Multi-agent learning

Multi-armed bandit algorithms

Gerard Vreeswijk, Intelligent Software Systems, Computer Science Department, Faculty of Sciences, Utrecht University, The Netherlands.

Tuesday 4th May, 2021

Contents

Contents

- Introduction, motivation, practical applications.

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.
- Simple (but common) MAB algorithms:

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.
- Simple (but common) MAB algorithms:
  - ϵ-Greedy.

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.
- Simple (but common) MAB algorithms:
  - ϵ-Greedy.
  - Q-learning with exploration ϵ and learning rate γ.

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.
- Simple (but common) MAB algorithms:
  - ϵ-Greedy.
  - Q-learning with exploration ϵ and learning rate γ.
  - Boltzmann (a.k.a. Softmax, Gibbs, mixed logit, quantal response) with temperature γ.

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.
- Simple (but common) MAB algorithms:
  - ϵ-Greedy.
  - Q-learning with exploration ϵ and learning rate γ.
  - Boltzmann (a.k.a. Softmax, Gibbs, mixed logit, quantal response) with temperature γ.
  - UCB (upper confidence bound). (Parameterless.)

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.
- Simple (but common) MAB algorithms:
  - ϵ-Greedy.
  - Q-learning with exploration ϵ and learning rate γ.
  - Boltzmann (a.k.a. Softmax, Gibbs, mixed logit, quantal response) with temperature γ.
  - UCB (upper confidence bound). (Parameterless.)
  - Thompson sampling.

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.
- Simple (but common) MAB algorithms:
  - ϵ-Greedy.
  - Q-learning with exploration ϵ and learning rate γ.
  - Boltzmann (a.k.a. Softmax, Gibbs, mixed logit, quantal response) with temperature γ.
  - UCB (upper confidence bound). (Parameterless.)
  - Thompson sampling.
- A well-known MAB algorithm what works well in adversarial circumstances.

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.
- Simple (but common) MAB algorithms:
  - ϵ-Greedy.
  - Q-learning with exploration ϵ and learning rate γ.
  - Boltzmann (a.k.a. Softmax, Gibbs, mixed logit, quantal response) with temperature γ.
  - UCB (upper confidence bound). (Parameterless.)
  - Thompson sampling.
- A well-known MAB algorithm what works well in adversarial circumstances.
  - Exp3 (exponential weight algorithm for exploration and exploitation) with egalitarian factor γ.

Contents

- Introduction, motivation, practical applications.
- Online vs. offline (batch) processing of data.
- Simple (but common) MAB algorithms:
  - ϵ-Greedy.
  - Q-learning with exploration ϵ and learning rate γ.
  - Boltzmann (a.k.a. Softmax, Gibbs, mixed logit, quantal response) with temperature γ.
  - UCB (upper confidence bound). (Parameterless.)
  - Thompson sampling.
- A well-known MAB algorithm what works well in adversarial circumstances.
  - Exp3 (exponential weight algorithm for exploration and exploitation) with egalitarian factor γ.
- Some remarks on the analysis unevenly spaced time series.

MAB algorithms are only interest in rewards per action

Row is protagonist. From

	a	b	c	d	e
A	1, 0	5, 6	1, 0	9, 7	7, 2
B	4, 6	4, 2	1, 8	7, 2	9, 7
C	1, 0	7, 2	9, 7	3, 4	4, 6
D	3, 7	5, 2	5, 3	9, 7	1, 8
E	1, 0	7, 2	4, 6	1, 2	2, 0

MAB algorithms are only interest in rewards per action

Row is protagonist. From

	a	b	c	d	e
A	1, 0	5, 6	1, 0	9, 7	7, 2
B	4, 6	4, 2	1, 8	7, 2	9, 7
C	1, 0	7, 2	9, 7	3, 4	4, 6
D	3, 7	5, 2	5, 3	9, 7	1, 8
E	1, 0	7, 2	4, 6	1, 2	2, 0

to

	a	b	c	d	e
A	1, ?	5, ?	1, ?	9, ?	7, ?
B	4, ?	4, ?	1, ?	7, ?	9, ?
C	1, ?	7, ?	9, ?	3, ?	4, ?
D	3, ?	5, ?	5, ?	9, ?	1, ?
E	1, ?	7, ?	4, ?	1, ?	2, ?

MAB algorithms are only interest in rewards per action

Row is protagonist. From

	a	b	c	d	e
A	1, 0	5, 6	1, 0	9, 7	7, 2
B	4, 6	4, 2	1, 8	7, 2	9, 7
C	1, 0	7, 2	9, 7	3, 4	4, 6
D	3, 7	5, 2	5, 3	9, 7	1, 8
E	1, 0	7, 2	4, 6	1, 2	2, 0

to

	a	b	c	d	e
A	1, ?	5, ?	1, ?	9, ?	7, ?
B	4, ?	4, ?	1, ?	7, ?	9, ?
C	1, ?	7, ?	9, ?	3, ?	4, ?
D	3, ?	5, ?	5, ?	9, ?	1, ?
E	1, ?	7, ?	4, ?	1, ?	2, ?

to

	don't care what the antagonist does
A	reward sequence $r_1, r_2, \dots$ ...
B	reward sequence $r_5, \dots$ ...
C	reward sequence $r_3, r_7, r_8, \dots$ ...
D	reward sequence $r_4, r_9, r_{10}, r_{11}, r_{12}, \dots$ ...
E	reward sequence $r_6, \dots$ ...

Introduction

The multi-armed bandit.
http://en.wikipedia.org/wiki/Multi-armed_bandit

The multi-armed bandit problem

Which slot machine to choose?

MAB problem: random questions

Given. An array of N slot machines.

MAB problem: random questions

Given. An array of N slot machines.

Random questions:

MAB problem: random questions

Given. An array of N slot machines.

Random questions:

1. How long do to stick with a slot machine?

MAB problem: random questions

Given. An array of N slot machines.

Random questions:

1. How long do to stick with a slot machine?
2. Try many machines, or opt for security?

MAB problem: random questions

Given. An array of N slot machines.

Random questions:

1. How long do to stick with a slot machine?
2. Try many machines, or opt for security?
3. Do you exploit success, or do you explore the possibilities?

MAB problem: random questions

Given. An array of N slot machines.

Random questions:

1. How long do to stick with a slot machine?
2. Try many machines, or opt for security?
3. Do you exploit success, or do you explore the possibilities?
4. Is it something we can assume about the distribution of the payouts? Constant mean? Constant variance? Stationary? Does a machine "shift gears" every now and then?

Experiment

	Yield Machine 1	Yield Machine 2	Yield Machine 3
	8	7	20
	8	11	1
	8	8	
	8	9	
	8		

Experiment

	Yield Machine 1	Yield Machine 2	Yield Machine 3
	8	7	20
	8	11	1
	8	8	
	8	9	
	8		
Average	8	8.75	10.5

Exploration vs. exploitation

Exploration vs. exploitation

Problem. You are at the beginning of a new study year. Every fellow student is interesting as a possible new friend.

How do you divide your time between your classmates to optimise your happiness?

Exploration vs. exploitation

Problem. You are at the beginning of a new study year. Every fellow student is interesting as a possible new friend.

How do you divide your time between your classmates to optimise your happiness?

Strategies:

Exploration vs. exploitation

Problem. You are at the beginning of a new study year. Every fellow student is interesting as a possible new friend.

How do you divide your time between your classmates to optimise your happiness?

Strategies:

1. Make friends whe{n|r}ever possible. You are an explorer.

Exploration vs. exploitation

Problem. You are at the beginning of a new study year. Every fellow student is interesting as a possible new friend.

How do you divide your time between your classmates to optimise your happiness?

Strategies:

1. Make friends whe{n|r}ever possible. You are an explorer.
2. Stick to the nearest fellow-student. You are an exploiter.

Exploration vs. exploitation

Problem. You are at the beginning of a new study year. Every fellow student is interesting as a possible new friend.

How do you divide your time between your classmates to optimise your happiness?

Strategies:

1. Make friends whe{n|r}ever possible. You are an explorer.
2. Stick to the nearest fellow-student. You are an exploiter.
3. What most people would do: first explore, then "exploit".

Exploration vs. exploitation

Problem. You are at the beginning of a new study year. Every fellow student is interesting as a possible new friend.

How do you divide your time between your classmates to optimise your happiness?

Strategies:

1. Make friends whe{n|r}ever possible. You are an explorer.
2. Stick to the nearest fellow-student. You are an exploiter.
3. What most people would do: first explore, then "exploit".

We ignore / abstract away from:

Exploration vs. exploitation

Problem. You are at the beginning of a new study year. Every fellow student is interesting as a possible new friend.

How do you divide your time between your classmates to optimise your happiness?

Strategies:

1. Make friends whe{n|r}ever possible. You are an explorer.
2. Stick to the nearest fellow-student. You are an exploiter.
3. What most people would do: first explore, then "exploit".

We ignore / abstract away from:

1. How quality of friendships is measured.

Exploration vs. exploitation

Problem. You are at the beginning of a new study year. Every fellow student is interesting as a possible new friend.

How do you divide your time between your classmates to optimise your happiness?

Strategies:

1. Make friends whe{n|r}ever possible. You are an explorer.
2. Stick to the nearest fellow-student. You are an exploiter.
3. What most people would do: first explore, then "exploit".

We ignore / abstract away from:

1. How quality of friendships is measured.
2. That personalities of friends may change (so-called "non-stationary search").

Other practical problems

Other practical problems

- Select a restaurant from N alternatives.

Other practical problems

- Select a restaurant from N alternatives.
- Select a movie channel from N recommendations.

Other practical problems

- Select a restaurant from N alternatives.
- Select a movie channel from N recommendations.
- Distribute load among servers.

Other practical problems

- Select a restaurant from N alternatives.
- Select a movie channel from N recommendations.
- Distribute load among servers.
- Choose a medical treatment from N alternatives.

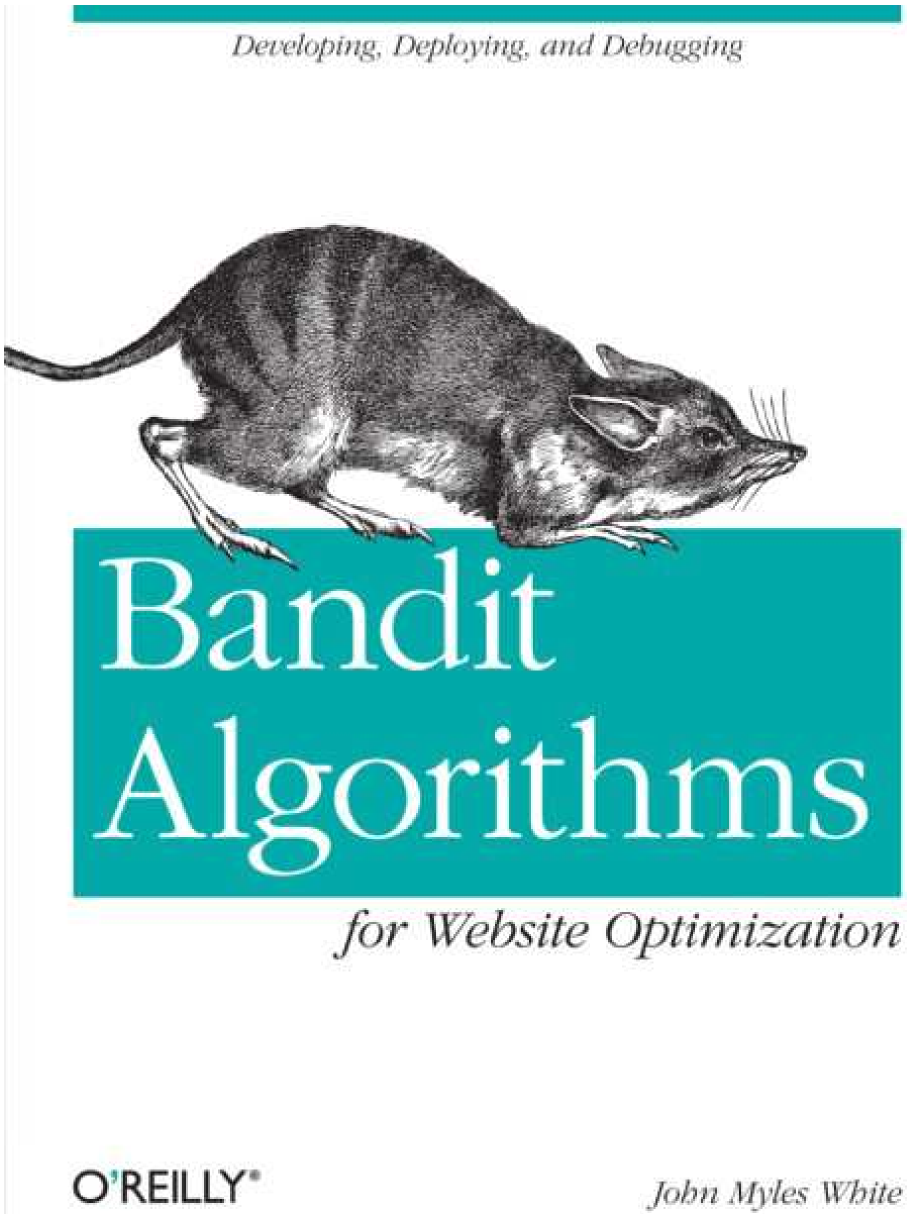
Other practical problems

- Select a restaurant from N alternatives.
- Select a movie channel from N recommendations.
- Distribute load among servers.
- Choose a medical treatment from N alternatives.
- Adaptive routing to optimize network flow.

Other practical problems

- Select a restaurant from N alternatives.
- Select a movie channel from N recommendations.
- Distribute load among servers.
- Choose a medical treatment from N alternatives.
- Adaptive routing to optimize network flow.
- Financial portfolio management.

Other practical problems



- Select a restaurant from N alternatives.
- Select a movie channel from N recommendations.
- Distribute load among servers.
- Choose a medical treatment from N alternatives.
- Adaptive routing to optimize network flow.
- Financial portfolio management.
- ...

Computation of the quality (off line version)

A reasonable measure for the quality of an action a after n tries, Q_n, would be its average payoff:

Formula for the quality of an action after n tries.

$$Q_n =_{Def} \frac{r_1 + \cdots + r_n}{n}$$

Computation of the quality (off line version)

A reasonable measure for the quality of an action a after n tries, Q_n, would be its average payoff:

Formula for the quality of an action after n tries.

$$Q_n =_{Def} \frac{r_1 + \cdots + r_n}{n}$$

Data comes in gradually.

Computation of the quality (off line version)

A reasonable measure for the quality of an action a after n tries, Q_n, would be its average payoff:

Formula for the quality of an action after n tries.

$$Q_n =_{Def} \frac{r_1 + \cdots + r_n}{n}$$

Data comes in gradually.

- This formula is correct. However, every time Q_n is computed, all $r_1, \ldots, r_n$ must be retrieved.

Computation of the quality (off line version)

A reasonable measure for the quality of an action a after n tries, Q_n, would be its average payoff:

Formula for the quality of an action after n tries.

$$Q_n =_{Def} \frac{r_1 + \cdots + r_n}{n}$$

Data comes in gradually.

- This formula is correct. However, every time Q_n is computed, all $r_1, \ldots, r_n$ must be retrieved. This is batch learning.

Computation of the quality (off line version)

A reasonable measure for the quality of an action a after n tries, Q_n, would be its average payoff:

Formula for the quality of an action after n tries.

$$Q_n =_{Def} \frac{r_1 + \cdots + r_n}{n}$$

Data comes in gradually.

- This formula is correct. However, every time Q_n is computed, all $r_1, \dots, r_n$ must be retrieved. This is batch learning.
- It would be better to have an update formula that computes the new average based on the old average and the new incoming value.

Computation of the quality (off line version)

A reasonable measure for the quality of an action a after n tries, Q_n, would be its average payoff:

Formula for the quality of an action after n tries.

$$Q_n =_{Def} \frac{r_1 + \cdots + r_n}{n}$$

Data comes in gradually.

- This formula is correct. However, every time Q_n is computed, all $r_1, \ldots, r_n$ must be retrieved. This is batch learning.
- It would be better to have an update formula that computes the new average based on the old average and the new incoming value. That would be online learning.

Computation of the quality (online version)

$$Q_n$$

Computation of the quality (online version)

$$Q_n = \frac{r_1 + \cdots + r_n}{n}$$

Computation of the quality (online version)

$$Q_n = \frac{r_1 + \cdots + r_n}{n} = \frac{r_1 + \cdots + r_{n-1}}{n} + \frac{r_n}{n}$$

Computation of the quality (online version)

$$
\begin{aligned}
Q_n &= \frac{r_1 + \cdots + r_n}{n} = \frac{r_1 + \cdots + r_{n-1}}{n} + \frac{r_n}{n} \\
&= \frac{r_1 + \cdots + r_{n-1}}{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n}
\end{aligned}
$$

Computation of the quality (online version)

$$
\begin{aligned}
Q_n &= \frac{r_1 + \cdots + r_n}{n} = \frac{r_1 + \cdots + r_{n-1}}{n} + \frac{r_n}{n} \\
&= \frac{r_1 + \cdots + r_{n-1}}{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n}
\end{aligned}
$$

Computation of the quality (online version)

$$\begin{aligned}
Q_n &= \frac{r_1 + \cdots + r_n}{n} = \frac{r_1 + \cdots + r_{n-1}}{n} + \frac{r_n}{n} \\
&= \frac{r_1 + \cdots + r_{n-1}}{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} - \left(\frac{1}{n}\right) Q_{n-1} + \left(\frac{1}{n}\right) r_n
\end{aligned}$$

Computation of the quality (online version)

$$
\begin{aligned}
Q_n &= \frac{r_1 + \cdots + r_n}{n} = \frac{r_1 + \cdots + r_{n-1}}{n} + \frac{r_n}{n} \\
&= \frac{r_1 + \cdots + r_{n-1}}{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} - \left(\frac{1}{n}\right) Q_{n-1} + \left(\frac{1}{n}\right) r_n \\
&= \underbrace{Q_{n-1}}_{\substack{\text{old} \\ \text{value}}}
\end{aligned}
$$

Computation of the quality (online version)

$$
\begin{aligned}
Q_n &= \frac{r_1 + \cdots + r_n}{n} = \frac{r_1 + \cdots + r_{n-1}}{n} + \frac{r_n}{n} \\
&= \frac{r_1 + \cdots + r_{n-1}}{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} - \left(\frac{1}{n}\right) Q_{n-1} + \left(\frac{1}{n}\right) r_n \\
&= \underbrace{Q_{n-1}}_{\substack{\text{old}\\\text{value}}} + \underbrace{\frac{1}{n}}_{\substack{\text{learning}\\\text{rate}}}
\end{aligned}
$$

Computation of the quality (online version)

$$
\begin{aligned}
Q_n &= \frac{r_1 + \cdots + r_n}{n} = \frac{r_1 + \cdots + r_{n-1}}{n} + \frac{r_n}{n} \\
&= \frac{r_1 + \cdots + r_{n-1}}{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} - \left(\frac{1}{n}\right) Q_{n-1} + \left(\frac{1}{n}\right) r_n \\
&= \underbrace{Q_{n-1}}_{\substack{\text{old} \\ \text{value}}} + \underbrace{\frac{1}{n}}_{\substack{\text{learning} \\ \text{rate}}} (\underbrace{r_n}_{\substack{\text{goal} \\ \text{value}}}
\end{aligned}
$$

Computation of the quality (online version)

$$
\begin{aligned}
Q_n &= \frac{r_1 + \cdots + r_n}{n} = \frac{r_1 + \cdots + r_{n-1}}{n} + \frac{r_n}{n} \\
&= \frac{r_1 + \cdots + r_{n-1}}{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} \cdot \frac{n-1}{n} + \frac{r_n}{n} \\
&= Q_{n-1} - \left(\frac{1}{n}\right) Q_{n-1} + \left(\frac{1}{n}\right) r_n \\
&= \underbrace{\underbrace{Q_{n-1}}_{\text{old value}} + \underbrace{\underbrace{\frac{1}{n}}_{\text{learning rate}} \underbrace{(\underbrace{r_n}_{\text{goal value}} - \underbrace{Q_{n-1}}_{\text{old value}})}_{\text{error}}}_{\text{correction}}}_{\text{new value}} .
\end{aligned}
$$

Progress of the quality of one action

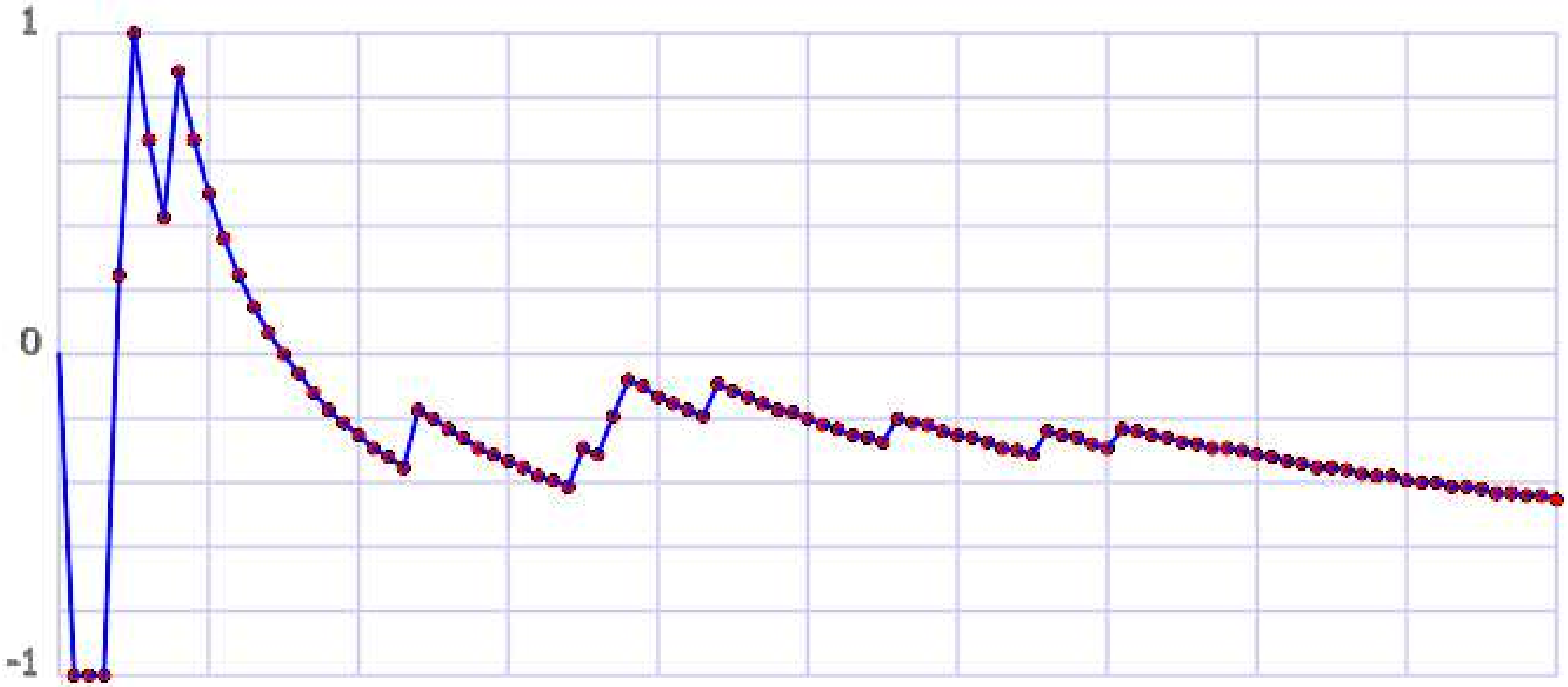
Progress of the quality of one action

- Amplitude of correction is determined by the learning rate.

Progress of the quality of one action

- Amplitude of correction is determined by the learning rate.
- To compute the average, the learning rate is $1/n$ (decreases!).

Progress of the quality of one action



- Amplitude of correction is determined by the learning rate.
- To compute the average, the learning rate is $1/n$ (decreases!).
- Learning rate can also be a constant $0 \leq \lambda \leq 1$

Progress of the quality of one action

- Amplitude of correction is determined by the learning rate.
- To compute the average, the learning rate is $1/n$ (decreases!).
- Learning rate can also be a constant $0 \leq \lambda \leq 1 \Rightarrow$ geometric average.

Action selection: greedy and epsilon-greedy

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- ϵ**-Greedy**: Let $0 < \epsilon \leq 1$ close to 0.

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- ϵ**-Greedy**: Let $0 < \epsilon \leq 1$ close to 0.

  1. Exploitation: choose $(1-\epsilon)\%$ of the time an optimal action.

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- ϵ**-Greedy**: Let $0 < \epsilon \leq 1$ close to 0.

1. Exploitation: choose $(1 - \epsilon)\%$ of the time an optimal action.

2. Exploration: at other times, choose a random action.

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- ϵ**-Greedy**: Let $0 < \epsilon \leq 1$ close to 0.

  1. Exploitation: choose $(1 - \epsilon)\%$ of the time an optimal action.
  2. Exploration: at other times, choose a random action.

  - Because $\sum_{i=1}^{\infty} \epsilon$ is infinite, it follows from the the second Borel-Cantelli lemma that every action is explored infinitely many times

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- ϵ**-Greedy**: Let $0 < \epsilon \leq 1$ close to 0.

  1. Exploitation: choose $(1 - \epsilon)\%$ of the time an optimal action.
  2. Exploration: at other times, choose a random action.

  - Because $\sum_{i=1}^{\infty} \epsilon$ is infinite, it follows from the the second Borel-Cantelli lemma that every action is explored infinitely many times a.s.

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- **ϵ-Greedy**: Let $0 < \epsilon \leq 1$ close to 0.

  1. Exploitation: choose $(1 - \epsilon)\%$ of the time an optimal action.
  2. Exploration: at other times, choose a random action.

  - Because $\sum_{i=1}^{\infty} \epsilon$ is infinite, it follows from the the second Borel-Cantelli lemma that every action is explored infinitely many times a.s.

    So, by the law of large numbers, the estimated value of an action converges to its true value.

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- **ϵ-Greedy**: Let $0 < \epsilon \leq 1$ close to 0.

1. Exploitation: choose $(1 - \epsilon)\%$ of the time an optimal action.
2. Exploration: at other times, choose a random action.

- Because $\sum_{i=1}^{\infty} \epsilon$ is infinite, it follows from the the second Borel-Cantelli lemma that every action is explored infinitely many times a.s.

  So, by the law of large numbers, the estimated value of an action converges to its true value.

- All this a.s.

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- **ϵ-Greedy**: Let $0 < \epsilon \leq 1$ close to 0.

  1. Exploitation: choose $(1 - \epsilon)\%$ of the time an optimal action.
  2. Exploration: at other times, choose a random action.

  - Because $\sum_{i=1}^{\infty} \epsilon$ is infinite, it follows from the the second Borel-Cantelli lemma that every action is explored infinitely many times a.s.

    So, by the law of large numbers, the estimated value of an action converges to its true value.

  - All this a.s. (= with probability 1).

Action selection: greedy and epsilon-greedy

- **Greedy**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- **ϵ-Greedy**: Let $0 < \epsilon \leq 1$ close to 0.

  1. Exploitation: choose $(1 - \epsilon)\%$ of the time an optimal action.

  2. Exploration: at other times, choose a random action.

  - Because $\sum_{i=1}^{\infty} \epsilon$ is infinite, it follows from the the second Borel-Cantelli lemma that every action is explored infinitely many times a.s.

    So, by the law of large numbers, the estimated value of an action converges to its true value.

  - All this a.s. (= with probability 1). In particular it is not certain.

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

- At the outset, an unrealistically high quality is attributed to every slot machine:

  $$Q_0^k = \text{high}$$

  for $1 \leq k \leq N$.

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

- At the outset, an unrealistically high quality is attributed to every slot machine:

$$Q_0^k = \text{high}$$

for $1 \leq k \leq N$.

- As usual, for every slot machine its average profit is maintained.

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

- At the outset, an unrealistically high quality is attributed to every slot machine:

  $$Q_0^k = \text{high}$$

  for $1 \leq k \leq N$.

- As usual, for every slot machine its average profit is maintained.

- Without exception, always exploit machines with highest Q-values.

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

- At the outset, an unrealistically high quality is attributed to every slot machine:

$$Q_0^k = \text{high}$$

for $1 \leq k \leq N$.

- As usual, for every slot machine its average profit is maintained.
- Without exception, always exploit machines with highest Q-values.

Some questions:

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

- At the outset, an unrealistically high quality is attributed to every slot machine:

$$Q_0^k = \text{high}$$

  for $1 \leq k \leq N$.

- As usual, for every slot machine its average profit is maintained.

- Without exception, always exploit machines with highest Q-values.

Some questions:

1. Initially, many actions are tried $\Rightarrow$ all actions are tried?

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

- At the outset, an unrealistically high quality is attributed to every slot machine:

$$Q_0^k = \text{high}$$

  for $1 \leq k \leq N$.

- As usual, for every slot machine its average profit is maintained.

- Without exception, always exploit machines with highest Q-values.

Some questions:

1. Initially, many actions are tried $\Rightarrow$ all actions are tried?

2. How high should "high" be?

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

- At the outset, an unrealistically high quality is attributed to every slot machine:

$$Q_0^k = \text{high}$$

  for $1 \leq k \leq N$.

- As usual, for every slot machine its average profit is maintained.
- Without exception, always exploit machines with highest Q-values.

Some questions:

1. Initially, many actions are tried $\Rightarrow$ all actions are tried?
2. How high should "high" be?
3. Can we still speak of exploration?

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

- At the outset, an unrealistically high quality is attributed to every slot machine:

$$Q_0^k = \text{high}$$

  for $1 \leq k \leq N$.

- As usual, for every slot machine its average profit is maintained.
- Without exception, always exploit machines with highest Q-values.

Some questions:

1. Initially, many actions are tried $\Rightarrow$ all actions are tried?
2. How high should "high" be?
3. Can we still speak of exploration?
4. ϵ-greedy: $\Pr(\text{ every action is explored infinitely many times }) = 1$. Also with optimism?

Action selection: optimistic initial values

An alternative for ϵ-greedy is to work with optimistic initial values.

- At the outset, an unrealistically high quality is attributed to every slot machine:

$$Q_0^k = \text{high}$$

  for $1 \leq k \leq N$.

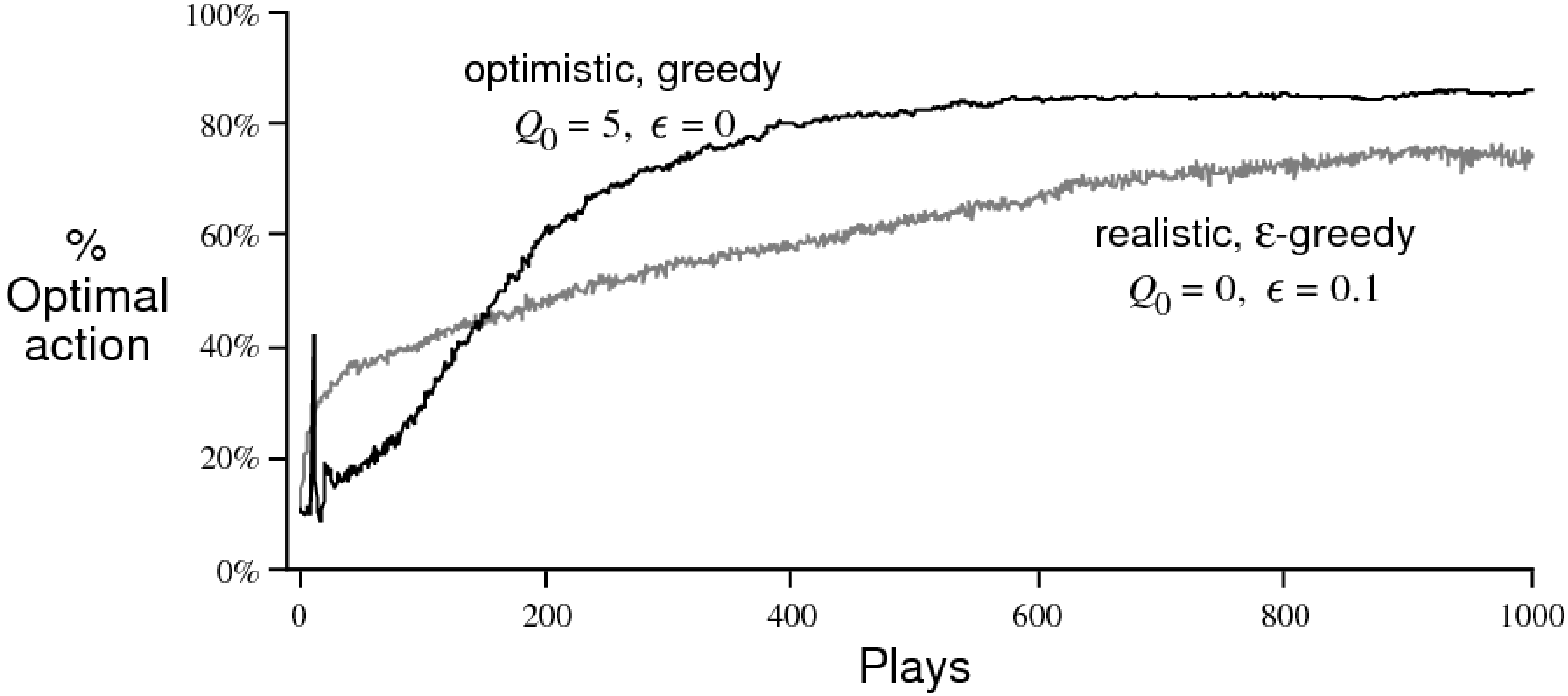
- As usual, for every slot machine its average profit is maintained.

- Without exception, always exploit machines with highest Q-values.

Some questions:

1. Initially, many actions are tried $\Rightarrow$ all actions are tried?
2. How high should "high" be?
3. Can we still speak of exploration?
4. ϵ-greedy: $\Pr(\text{ every action is explored infinitely many times }) = 1$. Also with optimism?
5. Is optimism (as a method) suitable to explore an array of (possibly) infinitely many slot machines? Why (not)?

Optimistic initial values vs. ϵ-greedy



From: "Reinforcement Learning (...)", Sutton and Barto, Sec. 2.8, p. 41.

Q-learning

Q-learning

- Q-learning is like ϵ-Greedy learning, but then with a moving average.
  Algorithm:

Q-learning

- Q-learning is like ϵ-Greedy learning, but then with a moving average.
  Algorithm:

  1. At round t choose an optimal action uniformly with probability $1 - \epsilon$.

Q-learning

- Q-learning is like ϵ-Greedy learning, but then with a moving average.

  Algorithm:

  1. At round t choose an optimal action uniformly with probability $1-\epsilon$.
  2. Update Arm i's estimate at round, $Q_i(t)$, as

$$Q_i(t) = \begin{cases} (1-\lambda)Q_i(t-1) + \lambda r_i, & \text{if Arm } i \text{ is pulled with reward } r_i \\ Q_i(t-1) & \text{otherwise.} \end{cases}$$

Q-learning

- Q-learning is like ϵ-Greedy learning, but then with a moving average.

  Algorithm:

  1. At round t choose an optimal action uniformly with probability $1 - \epsilon$.

  2. Update Arm i's estimate at round, $Q_i(t)$, as

  $$Q_i(t) = \begin{cases} (1-\lambda)Q_i(t-1) + \lambda r_i, & \text{if Arm } i \text{ is pulled with reward } r_i \\ Q_i(t-1) & \text{otherwise.} \end{cases}$$

- Q-learning possesses two parameters: an exploration rate, ϵ, and a learning- or adaptation rate, λ.

Q-learning

- Q-learning is like ϵ-Greedy learning, but then with a moving average.

  Algorithm:

  1. At round t choose an optimal action uniformly with probability $1 - \epsilon$.
  2. Update Arm i's estimate at round, $Q_i(t)$, as

  $$Q_i(t) = \begin{cases} (1-\lambda)Q_i(t-1) + \lambda r_i, & \text{if Arm } i \text{ is pulled with reward } r_i \\ Q_i(t-1) & \text{otherwise.} \end{cases}$$

- Q-learning possesses two parameters: an exploration rate, ϵ, and a learning- or adaptation rate, λ.

  A practical disadvantage of having two parameters is that tuning the algorithm takes more time.

Q-learning

- Q-learning is like ϵ-Greedy learning, but then with a moving average. Algorithm:

  1. At round t choose an optimal action uniformly with probability $1 - \epsilon$.

  2. Update Arm i's estimate at round, $Q_i(t)$, as

$$Q_i(t) = \begin{cases} (1-\lambda)Q_i(t-1) + \lambda r_i, & \text{if Arm } i \text{ is pulled with reward } r_i \\ Q_i(t-1) & \text{otherwise.} \end{cases}$$

- Q-learning possesses two parameters: an exploration rate, ϵ, and a learning- or adaptation rate, λ.

  A practical disadvantage of having two parameters is that tuning the algorithm takes more time.

- Exercise: what if ϵ is small and γ is large?

Q-learning

- Q-learning is like ϵ-Greedy learning, but then with a moving average.

  Algorithm:

  1. At round t choose an optimal action uniformly with probability $1-\epsilon$.
  2. Update Arm i's estimate at round, $Q_i(t)$, as

  $$Q_i(t) = \begin{cases} (1-\lambda)Q_i(t-1) + \lambda r_i, & \text{if Arm } i \text{ is pulled with reward } r_i \\ Q_i(t-1) & \text{otherwise.} \end{cases}$$

- Q-learning possesses two parameters: an exploration rate, ϵ, and a learning- or adaptation rate, λ.

  A practical disadvantage of having two parameters is that tuning the algorithm takes more time.

- Exercise: what if ϵ is small and γ is large? The other way?

Action selection

Action selection

- **Greedily**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

Action selection

- **Greedily**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- **Proportional**: select randomly and proportional to the expected payoff.

$$p_i =_{Def} \frac{Q_i}{\sum_{j=1}^{n} Q_j}.$$

Action selection

- **Greedily**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- **Proportional**: select randomly and proportional to the expected payoff.

$$p_i =_{Def} \frac{Q_i}{\sum_{j=1}^{n} Q_j}.$$

- **Through softmax** (or Boltzmann, or Gibbs, or mixed logit, or quantal response).

Action selection

- **Greedily**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- **Proportional**: select randomly and proportional to the expected payoff.

$$p_i =_{Def} \frac{Q_i}{\sum_{j=1}^{n} Q_j}.$$

- **Through softmax** (or Boltzmann, or Gibbs, or mixed logit, or quantal response).

$$p_i =_{Def} \frac{e^{Q_i/\tau}}{\sum_{j=1}^{n} e^{Q_j/\tau}},$$

Action selection

- **Greedily**: exploit the action that is optimal thus far.

$$p_i =_{Def} \begin{cases} 1 & \text{if } a_i \text{ is optimal thus far,} \\ 0 & \text{otherwise.} \end{cases}$$

- **Proportional**: select randomly and proportional to the expected payoff.

$$p_i =_{Def} \frac{Q_i}{\sum_{j=1}^{n} Q_j}.$$

- **Through softmax** (or Boltzmann, or Gibbs, or mixed logit, or quantal response).

$$p_i =_{Def} \frac{e^{Q_i/\tau}}{\sum_{j=1}^{n} e^{Q_j/\tau}},$$

where the parameter τ is often called the temperature.

Effect of the temperature parameter

The softmax function:

$$p_i =_{Def} \frac{e^{Q_i/\tau}}{\sum_{j=1}^{n} e^{Q_j/\tau}}.$$

Effect of the temperature parameter

The softmax function:

$$p_i =_{Def} \frac{e^{Q_i/\tau}}{\sum_{j=1}^{n} e^{Q_j/\tau}}.$$

This function favours successful actions.

Effect of the temperature parameter

The softmax function:

$$p_i =_{Def} \frac{e^{Q_i/\tau}}{\sum_{j=1}^{n} e^{Q_j/\tau}}.$$

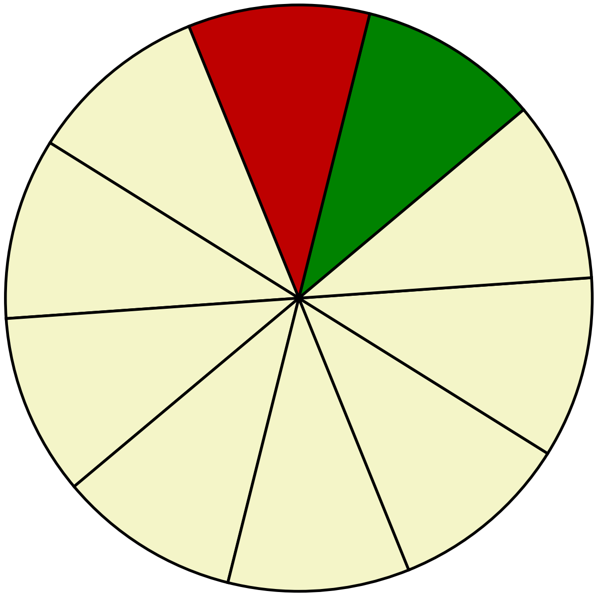
This function favours successful actions. How much depends on τ:

Effect of the temperature parameter

The softmax function:

$$p_i =_{Def} \frac{e^{Q_i/\tau}}{\sum_{j=1}^{n} e^{Q_j/\tau}}.$$

This function favours successful actions. How much depends on τ:



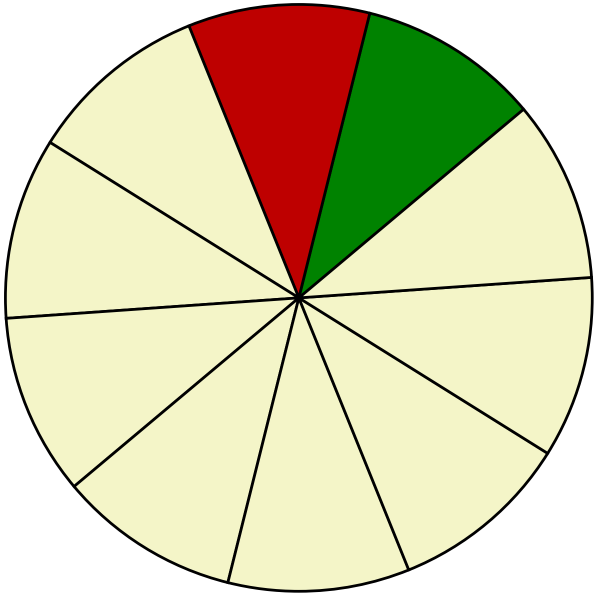
$\tau \to \infty$

Effect of the temperature parameter

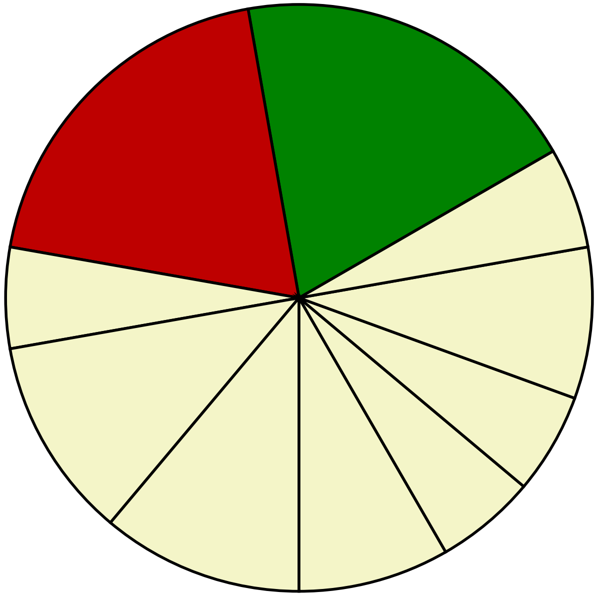
The softmax function:

$$p_i =_{Def} \frac{e^{Q_i/\tau}}{\sum_{j=1}^{n} e^{Q_j/\tau}}.$$

This function favours successful actions. How much depends on τ:



$\tau \to \infty$



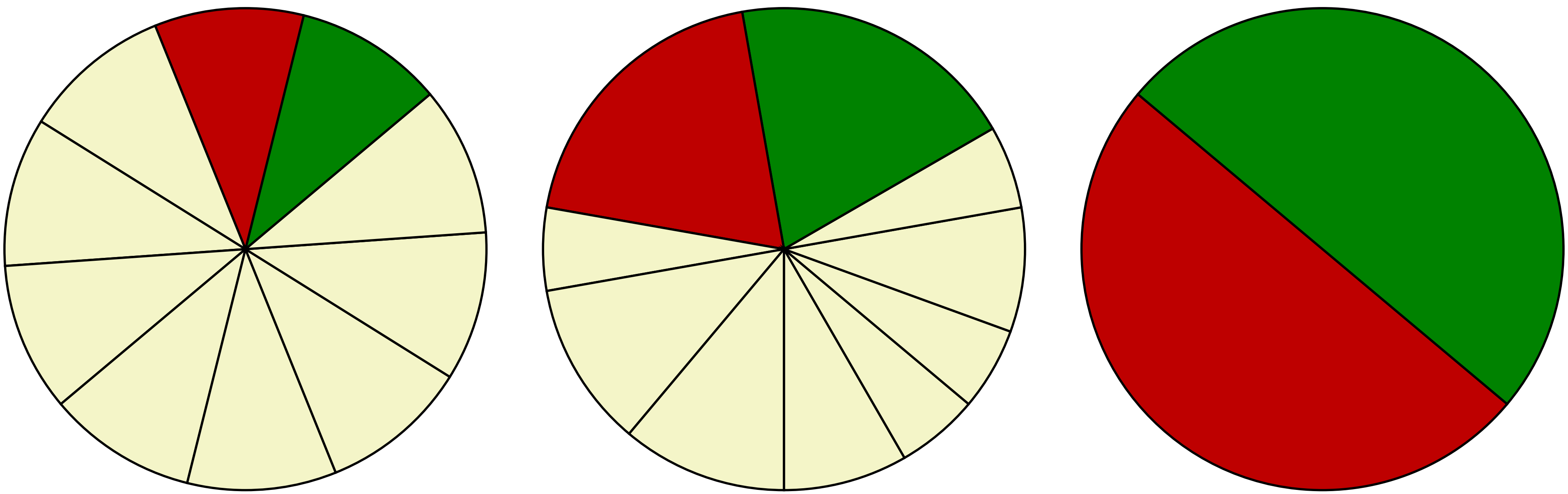
$\tau = 1$

Effect of the temperature parameter

The softmax function:

$$p_i =_{Def} \frac{e^{Q_i/\tau}}{\sum_{j=1}^{n} e^{Q_j/\tau}}.$$

This function favours successful actions. How much depends on τ:



$\tau \to \infty$ $\tau = 1$ $\tau \downarrow 0$

UCB

UCB

- UCB is short for upper confidence bounds.

UCB

- UCB is short for upper confidence bounds.
- Proposed by Lai *et al.* in "Asymptotically efficient adaptive allocation rules" in: *Advances in applied mathematics* Vol. **6**, Nr. 1 (1985), pp. 4-22.

UCB

- UCB is short for upper confidence bounds.
- Proposed by Lai *et al.* in "Asymptotically efficient adaptive allocation rules" in: *Advances in applied mathematics* Vol. **6**, Nr. 1 (1985), pp. 4-22.
- Idea: keep track of confidence intervals for each action. At any time, choose the action in that confidence interval with the highest upper bound.

UCB

- UCB is short for upper confidence bounds.
- Proposed by Lai *et al.* in "Asymptotically efficient adaptive allocation rules" in: *Advances in applied mathematics* Vol. **6**, Nr. 1 (1985), pp. 4-22.
- Idea: keep track of confidence intervals for each action. At any time, choose the action in that confidence interval with the highest upper bound. Often advocated as optimism in the face of uncertainty.

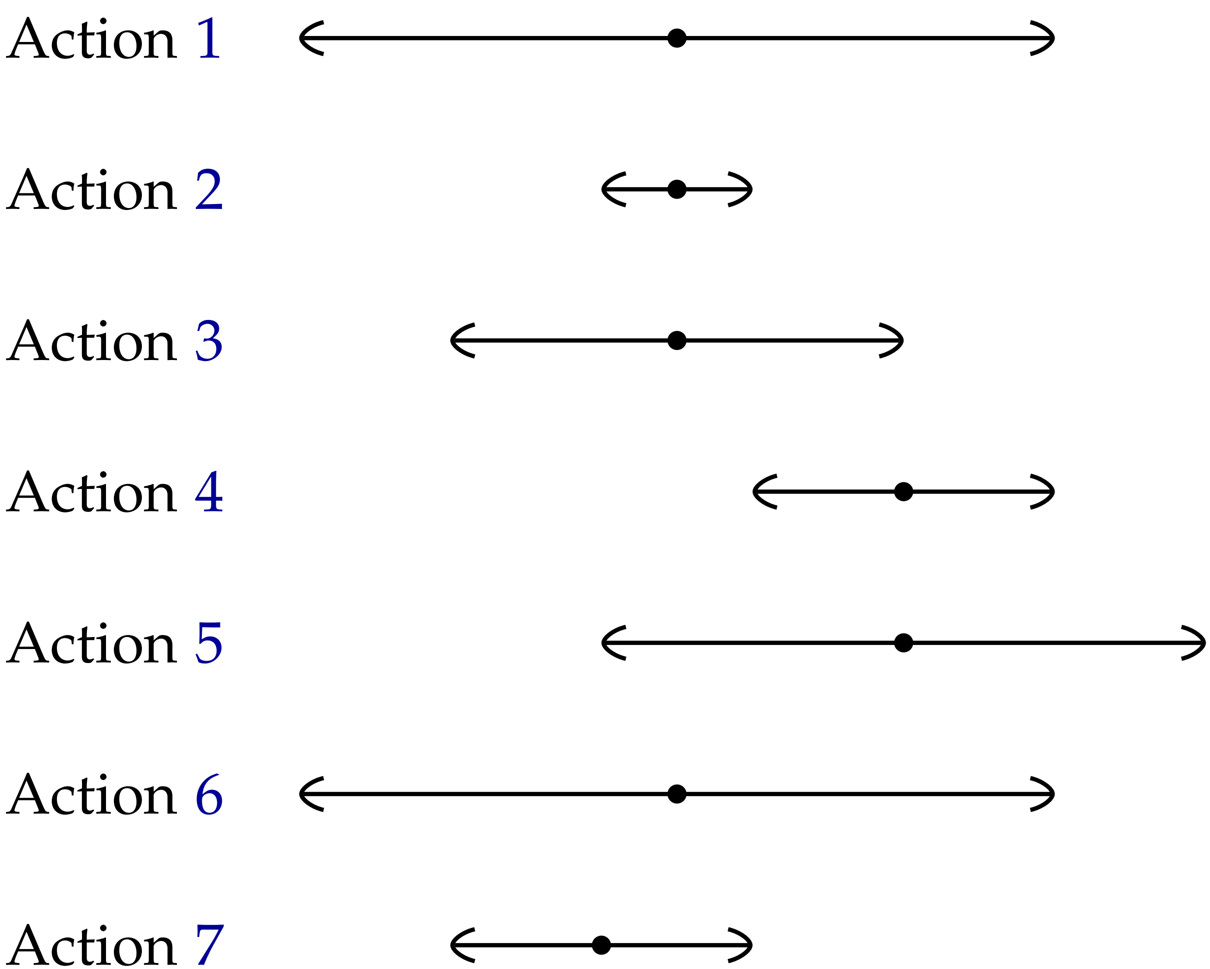
UCB

- UCB is short for upper confidence bounds.
- Proposed by Lai *et al.* in "Asymptotically efficient adaptive allocation rules" in: *Advances in applied mathematics* Vol. **6**, Nr. 1 (1985), pp. 4-22.
- Idea: keep track of confidence intervals for each action. At any time, choose the action in that confidence interval with the highest upper bound. Often advocated as optimism in the face of uncertainty.
- Algorithm: execute each action once. Then, at each round t, choose one of the actions that has highest

$$\bar{X}_t^i + \sqrt{\frac{2\ln(t)}{n_t^i}},$$

where $\bar{X}_t^i$ is action's i average at round t, and n_t^i is the number of times action i is executed at round t.

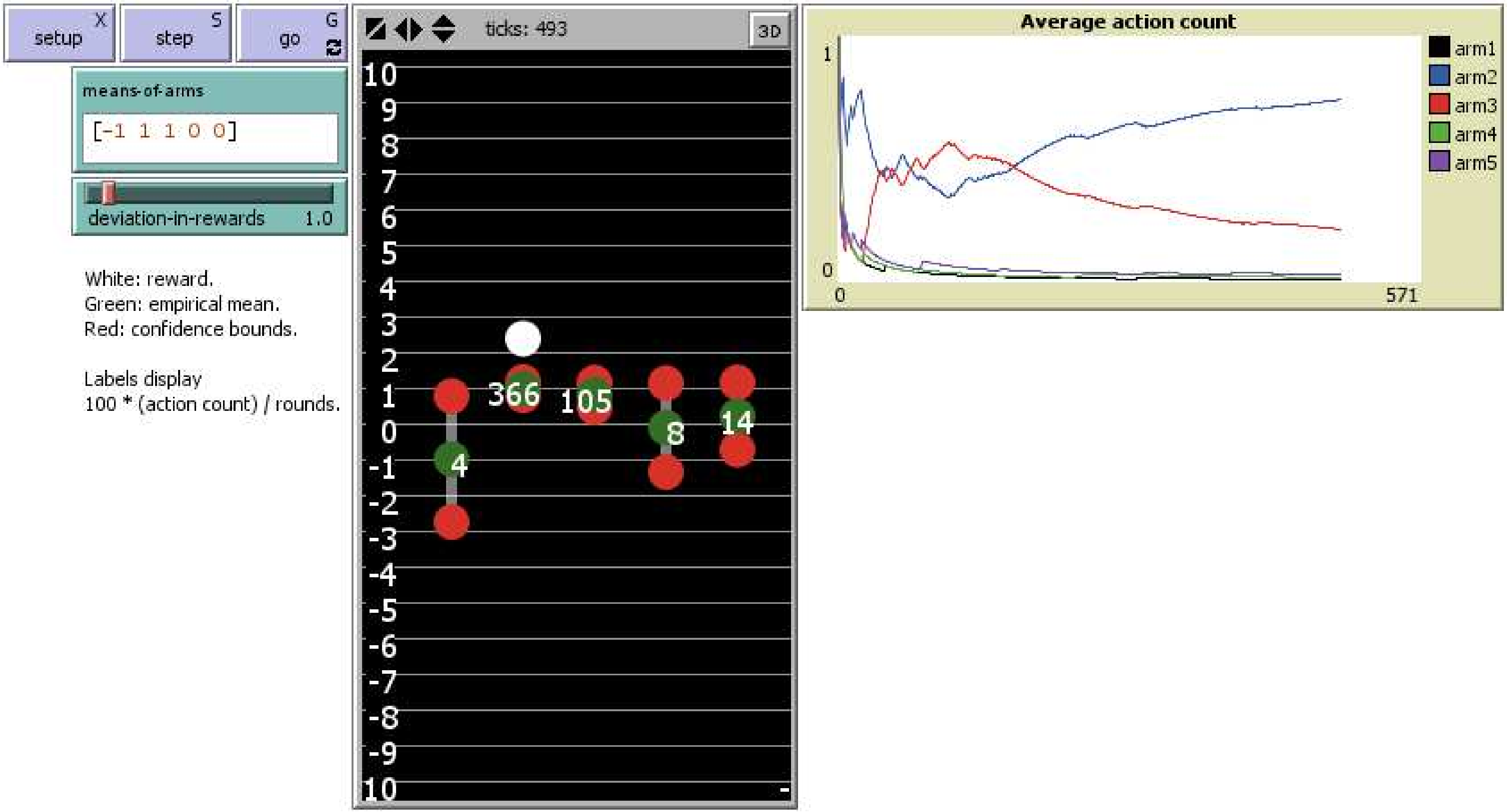
UCB: idea



Many actions have identical empirical means. On the basis of highest empirical mean only, Action 4 and Action 5 would be equally optimal.

However the variation in the rewards of Action 5 is higher, hence its confidence interval is wider, hence its UCB is higher, therefore, choose 5: optimism in the face of uncertainty.

UCB: demo



setup X
step S
go G
ticks: 493
3D
means-of-arms
[-1 1 1 0 0]
deviation-in-rewards 1.0
White: reward.
Green: empirical mean.
Red: confidence bounds.
Labels display
100 * (action count) / rounds.
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
-9
-10
4
366
105
8
14
Average action count
arm1
arm2
arm3
arm4
arm5
1
0
0
571

UCB: derivation

UCB: derivation

- Hoeffding's inequality for i.i.d. random variables $X_1, \dots, X_t$ with mean μ and values in $[0, 1]$ says that, for any $d \geq 0$

$$\Pr\{\mu \geq \bar{X}_t + d\} \leq \exp(-2td^2),$$

where $\bar{X}_t = (X_1, \dots, X_t)/t$ is the empirical mean of the random variables at round t.

UCB: derivation

- Hoeffding's inequality for i.i.d. random variables $X_1, \ldots, X_t$ with mean μ and values in $[0, 1]$ says that, for any $d \geq 0$

$$\Pr\{\mu \geq \bar{X}_t + d\} \leq \exp(-2td^2),$$

  where $\bar{X}_t = (X_1, \ldots, X_t)/t$ is the empirical mean of the random variables at round t.

- For action i this would be $\Pr\{\mu \geq \bar{X}^i_t + d\} \leq \exp(-2n^i_t d^2)$.

UCB: derivation

- Hoeffding's inequality for i.i.d. random variables $X_1, \ldots, X_t$ with mean μ and values in $[0,1]$ says that, for any $d \geq 0$

$$\Pr\{\mu \geq \bar{X}_t + d\} \leq \exp(-2td^2),$$

  where $\bar{X}_t = (X_1, \ldots, X_t)/t$ is the empirical mean of the random variables at round t.

- For action i this would be $\Pr\{\mu \geq \bar{X}^i_t + d\} \leq \exp(-2n^i_t d^2)$.
- The probability that the true mean lies outside $[\bar{X}^i_t - d, \bar{X}^i_t + d]$ goes to zero for $t \to \infty$ if we set $\exp(-2n^i_t d^2)$ equal to an expression that goes to zero if $t \to \infty$.

UCB: derivation

- Hoeffding's inequality for i.i.d. random variables $X_1, \dots, X_t$ with mean μ and values in $[0, 1]$ says that, for any $d \geq 0$

$$\Pr\{\mu \geq \bar{X}_t + d\} \leq \exp(-2td^2),$$

where $\bar{X}_t = (X_1, \dots, X_t)/t$ is the empirical mean of the random variables at round t.

- For action i this would be $\Pr\{\mu \geq \bar{X}^i_t + d\} \leq \exp(-2n^i_t d^2)$.

- The probability that the true mean lies outside $[\bar{X}^i_t - d, \bar{X}^i_t + d]$ goes to zero for $t \to \infty$ if we set $\exp(-2n^i_t d^2)$ equal to an expression that goes to zero if $t \to \infty$. The term t^{-4} is convenient here. Set

$$\exp(-2n^i_t d^2) = t^{-4}.$$

Isolating d yields $d = \sqrt{\frac{2\ln(t)}{n^i_t}}$.

UCB: discussion

UCB: discussion

- All arms are pulled infinitely often. (Hint: what happens with an arm’s UCB if it is not pulled?)

UCB: discussion

- All arms are pulled infinitely often. (Hint: what happens with an arm's UCB if it is not pulled?)
- Let $\mu_1, \dots, \mu_K$ be the means of the distributions of the K arms.

UCB: discussion

- All arms are pulled infinitely often. (Hint: what happens with an arm's UCB if it is not pulled?)
- Let $\mu_1, \ldots, \mu_K$ be the means of the distributions of the K arms.
- Let $I = \text{argmax}_{i=1,\ldots,K}\{\mu_i\}$. So I is the set of indices of all optimal arms.

UCB: discussion

- All arms are pulled infinitely often. (Hint: what happens with an arm's UCB if it is not pulled?)
- Let $\mu_1, \ldots, \mu_K$ be the means of the distributions of the K arms.
- Let $I = \text{argmax}_{i=1,\ldots,K}\{\mu_i\}$. So I is the set of indices of all optimal arms.
- It can be proven that the number of times a sub-optimal arm is played can be bounded from above by $C \ln N$, where C is some constant.

UCB: discussion

- All arms are pulled infinitely often. (Hint: what happens with an arm's UCB if it is not pulled?)
- Let $\mu_1, \ldots, \mu_K$ be the means of the distributions of the K arms.
- Let $I = \text{argmax}_{i=1,\ldots,K}\{\mu_i\}$. So I is the set of indices of all optimal arms.
- It can be proven that the number of times a sub-optimal arm is played can be bounded from above by $C \ln N$, where C is some constant.

  It has also been proven that this is the best possible bound, up to C.

UCB: discussion

- All arms are pulled infinitely often. (Hint: what happens with an arm's UCB if it is not pulled?)
- Let $\mu_1, \ldots, \mu_K$ be the means of the distributions of the K arms.
- Let $I = \text{argmax}_{i=1,\ldots,K}\{\mu_i\}$. So I is the set of indices of all optimal arms.
- It can be proven that the number of times a sub-optimal arm is played can be bounded from above by $C \ln N$, where C is some constant.

  It has also been proven that this is the best possible bound, up to C.
- Bounds can be loose. Suppose $C = 8$ and N is 20. Then $C \ln N = 11.82$. So it is possible to play 11 out of 20 times sub-optimal.

UCB: discussion

- All arms are pulled infinitely often. (Hint: what happens with an arm's UCB if it is not pulled?)
- Let $\mu_1, \ldots, \mu_K$ be the means of the distributions of the K arms.
- Let $I = \text{argmax}_{i=1,\ldots,K}\{\mu_i\}$. So I is the set of indices of all optimal arms.
- It can be proven that the number of times a sub-optimal arm is played can be bounded from above by $C \ln N$, where C is some constant.

  It has also been proven that this is the best possible bound, up to C.
- Bounds can be loose. Suppose $C = 8$ and N is 20. Then $C \ln N = 11.82$. So it is possible to play 11 out of 20 times sub-optimal.
- UCB comes in variants. UCB1 was discussed here.

Thompson sampling (posterior sampling)

Thompson sampling (posterior sampling)

- Idea: for each arm, use a parameterised probability density function as a hypothesis for its rewards.

Thompson sampling (posterior sampling)

- Idea: for each arm, use a parameterised probability density function as a hypothesis for its rewards.

  E.g., for the Bernoulli MAB problem[1] this could be the beta-distributions $\text{Beta}_i(\alpha_i, \beta_i)$, with parameters $\alpha_i, \beta_i > 0$.

Thompson sampling (posterior sampling)

- Idea: for each arm, use a parameterised probability density function as a hypothesis for its rewards.

  E.g., for the Bernoulli MAB problem[1] this could be the beta-distributions $\mathrm{Beta}_i(\alpha_i, \beta_i)$, with parameters $\alpha_i, \beta_i > 0$.

- For every action, start out with neutral distribution, the prior PDF.

Thompson sampling (posterior sampling)

- Idea: for each arm, use a parameterised probability density function as a hypothesis for its rewards.

  E.g., for the Bernoulli MAB problem[1] this could be the beta-distributions $\mathrm{Beta}_i(\alpha_i, \beta_i)$, with parameters $\alpha_i, \beta_i > 0$.

- For every action, start out with neutral distribution, the prior PDF.

  For B-MAB, this should be $\mathrm{Beta}(1, 1)$.

Thompson sampling (posterior sampling)

- Idea: for each arm, use a parameterised probability density function as a hypothesis for its rewards.

  E.g., for the Bernoulli MAB problem[1] this could be the beta-distributions $\text{Beta}_i(\alpha_i, \beta_i)$, with parameters $\alpha_i, \beta_i > 0$.

- For every action, start out with neutral distribution, the prior PDF.

  For B-MAB, this should be $\text{Beta}(1, 1)$.

- At every round, sample all hypotheses, and choose the arm which is sampled highest.

Thompson sampling (posterior sampling)

- Idea: for each arm, use a parameterised probability density function as a hypothesis for its rewards.

  E.g., for the Bernoulli MAB problem[1] this could be the beta-distributions $\text{Beta}_i(\alpha_i, \beta_i)$, with parameters $\alpha_i, \beta_i > 0$.

- For every action, start out with neutral distribution, the prior PDF.

  For B-MAB, this should be $\text{Beta}(1, 1)$.

- At every round, sample all hypotheses, and choose the arm which is sampled highest.

  B-MAB: the mean of $\text{Beta}(\alpha, \beta)$ is $1/(1 + \beta/\alpha)$.

Thompson sampling (posterior sampling)

- Idea: for each arm, use a parameterised probability density function as a hypothesis for its rewards.

  E.g., for the Bernoulli MAB problem[1] this could be the beta-distributions $\text{Beta}_i(\alpha_i, \beta_i)$, with parameters $\alpha_i, \beta_i > 0$.

- For every action, start out with neutral distribution, the prior PDF.

  For B-MAB, this should be $\text{Beta}(1, 1)$.

- At every round, sample all hypotheses, and choose the arm which is sampled highest.

  B-MAB: the mean of $\text{Beta}(\alpha, \beta)$ is $1/(1 + \beta/\alpha)$.

- Update: compute the action's posterior PDF by letting the reward change the parameters of its associated PDF, through a Bayesian update.

Thompson sampling (posterior sampling)

- Idea: for each arm, use a parameterised probability density function as a hypothesis for its rewards.

  E.g., for the Bernoulli MAB problem[1] this could be the beta-distributions $\text{Beta}_i(\alpha_i, \beta_i)$, with parameters $\alpha_i, \beta_i > 0$.

- For every action, start out with neutral distribution, the prior PDF.

  For B-MAB, this should be $\text{Beta}(1, 1)$.

- At every round, sample all hypotheses, and choose the arm which is sampled highest.

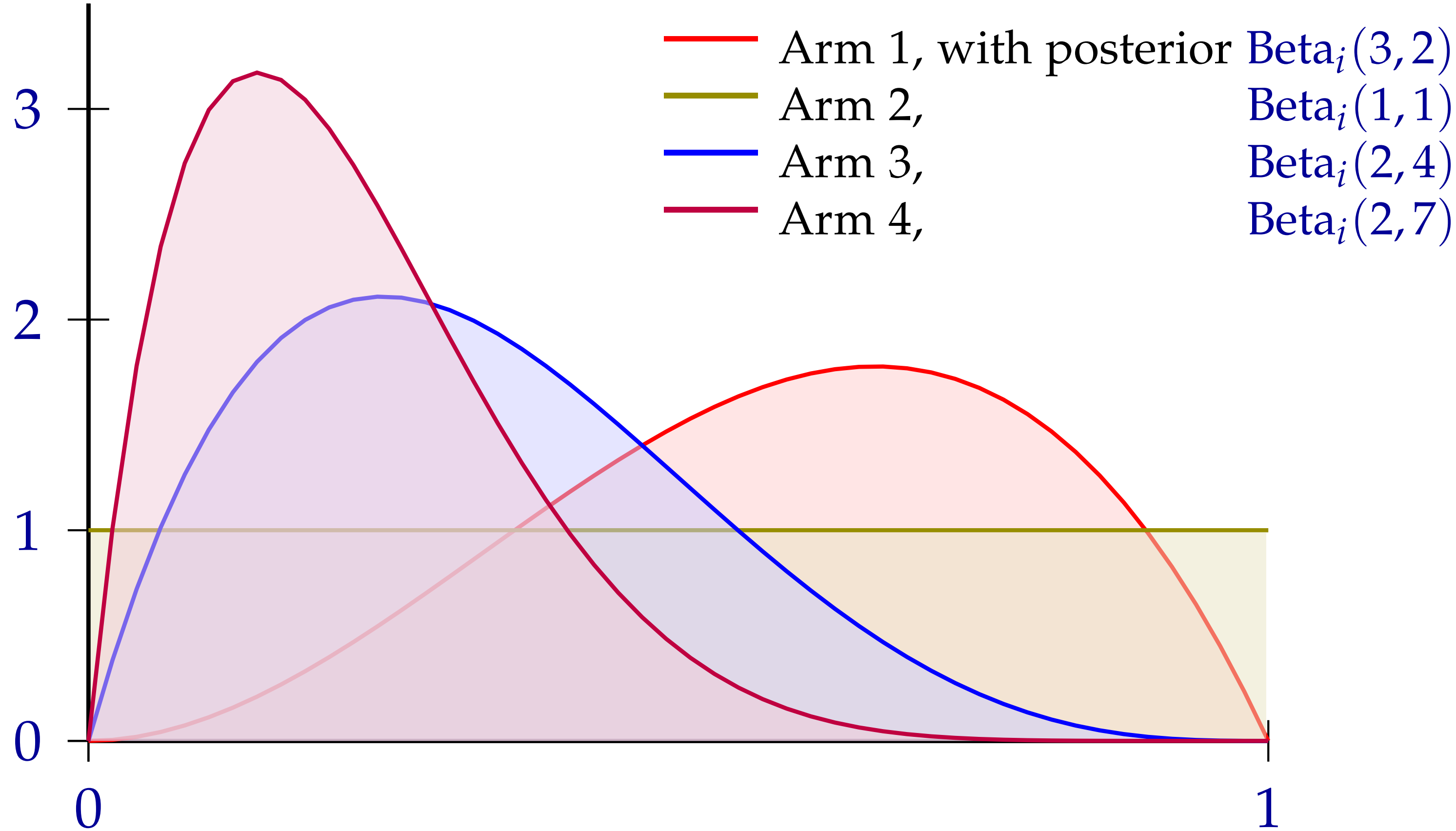
  B-MAB: the mean of $\text{Beta}(\alpha, \beta)$ is $1/(1 + \beta/\alpha)$.

- Update: compute the action's posterior PDF by letting the reward change the parameters of its associated PDF, through a Bayesian update.

  B-MAB: if arm 5 pays 0, do $\beta_5 = \beta_5 + 1$ for the associated PDF. (If arm 5 pays 1, do $\alpha_5 = \alpha_5 + 1$.)

[1] Each arm i is associated with the outcome of tossing a biased coin (heads = 1, tails = 0) with fixed (and hidden) bias $0 \leq \theta_i \leq 1$.

Thompson sampling on a Bernoulli bandit



Posterior PDFs after pulling Arm 1 three times with two successes, Arm 2 zero times, Arm 3 four times with one success, Arm 4 seven times with one success. (Notice: $\alpha = \#\text{successes} + 1$, $\beta = \#\text{failures} + 1$.)

Adversarial bandits

Adversarial bandits

- Adversarial behaviour seems to be the norm in multi-agent learning: opponents may change strategies at any time (called shifting gears in poker), for example to be unpredictable.

Adversarial bandits

- Adversarial behaviour seems to be the norm in multi-agent learning: opponents may change strategies at any time (called shifting gears in poker), for example to be unpredictable.

  Think of anti-coordination games, such as chicken, or matching pennies.

Adversarial bandits

- Adversarial behaviour seems to be the norm in multi-agent learning: opponents may change strategies at any time (called shifting gears in poker), for example to be unpredictable.

  Think of anti-coordination games, such as chicken, or matching pennies.

- An adversarial bandit (Auer and Cesa-Bianchi, 1998) may use any reward algorithm.

Adversarial bandits

- Adversarial behaviour seems to be the norm in multi-agent learning: opponents may change strategies at any time (called shifting gears in poker), for example to be unpredictable.

  Think of anti-coordination games, such as chicken, or matching pennies.

- An adversarial bandit (Auer and Cesa-Bianchi, 1998) may use any reward algorithm.

- In the worst case, an adversarial bandit is benign. If it is smart [omniscient], it may frustrate a learning algorithm [to the max].

Adversarial bandits

- Adversarial behaviour seems to be the norm in multi-agent learning: opponents may change strategies at any time (called shifting gears in poker), for example to be unpredictable.

  Think of anti-coordination games, such as chicken, or matching pennies.

- An adversarial bandit (Auer and Cesa-Bianchi, 1998) may use any reward algorithm.

- In the worst case, an adversarial bandit is benign. If it is smart [omniscient], it may frustrate a learning algorithm [to the max].

  Small exercise: suppose your are a benign adversarial playing against against a fictitious player who plays row. How would you play?

	H	T
H	$1, 0$	$-1, 0$
T	$-1, 0$	$1, 0$

Adversarial bandits

- Adversarial behaviour seems to be the norm in multi-agent learning: opponents may change strategies at any time (called shifting gears in poker), for example to be unpredictable.

  Think of anti-coordination games, such as chicken, or matching pennies.

- An adversarial bandit (Auer and Cesa-Bianchi, 1998) may use any reward algorithm.

- In the worst case, an adversarial bandit is benign. If it is smart [omniscient], it may frustrate a learning algorithm [to the max].

  Small exercise: suppose your are a benign adversarial playing against against a fictitious player who plays row. How would you play?

	H	T
H	$1, 0$	$-1, 0$
T	$-1, 0$	$1, 0$

- "Cumulative" algorithms, like ϵ-Greedy, UCB, or Thompson, respond slowly to sudden changes.

Exp3

Exp3

- Exp3 is short for exponential weight algorithm for exploration and exploitation.

Exp3

- Exp3 is short for exponential weight algorithm for exploration and exploitation.
- Proposed by Auer *et al*. (2002) in "The nonstochastic multiarmed bandit problem", in *Journal on Computing* Vol. **32**, Nr. 1.

Exp3

- Exp3 is short for exponential weight algorithm for exploration and exploitation.
- Proposed by Auer *et al.* (2002) in "The nonstochastic multiarmed bandit problem", in *Journal on Computing* Vol. **32**, Nr. 1.
- Exp3 may be seen as a volatile Softmax.

Exp3

- Exp3 is short for exponential weight algorithm for exploration and exploitation.
- Proposed by Auer *et al.* (2002) in "The nonstochastic multiarmed bandit problem", in *Journal on Computing* Vol. **32**, Nr. 1.
- Exp3 may be seen as a volatile Softmax.
- It is an adversarial algorithm, meaning that it should perform well in environments where payoffs for actions may suddenly change.

Exp3

- Exp3 is short for exponential weight algorithm for exploration and exploitation.
- Proposed by Auer *et al.* (2002) in "The nonstochastic multiarmed bandit problem", in *Journal on Computing* Vol. **32**, Nr. 1.
- Exp3 may be seen as a volatile Softmax.
- It is an adversarial algorithm, meaning that it should perform well in environments where payoffs for actions may suddenly change.
- Idea: maintain a vector of action weights $(w_1, \dots, w_K)$. Actions are chosen probabilistically, proportional to their weights:

$$p_k(t) =_{Def} (1-\gamma)\frac{w_k(t)}{\sum_{i=1}^{k} w_i(t)} + \gamma\frac{1}{K}$$

where $0 \leq \gamma \leq 1$ is the egalitarianism factor (check what if $\gamma = 0, 1$).

Exp3 (continued)

So far so good.

Exp3 (continued)

So far so good. But how are the weights computed?

Exp3 (continued)

So far so good. But how are the weights computed? If

$$\hat{r}_i(t) =_{Def} \begin{cases} \dfrac{r_i(t)}{p_i(t)} & \text{if } i \text{ is chosen at } t, \\ 0 & \text{otherwise.} \end{cases}$$

denotes the estimated reward, i.e., the reward of action i weighed by its surprise (i.e., multiplied by the reciprocal of its probability to occur), then weights are computed as

$$w_i(t+1) =_{Def} w_i(t) \exp\left(\gamma \frac{1}{K} \hat{r}_i(t)\right)$$

Exp3 (continued)

So far so good. But how are the weights computed? If

$$\hat{r}_i(t) =_{Def} \begin{cases} \dfrac{r_i(t)}{p_i(t)} & \text{if } i \text{ is chosen at } t, \\ 0 & \text{otherwise.} \end{cases}$$

denotes the estimated reward, i.e., the reward of action i weighed by its surprise (i.e., multiplied by the reciprocal of its probability to occur), then weights are computed as

$$w_i(t+1) =_{Def} w_i(t) \exp\left(\gamma \frac{1}{K} \hat{r}_i(t)\right)$$

Important: rewards are supposed to lie in $[0,1]$. (Scale payoffs if necessary.)

Exp3 (continued)

So far so good. But how are the weights computed? If

$$\hat{r}_i(t) =_{Def} \begin{cases} \dfrac{r_i(t)}{p_i(t)} & \text{if } i \text{ is chosen at } t, \\ 0 & \text{otherwise.} \end{cases}$$

denotes the estimated reward, i.e., the reward of action i weighed by its surprise (i.e., multiplied by the reciprocal of its probability to occur), then weights are computed as

$$w_i(t+1) =_{Def} w_i(t) \exp\left(\gamma \frac{1}{K} \hat{r}_i(t)\right)$$

Important: rewards are supposed to lie in $[0,1]$. (Scale payoffs if necessary.)

Exp3 is a so-called weak no-regret algorithm, which means that the average regrets are pressed out a.s.

Stationary time series

Stationary time series

A time series $\{X_t\}_t$

Stationary time series

A time series $\{X_t\}_t$ is called

Stationary time series

A time series $\{X_t\}_t$ is called

- strictly stationary if, for each fixed h, the vectors $(X_t, X_{t+1}, \dots, X_{t+h})$ are identically distributed for all t.

Stationary time series

A time series $\{X_t\}_t$ is called

- strictly stationary if, for each fixed h, the vectors $(X_t, X_{t+1}, \dots, X_{t+h})$ are identically distributed for all t.
- weakly stationary of order two if it has a fixed mean, and, for each fixed h, the covariances $\mathrm{Cov}(X_t, X_{t+h})$ are equal for all t.

Stationary time series

A time series $\{X_t\}_t$ is called

- strictly stationary if, for each fixed h, the vectors $(X_t, X_{t+1}, \ldots, X_{t+h})$ are identically distributed for all t.
- weakly stationary of order two if it has a fixed mean, and, for each fixed h, the covariances $\text{Cov}(X_t, X_{t+h})$ are equal for all t.

Strict: all statistics are time-independent.

Stationary time series

A time series $\{X_t\}_t$ is called

- strictly stationary if, for each fixed h, the vectors $(X_t, X_{t+1}, \ldots, X_{t+h})$ are identically distributed for all t.
- weakly stationary of order two if it has a fixed mean, and, for each fixed h, the covariances $\mathrm{Cov}(X_t, X_{t+h})$ are equal for all t.

Strict: all statistics are time-independent. (Unrealistic.)

Stationary time series

A time series $\{X_t\}_t$ is called

- strictly stationary if, for each fixed h, the vectors $(X_t, X_{t+1}, \dots, X_{t+h})$ are identically distributed for all t.
- weakly stationary of order two if it has a fixed mean, and, for each fixed h, the covariances $\mathrm{Cov}(X_t, X_{t+h})$ are equal for all t.

Strict: all statistics are time-independent. (Unrealistic.) Weak: constant mean, variance and covariance. Other statistics are free to change.

Stationary time series

A time series $\{X_t\}_t$ is called

- strictly stationary if, for each fixed h, the vectors $(X_t, X_{t+1}, \dots, X_{t+h})$ are identically distributed for all t.
- weakly stationary of order two if it has a fixed mean, and, for each fixed h, the covariances $\mathrm{Cov}(X_t, X_{t+h})$ are equal for all t.

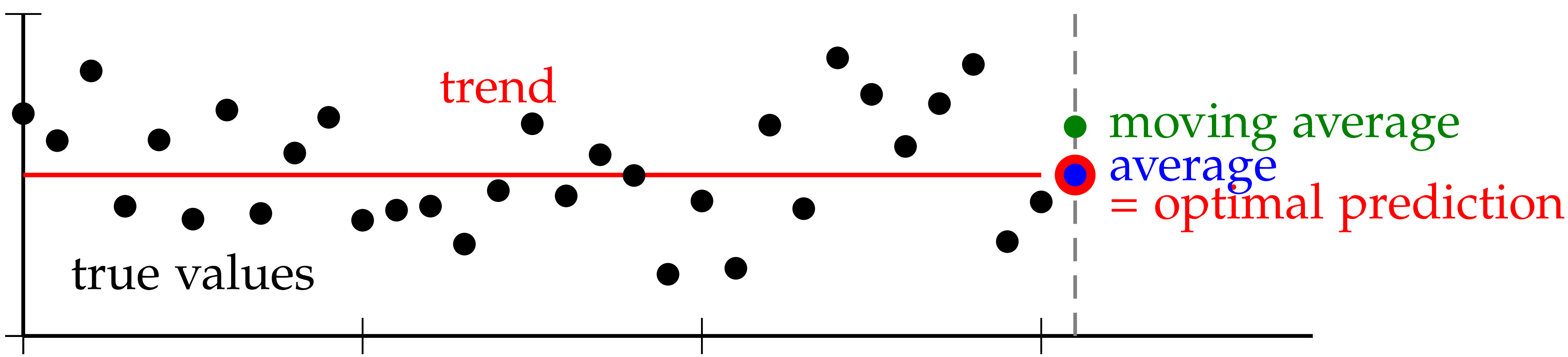
Strict: all statistics are time-independent. (Unrealistic.) Weak: constant mean, variance and covariance. Other statistics are free to change. (Common.)

Stationary time series

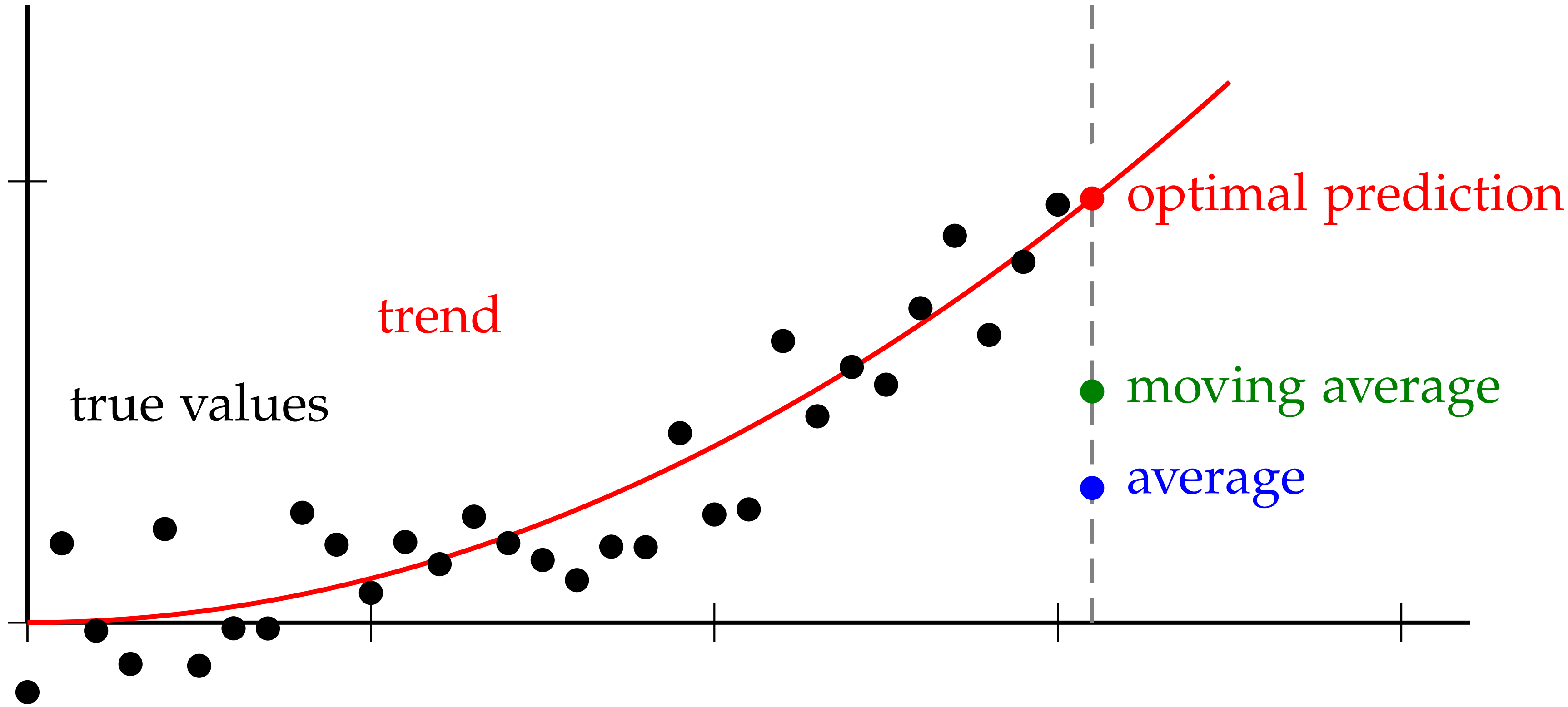
A time series $\{X_t\}_t$ is called

- strictly stationary if, for each fixed h, the vectors $(X_t, X_{t+1}, \dots, X_{t+h})$ are identically distributed for all t.
- weakly stationary of order two if it has a fixed mean, and, for each fixed h, the covariances $\mathrm{Cov}(X_t, X_{t+h})$ are equal for all t.

Strict: all statistics are time-independent. (Unrealistic.) Weak: constant mean, variance and covariance. Other statistics are free to change. (Common.)

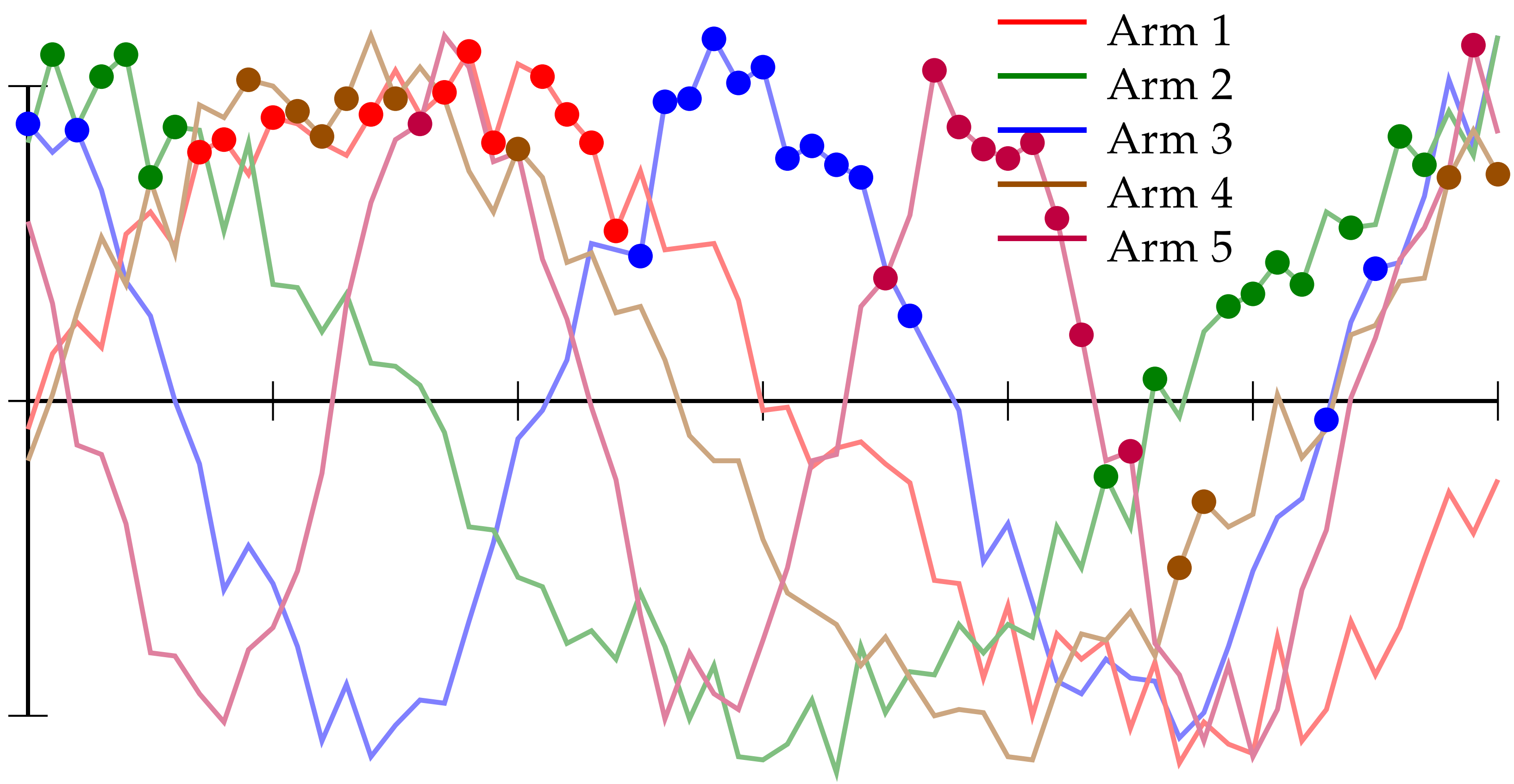


Non-stationary time series



With non-stationary time series, the average (a.k.a. empirical mean) $(n-1)/nT + r/n$ and moving average (a.k.a. rolling average, geometric mean, or exponentially smoothed mean) $(1-\gamma)T + \gamma r$ are bad predictors.

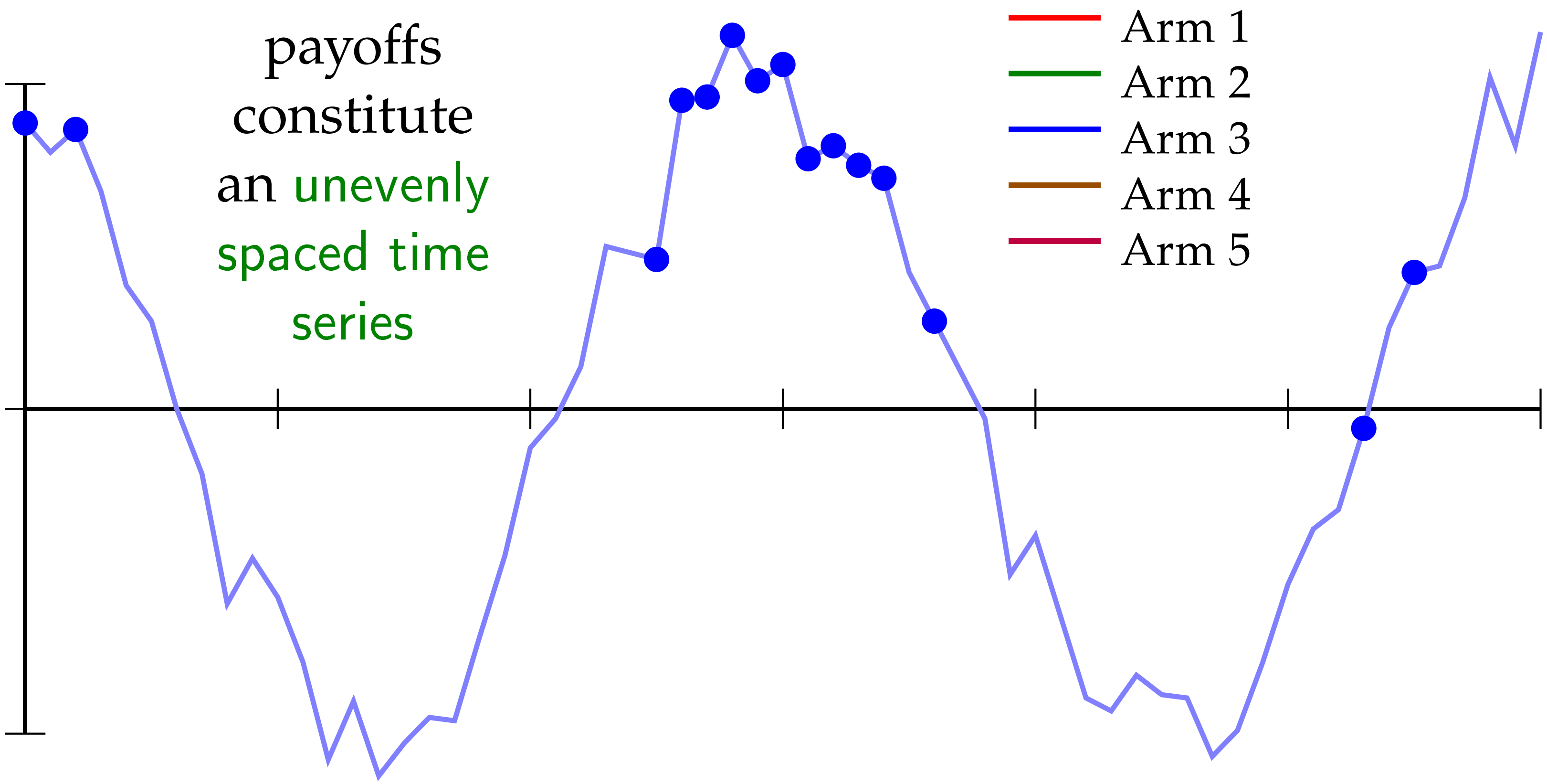
MAB with non-stationary payoffs



True payments per arm per round, had arm been pulled (solid lines).
Received payments, per arm per round (points).

Due to prediction errors, sometimes a “wrong” arm is pulled (verify!).

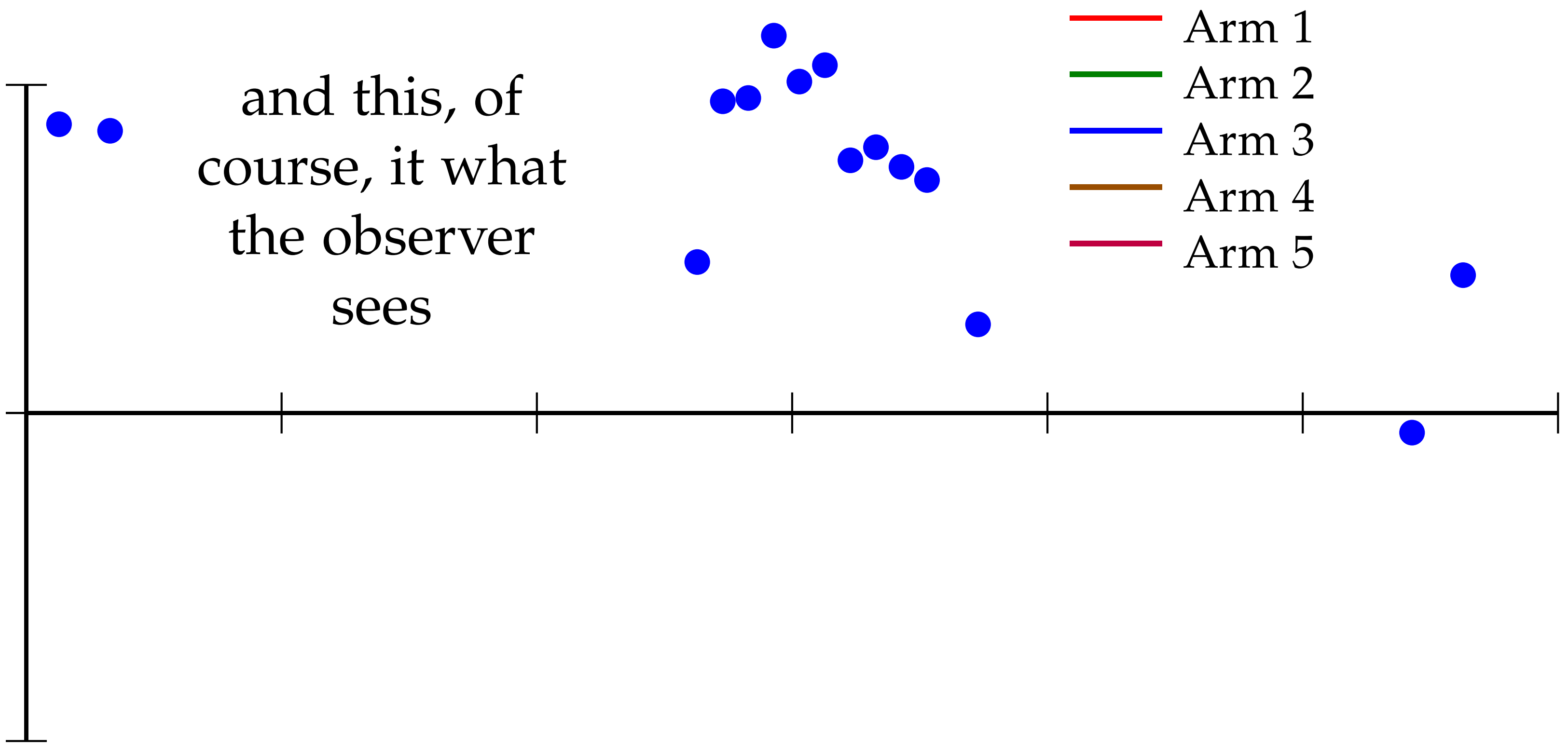
MAB with non-stationary payoffs



True payments per arm per round, had arm been pulled (solid lines).
Received payments, per arm per round (points).

Due to prediction errors, sometimes a “wrong” arm is pulled (verify!).

MAB with non-stationary payoffs



True payments per arm per round, had arm been pulled (solid lines).
Received payments, per arm per round (points).

Due to prediction errors, sometimes a “wrong” arm is pulled (verify!).

Status of MAB algorithms in MAL

Status of MAB algorithms in MAL

- Lots of theories, methods and techniques are known to analyse and predict time series.

Status of MAB algorithms in MAL

- Lots of theories, methods and techniques are known to analyse and predict time series.

  Regular, a.k.a. evenly spaced time series are best understood. This is probably due to its inherent properties and efforts in finance and stock prediction.

Status of MAB algorithms in MAL

- Lots of theories, methods and techniques are known to analyse and predict time series.

  Regular, a.k.a. evenly spaced time series are best understood. This is probably due to its inherent properties and efforts in finance and stock prediction.

- Unevenly spaced time series are less well understood.

Status of MAB algorithms in MAL

- Lots of theories, methods and techniques are known to analyse and predict time series.

  Regular, a.k.a. evenly spaced time series are best understood. This is probably due to its inherent properties and efforts in finance and stock prediction.

- Unevenly spaced time series are less well understood.

  Some techniques:

Status of MAB algorithms in MAL

- Lots of theories, methods and techniques are known to analyse and predict time series.

  Regular, a.k.a. evenly spaced time series are best understood. This is probably due to its inherent properties and efforts in finance and stock prediction.

- Unevenly spaced time series are less well understood.

  Some techniques:

  - Interpolate empty intervals / transform to evenly-spaced series. ("Traces" is a Python library based on this principle.)

Status of MAB algorithms in MAL

- Lots of theories, methods and techniques are known to analyse and predict time series.

  Regular, a.k.a. evenly spaced time series are best understood. This is probably due to its inherent properties and efforts in finance and stock prediction.

- Unevenly spaced time series are less well understood.

  Some techniques:

  - Interpolate empty intervals / transform to evenly-spaced series. ("Traces" is a Python library based on this principle.)
  - Techniques that take irregular time series "as they are" include state space analysis, Kalman filtering, autoregression, and stochastic differential equations, to name a few.

Status of MAB algorithms in MAL

- Lots of theories, methods and techniques are known to analyse and predict time series.

  Regular, a.k.a. evenly spaced time series are best understood. This is probably due to its inherent properties and efforts in finance and stock prediction.

- Unevenly spaced time series are less well understood.

  Some techniques:

  - Interpolate empty intervals / transform to evenly-spaced series. ("Traces" is a Python library based on this principle.)
  - Techniques that take irregular time series "as they are" include state space analysis, Kalman filtering, autoregression, and stochastic differential equations, to name a few.

- Rather than to overengineer MAB algorithms, for MAL it is perhaps better to take advantage of the game context (own payoff matrix, opponent moves, opponent's hypothesized strategy).

Appendix:
Confidence intervals

Confidence interval when the variance is known

Confidence interval when the variance is known

Let α be small, for example 0.05.

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1 - \alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1 - \alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1 - \alpha/2$ percentile of the standard normal distribution.

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution. Normally, we only use $z_{0.05/2} \approx 1.96$ and $z_{0.01/2} \approx 2.576$.

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution. Normally, we only use $z_{0.05/2} \approx 1.96$ and $z_{0.01/2} \approx 2.576$. E.g., in R `qnorm(1-0.01/2, 0, 1)` yields `2.575829`.

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1 - \alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1 - \alpha/2$ percentile of the standard normal distribution. Normally, we only use $z_{0.05/2} \approx 1.96$ and $z_{0.01/2} \approx 2.576$. E.g., in R `qnorm(1-0.01/2, 0, 1)` yields `2.575829`.

Example.

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution. Normally, we only use $z_{0.05/2} \approx 1.96$ and $z_{0.01/2} \approx 2.576$. E.g., in R `qnorm(1-0.01/2, 0, 1)` yields `2.575829`.

Example. Suppose the payoff of Arm 7 is normally distributed with deviation $\sigma = 1.5$.

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution. Normally, we only use $z_{0.05/2} \approx 1.96$ and $z_{0.01/2} \approx 2.576$. E.g., in R `qnorm(1-0.01/2, 0, 1)` yields `2.575829`.

Example. Suppose the payoff of Arm 7 is normally distributed with deviation $\sigma = 1.5$. Suppose five trials yield $1, 2, 3, 4, 5$, so $\bar{X} = 3$.

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1 - \alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1 - \alpha/2$ percentile of the standard normal distribution. Normally, we only use $z_{0.05/2} \approx 1.96$ and $z_{0.01/2} \approx 2.576$. E.g., in R `qnorm(1-0.01/2, 0, 1)` yields `2.575829`.

Example. Suppose the payoff of Arm 7 is normally distributed with deviation $\sigma = 1.5$. Suppose five trials yield $1, 2, 3, 4, 5$, so $\bar{X} = 3$. We want a confidence interval of 95%

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution. Normally, we only use $z_{0.05/2} \approx 1.96$ and $z_{0.01/2} \approx 2.576$. E.g., in R `qnorm(1-0.01/2, 0, 1)` yields `2.575829`.

Example. Suppose the payoff of Arm 7 is normally distributed with deviation $\sigma = 1.5$. Suppose five trials yield $1, 2, 3, 4, 5$, so $\bar{X} = 3$. We want a confidence interval of 95% so $\alpha = 0.05$ and $z_{\alpha/2} \approx 1.96$.

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1 - \alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1 - \alpha/2$ percentile of the standard normal distribution. Normally, we only use $z_{0.05/2} \approx 1.96$ and $z_{0.01/2} \approx 2.576$. E.g., in R `qnorm(1-0.01/2, 0, 1)` yields `2.575829`.

Example. Suppose the payoff of Arm 7 is normally distributed with deviation $\sigma = 1.5$. Suppose five trials yield $1, 2, 3, 4, 5$, so $\bar{X} = 3$. We want a confidence interval of 95% so $\alpha = 0.05$ and $z_{\alpha/2} \approx 1.96$. So a 95% confidence interval for Arm 7's revenue is
$[3 - 1.96 \cdot 1.5/\sqrt{5}, 3 - 1.96 \cdot 1.5/\sqrt{5}]$

Confidence interval when the variance is known

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed with standard deviation σ, and $\bar{X}$ is the sample mean, then $1 - \alpha$ of the time the interval

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

contains the population mean μ, where $z_{\alpha/2}$ is the $1 - \alpha/2$ percentile of the standard normal distribution. Normally, we only use $z_{0.05/2} \approx 1.96$ and $z_{0.01/2} \approx 2.576$. E.g., in R `qnorm(1-0.01/2, 0, 1)` yields `2.575829`.

Example. Suppose the payoff of Arm 7 is normally distributed with deviation $\sigma = 1.5$. Suppose five trials yield $1, 2, 3, 4, 5$, so $\bar{X} = 3$. We want a confidence interval of 95% so $\alpha = 0.05$ and $z_{\alpha/2} \approx 1.96$. So a 95% confidence interval for Arm 7's revenue is
$[3 - 1.96 \cdot 1.5/\sqrt{5}, 3 - 1.96 \cdot 1.5/\sqrt{5}] \approx [1.69, 4.31]$. $\square$

Confidence interval when the variance is unknown

Confidence interval when the variance is unknown

Let α be small, for example 0.05.

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1 - \alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2, n-1}$ is the $1 - \alpha/2$ percentile of the student t-distribution with $n - 1$ degrees of freedom.

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ percentile of the student t-distribution with $n-1$ degrees of freedom.

Example.

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1 - \alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2,n-1}$ is the $1 - \alpha/2$ percentile of the student t-distribution with $n - 1$ degrees of freedom.

Example. Suppose five times pulling Arm 7 yields $1, 2, 3, 4, 5$

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ percentile of the student t-distribution with $n-1$ degrees of freedom.

Example. Suppose five times pulling Arm 7 yields $1, 2, 3, 4, 5$, so $\bar{X} = 3$

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ percentile of the student t-distribution with $n-1$ degrees of freedom.

Example. Suppose five times pulling Arm 7 yields $1, 2, 3, 4, 5$, so $\bar{X} = 3$ and $S = \sqrt{\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5-1}}$

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ percentile of the student t-distribution with $n-1$ degrees of freedom.

Example. Suppose five times pulling Arm 7 yields $1, 2, 3, 4, 5$, so $\bar{X} = 3$ and $S = \sqrt{\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5-1}} \approx 1.58$.

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ percentile of the student t-distribution with $n-1$ degrees of freedom.

Example. Suppose five times pulling Arm 7 yields $1, 2, 3, 4, 5$, so $\bar{X} = 3$ and $S = \sqrt{\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5-1}} \approx 1.58$. We want a confidence interval of 95%

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ percentile of the student t-distribution with $n-1$ degrees of freedom.

Example. Suppose five times pulling Arm 7 yields $1,2,3,4,5$, so $\bar{X}=3$ and $S=\sqrt{\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5-1}} \approx 1.58$. We want a confidence interval of 95% so $\alpha=0.05$ and $t_{0.05/2,5-1}\approx 2.78$.

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1 - \alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \right]$$

contains the population mean μ, where $t_{\alpha/2, n-1}$ is the $1 - \alpha/2$ percentile of the student t-distribution with $n - 1$ degrees of freedom.

Example. Suppose five times pulling Arm 7 yields $1, 2, 3, 4, 5$, so $\bar{X} = 3$ and $S = \sqrt{\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5-1}} \approx 1.58$. We want a confidence interval of 95% so $\alpha = 0.05$ and $t_{0.05/2, 5-1} \approx 2.78$. E.g., in R `qt(1-0.05/2, df=4)` yields `2.776445`.

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ percentile of the student t-distribution with $n-1$ degrees of freedom.

Example. Suppose five times pulling Arm 7 yields $1, 2, 3, 4, 5$, so $\bar{X} = 3$ and $S = \sqrt{\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5-1}} \approx 1.58$. We want a confidence interval of 95% so $\alpha = 0.05$ and $t_{0.05/2,5-1} \approx 2.78$. E.g., in R `qt(1-0.05/2, df=4)` yields `2.776445`. So a 95% confidence interval for Arm 7's revenue without knowing its variance is
$[3 - 2.78 \cdot 1.58/\sqrt{5}, 3 - 2.78 \cdot 1.58/\sqrt{5}]$

Confidence interval when the variance is unknown

Let α be small, for example 0.05. If we sample n individuals from a population that is normally distributed, $\bar{X}$ is the sample mean, and S is the sample deviation, then $1-\alpha$ of the time the interval

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right]$$

contains the population mean μ, where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ percentile of the student t-distribution with $n-1$ degrees of freedom.

Example. Suppose five times pulling Arm 7 yields $1, 2, 3, 4, 5$, so $\bar{X} = 3$ and $S = \sqrt{\frac{(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2}{5-1}} \approx 1.58$. We want a confidence interval of 95% so $\alpha = 0.05$ and $t_{0.05/2,5-1} \approx 2.78$. E.g., in R `qt(1-0.05/2, df=4)` yields `2.776445`. So a 95% confidence interval for Arm 7's revenue without knowing its variance is
$[3-2.78\cdot 1.58/\sqrt{5}, 3-2.78\cdot 1.58/\sqrt{5}] \approx [1.04, 4.96]$. $\square$

Appendix:
The two Borel-Cantelli lemma's

The two Borel-Cantelli lemma's

The two Borel-Cantelli lemma's

BC1.

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2.

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin.

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time.

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success".

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success". So

$$P\{E_1\} + P\{E_2\} + P\{E_3\} + \dots$$

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success". So

$$P\{E_1\} + P\{E_2\} + P\{E_3\} + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots$$

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success". So

$$P\{E_1\} + P\{E_2\} + P\{E_3\} + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$$

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success". So

$$P\{E_1\} + P\{E_2\} + P\{E_3\} + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$$

which is finite.

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success". So

$$P\{E_1\} + P\{E_2\} + P\{E_3\} + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$$

which is finite. From **BC1** it follows that, if these experiments are executed

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success". So

$$P\{E_1\} + P\{E_2\} + P\{E_3\} + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$$

which is finite. From **BC1** it follows that, if these experiments are executed (in any order)

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success". So

$$P\{E_1\} + P\{E_2\} + P\{E_3\} + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$$

which is finite. From **BC1** it follows that, if these experiments are executed (in any order), only finitely many of them succeed

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success". So

$$P\{E_1\} + P\{E_2\} + P\{E_3\} + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$$

which is finite. From **BC1** it follows that, if these experiments are executed (in any order), only finitely many of them succeed a.s.

The two Borel-Cantelli lemma's

BC1. If the sum of the probabilities of a countably infinite number of events is finite, then only finitely many of these events occur a.s.

BC2. If this same sum is infinite, and events are independent, then infinitely many of these events occur a.s.

BC1 is a.k.a. the convergence BC lemma, or the trivial BC lemma; **BC2** is a.k.a. the divergence BC lemma, or the non-trivial BC lemma.

Example for BC1: experiment n consists of n times flipping a coin. It succeeds if heads show up all the time. E_n = "success". So

$$P\{E_1\} + P\{E_2\} + P\{E_3\} + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$$

which is finite. From **BC1** it follows that, if these experiments are executed (in any order), only finitely many of them succeed a.s. $\square$

The online dating site

Example for BC2:

The online dating site

Example for BC2: we play a game.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{"finger i is raised in round } n\text{"},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent,

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{"finger i is raised in round } n\text{"},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n \geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{"finger i is raised in round } n\text{"},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{"finger i is raised in round } n\text{"},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{"finger i is raised in round } n\text{"},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n \geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back and throw in a red ball as well.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back and throw in a red ball as well.

From **BC2** it follows that, notwithstanding the great many red balls in later rounds, the blue ball will be picked infinitely many times

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back and throw in a red ball as well.

From **BC2** it follows that, notwithstanding the great many red balls in later rounds, the blue ball will be picked infinitely many times a.s.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{"finger i is raised in round } n\text{"},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n \geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back and throw in a red ball as well.

From **BC2** it follows that, notwithstanding the great many red balls in later rounds, the blue ball will be picked infinitely many times a.s. Indeed, let E_n = "ball drawn in round n is blue", $n \geq 1$.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back and throw in a red ball as well.

From **BC2** it follows that, notwithstanding the great many red balls in later rounds, the blue ball will be picked infinitely many times a.s. Indeed, let E_n = “ball drawn in round n is blue”, $n \geq 1$. So $P\{E_n\} = 1/n$.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back and throw in a red ball as well.

From **BC2** it follows that, notwithstanding the great many red balls in later rounds, the blue ball will be picked infinitely many times a.s. Indeed, let E_n = “ball drawn in round n is blue”, $n \geq 1$. So $P\{E_n\} = 1/n$. The sum of these probabilities is

$$1 + 1/2 + 1/3 + 1/4 + \ldots$$

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{“finger i is raised in round } n\text{”},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back and throw in a red ball as well.

From **BC2** it follows that, notwithstanding the great many red balls in later rounds, the blue ball will be picked infinitely many times a.s. Indeed, let E_n = “ball drawn in round n is blue”, $n \geq 1$. So $P\{E_n\} = 1/n$. The sum of these probabilities is

$$1 + 1/2 + 1/3 + 1/4 + \dots$$

which is the so-called harmonic sum which is know to diverge.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{"finger i is raised in round } n\text{"},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back and throw in a red ball as well.

From **BC2** it follows that, notwithstanding the great many red balls in later rounds, the blue ball will be picked infinitely many times a.s. Indeed, let E_n = "ball drawn in round n is blue", $n \geq 1$. So $P\{E_n\} = 1/n$. The sum of these probabilities is

$$1 + 1/2 + 1/3 + 1/4 + \ldots$$

which is the so-called harmonic sum which is know to diverge. Hence **BC2** applies.

The online dating site

Example for BC2: we play a game. Each time an arbitrary finger of one hand is raised with probability $1/5$, independent of the previous fingers raised. So if

$$E_{i,n} = \text{"finger i is raised in round } n\text{"},$$

then $P\{E_{i,n}\} = 1/5$. Fix a finger i. The $\{E_{i,n} \mid n \geq 1\}$ are independent, and $\sum_{n\geq 1} P\{E_{i,n}\}$ is unbounded. From **B2**: $E_{i,n}$ occurs infinitely often a.s.

Example for BC2: Start with an urn with one blue bal. Repeat: draw a bal, observe its color, put it back and throw in a red ball as well.

From **BC2** it follows that, notwithstanding the great many red balls in later rounds, the blue ball will be picked infinitely many times a.s. Indeed, let E_n = "ball drawn in round n is blue", $n \geq 1$. So $P\{E_n\} = 1/n$. The sum of these probabilities is

$$1 + 1/2 + 1/3 + 1/4 + \dots$$

which is the so-called harmonic sum which is know to diverge. Hence **BC2** applies. $\square$

The online dating site

Example for BC2: suppose a dating site has 50 male members,

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

$E_n =$ "André will be selected in week n".

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

If Marian subscribed in Week k,

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”.}$$

If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\}$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

$$E_n = \text{"André will be selected in week } n\text{"}.$$

If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{\underbrace{50 + \cdots + 50}_{n}}$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{\underbrace{50 + \cdots + 50}_{n}} = \sum_{n=k}^{\infty} \frac{1}{50n}$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{\underbrace{50 + \cdots + 50}_{n}} = \sum_{n=k}^{\infty} \frac{1}{50n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{n}$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

$$E_n = \text{"André will be selected in week } n\text{"}.$$

If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{\underbrace{50 + \cdots + 50}_{n}} = \sum_{n=k}^{\infty} \frac{1}{50n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{n}$$

Since the harmonic series $\sum_n 1/n$, hence every tail of it, diverges, **BC2** applies:

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”.}$$

If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{\underbrace{50 + \cdots + 50}_{n}} = \sum_{n=k}^{\infty} \frac{1}{50n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{n}$$

Since the harmonic series $\sum_n 1/n$, hence every tail of it, diverges, **BC2** applies: at any point in time, and no matter how many males, André will be selected eventually.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. This dating site welcomes 50 new male subscribers every week. After some time, which time we don't know, Marian subscribes. From then on, every week Marian selects a male candidate randomly. She and André are destined for each other and for no one else. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{\underbrace{50 + \cdots + 50}_{n}} = \sum_{n=k}^{\infty} \frac{1}{50n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{n}$$

Since the harmonic series $\sum_n 1/n$, hence every tail of it, diverges, **BC2** applies: at any point in time, and no matter how many males, André will be selected eventually. $\square$

The online dating site

Example for BC2: suppose a dating site has 50 male members

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”.}$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\}$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{50(1.0002)^n}$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{50(1.0002)^n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{1.0002^n}$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\begin{aligned}
\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{50(1.0002)^n} &= \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{1.0002^n} \\
&\approx \frac{1}{50} \sum_{n=k}^{\infty} (0.9998)^n
\end{aligned}$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{50(1.0002)^n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{1.0002^n}$$

$$\approx \frac{1}{50} \sum_{n=k}^{\infty} (0.9998)^n \leq \frac{1}{50} \sum_{n=1}^{\infty} (0.9998)^n$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{50(1.0002)^n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{1.0002^n}$$
$$\approx \frac{1}{50} \sum_{n=k}^{\infty} (0.9998)^n \leq \frac{1}{50} \sum_{n=1}^{\infty} (0.9998)^n = \frac{1}{50} \frac{0.9998}{1 - 0.9998},$$

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\begin{aligned}
\sum_{n=k}^{\infty} P\{E_n\} &= \sum_{n=k}^{\infty} \frac{1}{50(1.0002)^n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{1.0002^n} \\
&\approx \frac{1}{50} \sum_{n=k}^{\infty} (0.9998)^n \leq \frac{1}{50} \sum_{n=1}^{\infty} (0.9998)^n = \frac{1}{50} \frac{0.9998}{1 - 0.9998},
\end{aligned}$$

because $\sum_{n=1}^{\infty} x^n = x/(1-x)$ for all $0 \leq x < 1$.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\sum_{n=k}^{\infty} P\{E_n\} = \sum_{n=k}^{\infty} \frac{1}{50(1.0002)^n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{1.0002^n}$$
$$\approx \frac{1}{50} \sum_{n=k}^{\infty} (0.9998)^n \leq \frac{1}{50} \sum_{n=1}^{\infty} (0.9998)^n = \frac{1}{50} \frac{0.9998}{1 - 0.9998},$$

because $\sum_{n=1}^{\infty} x^n = x/(1-x)$ for all $0 \leq x < 1$. **BC1** applies: André will be selected only finitely many times

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{“André will be selected in week } n\text{”}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\begin{aligned}\sum_{n=k}^{\infty} P\{E_n\} &= \sum_{n=k}^{\infty} \frac{1}{50(1.0002)^n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{1.0002^n} \\ &\approx \frac{1}{50} \sum_{n=k}^{\infty} (0.9998)^n \leq \frac{1}{50} \sum_{n=1}^{\infty} (0.9998)^n = \frac{1}{50} \frac{0.9998}{1-0.9998},\end{aligned}$$

because $\sum_{n=1}^{\infty} x^n = x/(1-x)$ for all $0 \leq x < 1$. **BC1** applies: André will be selected only finitely many times—likely never if k is large.

The online dating site

Example for BC2: suppose a dating site has 50 male members, of which André is one of them. On the male part, this dating site grows 1% per year. After some time, Marian selects a male candidate randomly every week. Will André be selected by Marian eventually?

$$E_n = \text{"André will be selected in week } n\text{"}.$$

Factor 1.01 a year is $\sqrt[52]{1.01} \approx 1.0002$ per week. If Marian subscribed in Week k,

$$\begin{aligned}\sum_{n=k}^{\infty} P\{E_n\} &= \sum_{n=k}^{\infty} \frac{1}{50(1.0002)^n} = \frac{1}{50} \sum_{n=k}^{\infty} \frac{1}{1.0002^n} \\ &\approx \frac{1}{50} \sum_{n=k}^{\infty} (0.9998)^n \leq \frac{1}{50} \sum_{n=1}^{\infty} (0.9998)^n = \frac{1}{50} \frac{0.9998}{1 - 0.9998},\end{aligned}$$

because $\sum_{n=1}^{\infty} x^n = x/(1-x)$ for all $0 \leq x < 1$. **BC1** applies: André will be selected only finitely many times—likely never if k is large. $\square$

Pairwise vs. mutual independence

Pairwise vs. mutual independence

Definition.

Pairwise vs. mutual independence

Definition.

- Events $E_1, \dots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.

Pairwise vs. mutual independence

Definition.

- Events $E_1, \ldots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \ldots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Pairwise vs. mutual independence

Definition.

- Events $E_1, \dots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \dots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Example.

Pairwise vs. mutual independence

Definition.

- Events $E_1, \dots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \dots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Example. Consider $A = \{$the first die rolled is a $3\}$, $B = \{$the second die rolled is a $4\}$, $C = \{$their sum is $7\}$.

Pairwise vs. mutual independence

Definition.

- Events $E_1, \ldots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \ldots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Example. Consider $A = \{$the first die rolled is a $3\}$, $B = \{$the second die rolled is a $4\}$, $C = \{$their sum is $7\}$.

You may verify that A, B and C are pairwise independent.

Pairwise vs. mutual independence

Definition.

- Events $E_1, \dots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \dots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Example. Consider $A = \{$the first die rolled is a $3\}$, $B = \{$the second die rolled is a $4\}$, $C = \{$their sum is $7\}$.

You may verify that A, B and C are pairwise independent. Indeed, knowing that one of the dices shows 3 or 4 does not affect the odds of having 7 points

Pairwise vs. mutual independence

Definition.

- Events $E_1, \ldots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \ldots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Example. Consider $A = \{$the first die rolled is a $3\}$, $B = \{$the second die rolled is a $4\}$, $C = \{$their sum is $7\}$.

You may verify that A, B and C are pairwise independent. Indeed, knowing that one of the dices shows 3 or 4 does not affect the odds of having 7 points, and knowing that the eyes add up to 7 does not give extra information about the odds on having a 3 or a 4.

Pairwise vs. mutual independence

Definition.

- Events $E_1, \dots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \dots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Example. Consider $A = \{\text{the first die rolled is a } 3\}$, $B = \{\text{the second die rolled is a } 4\}$, $C = \{\text{their sum is } 7\}$.

You may verify that A, B and C are pairwise independent. Indeed, knowing that one of the dices shows 3 or 4 does not affect the odds of having 7 points, and knowing that the eyes add up to 7 does not give extra information about the odds on having a 3 or a 4. However, A, B and C do depend on one another.

Pairwise vs. mutual independence

Definition.

- Events $E_1, \dots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \dots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Example. Consider $A = \{$the first die rolled is a $3\}$, $B = \{$the second die rolled is a $4\}$, $C = \{$their sum is $7\}$.

You may verify that A, B and C are pairwise independent. Indeed, knowing that one of the dices shows 3 or 4 does not affect the odds of having 7 points, and knowing that the eyes add up to 7 does not give extra information about the odds on having a 3 or a 4. However, A, B and C do depend on one another. E.g., $P\{C \mid AB\} = 1 \neq P\{C\} < 1$.

Pairwise vs. mutual independence

Definition.

- Events $E_1, \dots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \dots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Example. Consider $A = \{$the first die rolled is a $3\}$, $B = \{$the second die rolled is a $4\}$, $C = \{$their sum is $7\}$.

You may verify that A, B and C are pairwise independent. Indeed, knowing that one of the dices shows 3 or 4 does not affect the odds of having 7 points, and knowing that the eyes add up to 7 does not give extra information about the odds on having a 3 or a 4. However, A, B and C do depend on one another. E.g., $P\{C \mid AB\} = 1 \neq P\{C\} < 1$. $\square$

Pairwise vs. mutual independence

Definition.

- Events $E_1, \dots, E_n$ are pairwise independent if $P\{E_i \mid E_j\} = P\{E_i\}$ for all i and j such that $j \neq i$.
- Events $E = \{E_1, \dots, E_n\}$ are (just, or, mutual) independent if $P\{E_i \mid D\} = P\{E_i\}$ for all i and non-empty $D \subseteq E \backslash \{E_i\}$.

Example. Consider $A = \{$the first die rolled is a $3\}$, $B = \{$the second die rolled is a $4\}$, $C = \{$their sum is $7\}$.

You may verify that A, B and C are pairwise independent. Indeed, knowing that one of the dices shows 3 or 4 does not affect the odds of having 7 points, and knowing that the eyes add up to 7 does not give extra information about the odds on having a 3 or a 4. However, A, B and C do depend on one another. E.g., $P\{C \mid AB\} = 1 \neq P\{C\} < 1$. $\square$

BC2 holds even if the events involved are pairwise independent.

Appendix: Numerical trace of UCB on five arms

Sample run UCB arm variance 5, rounds 1-2

Sample run UCB arm variance 5, rounds 1-2

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	1	1	1	1
total empirical	2.70	12.73	2.20	6.91	3.80
empirical mean	2.70	12.73	2.20	6.91	3.80
upper confidence	2.70	12.73	2.20	6.91	3.80
value of pulled		9.40			

Sample run UCB arm variance 5, rounds 1-2

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	1	1	1	1
total empirical	2.70	12.73	2.20	6.91	3.80
empirical mean	2.70	12.73	2.20	6.91	3.80
upper confidence	2.70	12.73	2.20	6.91	3.80
value of pulled		9.40			

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	2	1	1	1
total empirical	2.70	22.13	2.20	6.91	3.80
empirical mean	2.70	11.07	2.20	6.91	3.80
upper confidence	3.87	11.90	3.38	8.09	4.98
value of pulled		6.20			

Sample run UCB arm variance 5, rounds 3-4

Sample run UCB arm variance 5, rounds 3-4

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	3	1	1	1
total empirical	2.70	28.33	2.20	6.91	3.80
empirical mean	2.70	9.44	2.20	6.91	3.80
upper confidence	4.18	10.30	3.68	8.39	5.29
value of pulled		5.40			

Sample run UCB arm variance 5, rounds 3-4

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	3	1	1	1
total empirical	2.70	28.33	2.20	6.91	3.80
empirical mean	2.70	9.44	2.20	6.91	3.80
upper confidence	4.18	10.30	3.68	8.39	5.29
value of pulled		5.40			

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	4	1	1	1
total empirical	2.70	33.73	2.20	6.91	3.80
empirical mean	2.70	8.43	2.20	6.91	3.80
upper confidence	4.36	9.27	3.87	8.58	5.47
value of pulled		6.98			

Sample run UCB arm variance 5, rounds 5-6

Sample run UCB arm variance 5, rounds 5-6

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	5	1	1	1
total empirical	2.70	40.71	2.20	6.91	3.80
empirical mean	2.70	8.14	2.20	6.91	3.80
upper confidence	4.49	8.94	4.00	8.71	5.60
value of pulled		11.68			

Sample run UCB arm variance 5, rounds 5-6

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	5	1	1	1
total empirical	2.70	40.71	2.20	6.91	3.80
empirical mean	2.70	8.14	2.20	6.91	3.80
upper confidence	4.49	8.94	4.00	8.71	5.60
value of pulled		11.68			

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	6	1	1	1
total empirical	2.70	52.39	2.20	6.91	3.80
empirical mean	2.70	8.73	2.20	6.91	3.80
upper confidence	4.59	9.50	4.09	8.81	5.70
value of pulled		2.59			

Sample run UCB arm variance 5, rounds 7-8

Sample run UCB arm variance 5, rounds 7-8

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	7	1	1	1
total empirical	2.70	54.98	2.20	6.91	3.80
empirical mean	2.70	7.85	2.20	6.91	3.80
upper confidence	4.67	8.60	4.17	8.88	5.78
value of pulled				4.84	

Sample run UCB arm variance 5, rounds 7-8

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	7	1	1	1
total empirical	2.70	54.98	2.20	6.91	3.80
empirical mean	2.70	7.85	2.20	6.91	3.80
upper confidence	4.67	8.60	4.17	8.88	5.78
value of pulled				4.84	

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	7	1	2	1
total empirical	2.70	54.98	2.20	11.75	3.80
empirical mean	2.70	7.85	2.20	5.88	3.80
upper confidence	4.74	8.63	4.24	7.32	5.84
value of pulled		4.76			

Sample run UCB arm variance 5, rounds 9-10

Sample run UCB arm variance 5, rounds 9-10

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	8	1	2	1
total empirical	2.70	59.75	2.20	11.75	3.80
empirical mean	2.70	7.47	2.20	5.88	3.80
upper confidence	4.79	8.21	4.30	7.36	5.90
value of pulled		2.22			

Sample run UCB arm variance 5, rounds 9-10

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	8	1	2	1
total empirical	2.70	59.75	2.20	11.75	3.80
empirical mean	2.70	7.47	2.20	5.88	3.80
upper confidence	4.79	8.21	4.30	7.36	5.90
value of pulled		2.22			

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	9	1	2	1
total empirical	2.70	61.96	2.20	11.75	3.80
empirical mean	2.70	6.88	2.20	5.88	3.80
upper confidence	4.84	7.60	4.35	7.39	5.95
value of pulled		14.34			

Sample run UCB arm variance 5, rounds 11-12

Sample run UCB arm variance 5, rounds 11-12

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	10	1	2	1
total empirical	2.70	76.30	2.20	11.75	3.80
empirical mean	2.70	7.63	2.20	5.88	3.80
upper confidence	4.89	8.32	4.39	7.43	5.99
value of pulled		10.93			

Sample run UCB arm variance 5, rounds 11-12

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	10	1	2	1
total empirical	2.70	76.30	2.20	11.75	3.80
empirical mean	2.70	7.63	2.20	5.88	3.80
upper confidence	4.89	8.32	4.39	7.43	5.99
value of pulled		10.93			

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	11	1	2	1
total empirical	2.70	87.23	2.20	11.75	3.80
empirical mean	2.70	7.93	2.20	5.88	3.80
upper confidence	4.93	8.60	4.43	7.45	6.03
value of pulled		14.11			

Sample run UCB arm variance 5, rounds 13-14

Sample run UCB arm variance 5, rounds 13-14

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	12	1	2	1
total empirical	2.70	101.33	2.20	11.75	3.80
empirical mean	2.70	8.44	2.20	5.88	3.80
upper confidence	4.96	9.10	4.47	7.48	6.07
value of pulled		1.72			

Sample run UCB arm variance 5, rounds 13-14

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	12	1	2	1
total empirical	2.70	101.33	2.20	11.75	3.80
empirical mean	2.70	8.44	2.20	5.88	3.80
upper confidence	4.96	9.10	4.47	7.48	6.07
value of pulled		1.72			

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	13	1	2	1
total empirical	2.70	103.06	2.20	11.75	3.80
empirical mean	2.70	7.93	2.20	5.88	3.80
upper confidence	4.99	8.56	4.50	7.50	6.10
value of pulled		5.56			

Sample run UCB arm variance 5, rounds 15-16

Sample run UCB arm variance 5, rounds 15-16

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	14	1	2	1
total empirical	2.70	108.62	2.20	11.75	3.80
empirical mean	2.70	7.76	2.20	5.88	3.80
upper confidence	5.02	8.38	4.53	7.52	6.13
value of pulled		0.87			

Sample run UCB arm variance 5, rounds 15-16

	A_0	A_1	A_2	A_3	A_4
’intrinsic’ value	5	6	7	8	9
times pulled	1	14	1	2	1
total empirical	2.70	108.62	2.20	11.75	3.80
empirical mean	2.70	7.76	2.20	5.88	3.80
upper confidence	5.02	8.38	4.53	7.52	6.13
value of pulled		0.87			

	A_0	A_1	A_2	A_3	A_4
’intrinsic’ value	5	6	7	8	9
times pulled	1	15	1	2	1
total empirical	2.70	109.49	2.20	11.75	3.80
empirical mean	2.70	7.30	2.20	5.88	3.80
upper confidence	5.05	7.91	4.56	7.54	6.16
value of pulled		-5.30			

Sample run UCB arm variance 5, rounds 17-18

Sample run UCB arm variance 5, rounds 17-18

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	16	1	2	1
total empirical	2.70	104.19	2.20	11.75	3.80
empirical mean	2.70	6.51	2.20	5.88	3.80
upper confidence	5.08	7.11	4.58	7.56	6.18
value of pulled				4.70	

Sample run UCB arm variance 5, rounds 17-18

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	16	1	2	1
total empirical	2.70	104.19	2.20	11.75	3.80
empirical mean	2.70	6.51	2.20	5.88	3.80
upper confidence	5.08	7.11	4.58	7.56	6.18
value of pulled				4.70	

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	16	1	3	1
total empirical	2.70	104.19	2.20	16.46	3.80
empirical mean	2.70	6.51	2.20	5.49	3.80
upper confidence	5.10	7.11	4.61	6.87	6.21
value of pulled		9.23			

Sample run UCB arm variance 5, rounds 19-20

Sample run UCB arm variance 5, rounds 19-20

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	17	1	3	1
total empirical	2.70	113.42	2.20	16.46	3.80
empirical mean	2.70	6.67	2.20	5.49	3.80
upper confidence	5.12	7.26	4.63	6.89	6.23
value of pulled		-5.48			

Sample run UCB arm variance 5, rounds 19-20

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	17	1	3	1
total empirical	2.70	113.42	2.20	16.46	3.80
empirical mean	2.70	6.67	2.20	5.49	3.80
upper confidence	5.12	7.26	4.63	6.89	6.23
value of pulled		-5.48			

	A_0	A_1	A_2	A_3	A_4
'intrinsic' value	5	6	7	8	9
times pulled	1	18	1	3	1
total empirical	2.70	107.94	2.20	16.46	3.80
empirical mean	2.70	6.00	2.20	5.49	3.80
upper confidence	5.14	6.57	4.65	6.90	6.25
value of pulled				9.63	