta$ $(\zeta > 0)$. We have that:

$$\begin{array}{rcl} Pr[\overline{X} - p_0 < \varepsilon_N] & = & Pr[p_0 - \overline{X} > -\varepsilon_N] \\ & = & Pr[p_0 - \overline{X} + \zeta > \zeta - \varepsilon_N] \\ & \leq & e^{-2N(\zeta - \varepsilon)^2} \end{array}$$

The worst-case number of samples outside the power-indifference region can be computed as:

$$N_C = \frac{2\sqrt{\log(\beta)\log(\alpha)} - \log(\alpha\beta)}{2\zeta^2}$$

Chernoff-Hoeffding vs Gaussian CI

$\alpha = \beta$

α	ζ	$p_0 = 0.5$		$p_0 = 0.2$	
		N_C	N_G	N_C	N_G
0.05	0.1	600	259	600	189
	0.025	9,587	4,199	9,587	2,785
	0.01	59,915	26,265	59,915	17,056
0.025	0.1	738	372	738	273
	0.025	11,805	6,035	11,805	4,012
	0.01	73,778	37,752	73,778	24,540

Sequential probability ratio test

The sequential probability radio test (SPRT) is based on the idea to sequentially test which of the following two hypothesis is true:

- $H_{+1} : p \geq p_{+1}$
- $H_{-1} : p \geq p_{-1}$

where $p_{-1} < p_{+1}$.

Sequential probability ratio test

The sequential probability radio test (SPRT) is based on the idea to sequentially test which of the following two hypothesis is true:

- $H_{+1} : p \geq p_{+1}$
- $H_{-1} : p \geq p_{-1}$

where $p_{-1} < p_{+1}$.

The hypotheses' likelihood ratio T_N is:

$$T_N = \frac{p_{+1}^{S_N}(1 - p_{+1})^{N-S_N}}{p_{-1}^{S_N}(1 - p_{-1})^{N-S_N}}$$

Sequential probability ratio test

The sequential probability radio test (SPRT) is based on the idea to sequentially test which of the following two hypothesis is true:

- $H_{+1} : p \geq p_{+1}$
- $H_{-1} : p \geq p_{-1}$

where $p_{-1} < p_{+1}$.

The hypotheses' likelihood ratio T_N is:

$$T_N = \frac{p_{+1}^{S_N}(1 - p_{+1})^{N-S_N}}{p_{-1}^{S_N}(1 - p_{-1})^{N-S_N}}$$

Small values of T_N speak in favour of H_{-1}, large values of T_N speak in favour of H_{+1}.

Sequential probability ratio test

The idea of the test is to found boundaries u' and l' such that when T_N crosses either of these boundaries we accept the corresponding hypothesis.

Sequential probability ratio test

The idea of the test is to found boundaries u' and l' such that when T_N crosses either of these boundaries we accept the corresponding hypothesis.

If we consider:

- $l' = \alpha_1/(1-\alpha_2)$
- $u' = (1-\alpha_1)/\alpha_2)$

we have that:

- $Pr[H_{-1}|H_{+1}] \leq \alpha_2$
- $Pr[H_{+1}|H_{-1}] \leq \alpha_1$

Sequential probability ratio test

The idea of the test is to found boundaries u' and l' such that when T_N crosses either of these boundaries we accept the corresponding hypothesis.

If we consider:

- $l' = \alpha_1/(1-\alpha_2)$
- $u' = (1-\alpha_1)/\alpha_2$

we have that:

- $Pr[H_{-1}|H_{+1}] \leq \alpha_2$
- $Pr[H_{+1}|H_{-1}] \leq \alpha_1$

We identify an indifference level δ and we set $p_{+1} = p_0 + \delta$ and $p_{-1} = p_0 - \delta$ and:

- $H_{+1} : p \geq p_0 + \delta$
- $H_{-1} : p \geq p_0 - \delta$

Sequential probability ratio test

We can observe that instead of the test statistic T_N we could use:

$$\log T_N = q_1 S_N + q_2 N$$

where:

$$q_1 = \log\left(\frac{p_{+1}(1-p_1)}{(1-p_{+1})p_{-1}}\right) \qquad q_2 = \log\left(\frac{1-p_{+1}}{1-p_{-1}}\right)$$

Sequential probability ratio test

We can observe that instead of the test statistic T_N we could use:

$$\log T_N = q_1 S_N + q_2 N$$

where:

$$q_1 = \log\left(\frac{p_{+1}(1-p_1)}{(1-p_{+1})p_{-1}}\right) \qquad q_2 = \log\left(\frac{1-p_{+1}}{1-p_{-1}}\right)$$

Finally:

$$\begin{aligned} l(N) &= \tfrac{1}{q_1}(\log l' - q_2 N) - N p_0 \\ u(N) &= \tfrac{1}{q_1}(\log u' - q_2 N) - N p_0 \end{aligned}$$

Sequential probability ratio test

sequentialTest$(\mathcal{M},\alpha_1,\alpha_2,\delta,N_m,p_0,\phi,t,\psi)$

1. $q_1 = \ldots$
2. $q_2 = \ldots$
3. $s = 0$
4. $N = 0$
5. $logT = 0$
6. while $N < N_m$ do:
   - $\pi = simulate(\mathcal{M})$
   - if $\pi \in \mathcal{R}(\phi, t, \psi)$ then $s = s + 1$
   - $N = N + 1$
   - $logT = q_1 \cdot s - q_2 \cdot N$
   - if $logT < l(N)$ return $\mathcal{L}$
   - if $logT > u(N)$ return $\mathcal{U}$
7. return $\mathcal{I}$

To be continued. . .