A statistical analysis of the accusations made by Jonas Egeberg about the game Carlo Metta vs Reem Ben David played on 29 November 2017

Francesco Morandin, defender expert

23 April 2018

We show that the measurement performed by the expert witness (98% of similarity) was wrong, and that the correct stable measurement is less than 94%. Since the expert witness did not provide a statistical assessment of how often an innocent would be wrongly convicted on the basis of his methodology, we perform a theoretical estimate: out of 100 innocent regular players, every year at least one would be wrongly convicted with 98%, and more than 30 would be wrongly convicted with 94%. Indeed, by looking at games from this year's PGETC, we can find several games of other players with very high percentages of similarity.

Contents

1 The numbers generated this way are unstable and support wrong conclusions

The expert witness described his method in this way. *As for the method used: We checked moves 50-150 and noted the moves as similar if they were within Leela's top 3 moves, and no further than 5% away from its top move. We usually let it run around 50k nodes, though this varied a bit depending on the move played.* The referee claimed that he based his decision on the fact that this methodology applied to the game Metta vs Ben David (the contested game) gave a similarity of 98%.

In this document, we used the same method:

1. We analyze moves 51-150, that is, 50 moves for Black and 50 moves for White.
2. For every move, we say it is *similar* if the move is within Leela's top 3 moves and no farther than 5% from Leela's top move.
3. All the moves excluded by the point above are called *missed.*
4. Hardware: Intel Core i7, 2.60 GHZ, RAM 16GB, GPU NVIDIA GeForce GTX 960M. Operating system: Windows 10.
5. Leela 0.11 with GPU support, downloaded from https://www.sjeng.org/leela.html.

For clarity, 4 moves of Leela missed by the player mean 92% of similarity, 10 missed mean 80%, and so on.

As for computation time, we made several choices: 30", 60", 75" per move, 50k, 100k, 150k nodes, exact time used by the players (*actual time*). For the analysis of live games with no time recorded, we used 30", which corresponds, with the given hardware, approximately with 100k nodes.

1.1 Extensive analysis provide different numbers, and the contested game had less than 94% similarity

If the methodology had been applied repeatedly, it would have resulted clear that it provides every time different numbers. Therefore, running one or two runs does not prove anything about which hints a cheater player would have received by Leela during a cheated game. The reason why different runs provide different numbers is that when Leela makes a choice, she plays many games to assess which choice is the best. The choice of games is partially **random**, and this is why the output changes from one run to another. More details on how Leela decides her moves can be found in the statistical appendix A.

Here is an extensive analysis of the contested game Metta vs Ben David.

We were at first unable to replicate the value found in the expert witness' runs, using various time settings.

1. [1] First run. Actual time, **90%** (5 missed: moves 97, 101, 117, 125, 139).
2. [2] Second run. Actual time, **86%** (7 missed: moves 97, 101, 105, 117, 121, 125, 129).
3. [3] Third run. 60" time, **92%** (4 missed: moves 97, 101, 121, 149).
4. [4] Fourth run. 75" time, **92%** (4 missed: moves 55, 97, 101, 117).
5. [5] Fifth run. 50k nodes, **92%** (4 missed: moves 55, 101, 105, 121).

In order to reproduce the result of the expert witness, we isolated 10 moves that in at least one of the previous analyses were not in the top 3: moves 55, 97, 101, 105, 117, 121, 125, 129, 139, 149. We made an extensive analysis of those moves only, running Leela 105 times on each of them, at 50k nodes.

- **90%** (5 missed): 4 times.
- **92%** (4 missed): 30 times.
- **94%** (3 missed): 57 times.
- **96%** (2 missed): 12 times.
- **98%** (1 missed): 2 times.
- **100%** (0 missed): 0 times.

Please note that in 105 runs, only twice were we able to replicate the expert witness' figure (1 missed move), that is, 1.9%. The average similarity from those 105 runs was **93.6%**, that is, **less than 94%**: this is the number that we consider more accurate for the contested game.

In support of the fact that an extensive analysis provides different numbers, we annotated the way the positions in Leela of the above 10 moves changed.

- Move 55: position 2 or 3, rarely 4.
- Move 97: position 3, 4, 5 or 6, often 4.
- Move 101: position 2, 3, 4, 5, or 6.
- Move 105: position 3. Please note that letting the analysis run for more than 1 minute, this move goes to position 4, and Metta used almost 3 minutes for this move.
- Move 117: position 1, 2 or 3, often 2.
- Move 121: position 4, 6, 8 or out of top 20, rarely 1. This is the most unstable move.
- Move 125: position 2, 3 or 4.
- Move 129: position 2 or 3.
- Moves 139 and 149: position 1 or 2, often 1.

Beyond extensively analysing the contested game itself, we also tried to reproduce the other figures mentioned by the expert witness, and even in those cases we obtained different numbers.

- [6] Metta vs Pankoke, EGC 2017. Result mentioned by the expert witness: **70%-80%**. Our run: **90%**.
- [7] Metta vs Grigoriu, EGC 2017. Result mentioned by the expert witness: **74%**. Our run: **84%**.
- [8], [9] Metta vs Kruml, PGETC 2017-2018. Result mentioned by the expert witness: **90%**. Our runs: **84%**. Please note that we made 2 runs for this games, obtaining 84% on both of them.

We obtained a similar number as the expert witness' only in one case: [10] Metta vs Shakov, EGC 2016. Result mentioned by the expert witness: **80%**. Our run: **82%**.

As a general remark, the analysis with 50k nodes appears very unstable: if Leela runs longer, it takes consistent decisions much more often. In [11] you can find a table with Leela's position of the 10 moves in the contested game with 50k, 100k and 150k runs each. If 100k or 150k are chosen, for instance, similarity between Leela and Metta in the contested game decreases.

1.2 In the contested game, Leela judges that Metta's moves were all human

We analysed another number that Leela provides: the probability that a strong amateur player would play this move. This number is called *network probability*. We put this in a table [12], where the choices of Metta are underlined and in red. We found that **Metta's choices were 49 out of 50 times in the human top 3** (and 46 out of 50 times in the human top 2). The only time his choice was "not human" (move 121) was a mistake according to Leela (Leela ranked this move in position 7 after indefinite time).

1.3 Metta had similarly high percentages in games where cheating was not possible

In the following, we give numbers on other games played by Metta, where cheating was not possible (live games or online games played before superhuman Go softwares were available).

- The following game was used by the referee to support his decision: [6] Metta vs Pankoke, EGC 2017. We found a similarity of **90%** (although the expert witness stated he had found 70%-80%).
- [13] Metta vs Kral, PGETC 2016-2017, February 2016, 4 months before creation of Leela, 3 months before Crazy Stones, 4 months before Zen 6. Similarity: **88%**.
- [14] Metta vs Budahn, EGC 2017. Similarity: **86%**.
- [15] Metta vs Yi, EGC 2017. Similarity: **90%**.

1.4 It's easy to find games with high percentages

The chosen methodology was very poor (see next sections). As a consequence, one can easily choose a game where the similarity index is high: by just choosing a calm game, with no big fights. We searched some games with those characteristics from PGETC league A of this year, and we found very high similarity indexes.

- [16] Bajenaru vs Habu, Romania vs Poland, PGETC 2017-2018 board 3. Bajenaru similarity: **96%**.
- [17] Bajenaru vs Fionin, Romania vs Russia, PGETC 2017-2018 board 3. Bajenaru similarity: **94%**.
- [18] Habu vs Csizmadia, Poland vs Hungary, PGETC 2017-2018 board 3. Habu similarity: **94%**.

2 The methodology was statistically flawed in the first place

In appendix to this document we include a general description of the risks of amateurish use of statistics and a thorough analysis of the methodology chosen by the expert witness, that shows why it is flawed.

We summarize here the main findings.

- The analysis of a single game for each of 3 players was completely insufficient to give a reasonable statistical estimate of the probability of wrongly convicting an innocent: at the very least, 10 players and 10 games each should have been analysed.
- In the absence of a statistical estimate, we provided a theoretical estimate, which shows that the methodology has a probability of wrongly convicting an innocent player that is unacceptably high: using 98% as a threshold, among 100 innocent players who play regularly online, at least one would be wrongly convicted every year; if the similarity that Metta truly had in the contested game (94% or less) was considered enough to rule a player guilty, **every year out of 100 innocent regular players, more than 30 would be wrongly convicted**.
- There was no attempt at learning from state-of-the-art cheat analysis for online chess games, where moves are given different weights based on the amount of choices available in the situation.
- There was no attempt to minimize the risk of wrongly convicting an innocent player, based on what is known of neural networks and of Leela in particular. For instance:
  - Leela cannot be automatically frozen at exactly 50k nodes, since the only settings are unlimited, 20k or smaller numbers: why were 50k nodes chosen? Why not choosing 20k, or unlimited, or the time the move took in the real game?
  - AlphaGo Zero and LeelaZero start the middle game at move 31, not at move 51: what was the basis of choosing 51?
  - When Leela is stopped in her computations, she provides many pieces of information about the moves she was assessing: the ordered list of moves at that time, the estimated probability of victory for each of them, the probability that a strong amateur player would make that move before Leela was created; why choosing the top 3, and not the first, or the top 5? Why not choosing less-probably-human moves?
- The methodology was chosen by people who were not blind to the moves that Metta had made during the contested game: since there wasn't any attempt to minimise the probability of accusing an innocent, this carries the risk of involuntarily picking a methodology exactly because it confirmed the accusations against Metta (see "cherry picking" in Section A.2).

References

[1] Metta vs Ben David, PGETC 2017-2018, actual time, 90%.

[2] Metta vs Ben David, PGETC 2017-2018, actual time, 86%.

[3] Metta vs Ben David, PGETC 2017-2018, 60", 92%.

[4] Metta vs Ben David, PGETC 2017-2018, 75", 92%.

[5] Metta vs Ben David, PGETC 2017-2018, 50k nodes, 92%.

[6] Metta vs Pankoke, EGC 2017, 30", 90%.

[7] Metta vs Grigoriu, EGC 2017, 30", 84%.

[8] Metta vs Kruml, PGETC 2017-2018, 50k nodes, 84%.

[9] Metta vs Kruml, PGETC 2017-2018, actual time, 84%.

[10] Metta vs Shakov, EGC 2016, 50k nodes, 82%.

[11] Attachment *worsting.ods*, position in Leela of 10 critical moves, with runs at 50k, 100k and 150k nodes.

[12] Attachment *humanity.ods*, network probability of humans in the contested game. Fixed, doesn't change with different runs.

[13] Metta vs Kral, PGETC 2016-2017 before superhuman Go software, 30", 88%.

[14] Metta vs Budahn, EGC 2017, 30", 86%.

[15] Metta vs Yi, EGC 2017, 30", 90%.

[16] Bajenaru vs Habu, PGETC 2017-2018, actual time, 96%.

[17] Bajenaru vs Fionin, PGETC 2017-2018, actual time, 94%.

[18] Habu vs Csizmadia, PGETC 2017-2018, actual time, 94%.

[19] *The Parable of the Golfers*, https://www.cse.buffalo.edu//~regan/chess/fidelity/Golfers.html

[20] *Review of the contested game by Stanisław Frejlak*, https://www.facebook.com/groups/go.igo.weiqi.baduk/permalink/10156305848221514/?comment_id=10156310120581514&comment_tracking=%7B%22tn%22%3A%22R%22%7D

A Appendix: statistical considerations

The following discussion is exposed in informal, non-technical language for the sake of clarity. The author of the discussion is an expert in the field of statistics and can provide extensive scientific bibliography supporting it, if required.

A.1 Preamble on statistical tests

Taking decisions in the presence of uncertainty is the definition of a **statistical test**. In its basic form, there are two opposite hypothesis (e.g. a coin is fair or it is not) and after observing some experiments (e.g. 100 coin tosses) a decision is made, based on them (e.g. 80% of the tosses yielded heads, so the coin is ruled as not fair).

One must understand that the test itself is a probabilistic, uncertain tool, and in particular that it is impossible to design it as to avoid all decision errors. This is why it is very important to know the performance of the test in terms of the two kind of errors that may occur: false positive and false negative.

The correct, ethical recommendation is to measure and understand the performance of the test *before* applying it to make any actual decision.

In order to do so, the very first steps are to identify:

1. how we will perform the experiments (e.g. 100 coin tosses),
2. what we will measure in the experiments (e.g. the number of heads),
3. what will be the threshold to discern between the two hypothesis.

Notice that even the classic example of the coin becomes quite complicated when we get to the last point. For example we may rule that the coin is unfair if the number of heads is too high, or too low, or if we see too many consecutive tosses with the same result, or even all of the above together.

Even if we have a basic one-sided hypothesis (e.g. we don't want the coin to yield too many heads) it may be difficult to decide the exact threshold: is 60% too much? or is 70%? 80%?

Nevertheless it is possible to compute the performance of the test beforehand in each of these cases: if the threshold of rejection is 60, 70 or 80, the probability of a false positive is $\alpha = 0.0284$, $\alpha = 4 \cdot 10^{-5}$, $\alpha = 6 \cdot 10^{-10}$. The latter is thus the probability that a fair coin fails the test and is incorrectly marked as unfair.

The probability of a false negative is more of a problem. It is the probability that an unfair coin is marked as fair, but this can never be a precise number, because it really depends on *how* the coin is unfair. In our example, if the coin favors heads with a probability of 75% (say), then it is $\beta = 0.00324$ (good), $\beta = 0.1038$ (ok), $\beta = 0.851$ (terrible); but if the coin favors heads with a probability of 55%, then it is $\beta = 0.817$, $\beta = 0.998$, $\beta = 0.9999999$ (all terrible).

There are two lessons here: there is always a trade-off between false positive and false negative (if you change the threshold one goes down and the other goes up), and it is difficult if not impossible to devise a test to distinguish fair coins from coins who are only *slightly* unfair.

When designing a statistical test, it is hence also essential to

4. compute the expected cost of the two kind of errors

in such a way that the threshold of (3) may be tuned in order to avoid potential pitfalls.

Leaving the example of the coin and moving on to the case of the decision in discussion, a quick analysis of the (social) costs of the errors is the following:

(a) the consequence of a false positive is that an honest player is deemed a cheater, his rightful wins are forfeited and his career is stained forever;

(b) the consequence of a false negative is that a cheater gets away with it, at least for the time being. **He may be caught next time**.

This situation calls for a small probability of false positive, even if this attains a high probability of false negative (e.g. threshold 70 for the coin above, yielding $4 \cdot 10^{-5}$ false positive and 10.38% false negative for 75% coins).

Notice that this may well result in an impossibility to catch a cheater who has a light touch (99.8% false negative for 55% coins), but this is no argument at all to raise the false positives. If this is the concern, the test needs to be reconsidered completely.

A.2 Analysis of the test used for this case

The decision of the referee that Carlo Metta used Leela in the match Italy vs Israel is the consequence of a statistical test procedure. It is known that:

1. the experiments consists of black moves $51, 53, \ldots, 149$ (50 moves)

2. the quantity measured is the *similarity index*, i.e. the number of moves deemed similar to Leela (among the first 3 choices of Leela without a drop of 5% probability from best)

3. the measurement attained for the similarity index is 98%, that is 49 moves out of 50 are similar, and this is deemed too high

4. there is no estimate of false positive or false negative; there is no estimate of costs of errors.

Choices (1)-(3) might seem common sense, but great care must be used here. In particular the correct methodology is to make all choices about the test **before** looking at the experiment. Those who have designed the test should have given details of how and why these choices have been made and guarantee that the test was not designed by someone who had already looked at the game under examination.

It must be stressed that it is unacceptable to choose the details of the test after looking at the experiment, as this increases in an uncontrolled way the probability of false positives. (In non-technical terms, there is the risk of unintentional *cherry-picking*.)

We make here a list of parameters and technical choices which could have been different and needed to be properly justified before even seeing the moves of the match under examination.

1. The choice of the window 51-149 appears arbitrary and was not properly justified before looking at Carlo's moves.

    - A more natural choice could have been to start from move 31, as a more reasonable start of the middle-game (as an example, AlphaGoZero and LeelaZero while training play the first 30 moves more randomly, not the first 50).
    - It would have also been reasonable that moves that are forced at some level were removed from the count.

2. The choice to look at the first 3 moves suggested by Leela appears arbitrary and was not properly justified before looking at Carlo's moves.

    - Many different rules could have been chosen: there are the moves suggested by the bare neural network, the moves who have the highest probability of winning according to different measurements, the moves on which the program spends the most time.
    - It could have been meaningful to consider the moves for which the neural network gives low probability, but that Leela likes after computation.
    - Moves could have been given different weights based on the amount of choices available in the situation. (This is what is done in the state-of-the-art cheat analysis for online chess games.)
    - The fact that Leela analysis is not deterministic was not taken into account. We understand that the value 98% of the statistics comes from a single simulation with Leela. Our own experiments show that this number can vary a lot (for the same game and same hardware!) and it is typically much smaller (see 1.1).
    - The fact that Leela analysis evolves with time was not taken properly into account. The person who made the analysis "usually let [Leela] run at 50k nodes, though this varied a bit depending on the move played". It is not correct to change the protocol in an uncontrolled way, according to the move, as this could potentially lead again to unintentional cherry-picking. Moreover, during the simulation time, the scoreboard of the best moves changes quite a bit, and more than 3 moves can be found briefly among the first three. So it should have been clearly stated that the 3 suggested moves considered for the statistics are only those read at the end of (say) exactly 50k nodes of simulation. However we notice that this cannot even be attained with Leela, since the only settings available are unlimited, 20k or smaller numbers. It is not possible to stop it at exactly 50k nodes.
    - An arbitrary choice of protocol may lead to involuntary cherry-picking. For example if the chosen move exits from the top three suggested by Leela just before 50k nodes, or if it enters just after 50k nodes, one may be tempted to mark the move as *similar* nevertheless. In cases like this, the statistics should be computed by someone who does not know Carlo's moves in advance (like in blind experiments from health sciences) or by an automated script, otherwise the results are at best dubious, and at worst fallacious.

A.3 Probability of errors

As explained in the preamble, a statistical test also needs to be tuned in view of the unavoidable errors and of their costs.

This wasn't done by those who devised this test.

3. The value of 98% was deemed too high, but the actual threshold of the test was not given.

- Would the referee have taken the same decision with a 96% similarity? And with 94%? What about 92%? This is not a minor point, as the choice seems psychological and not quantitative, but each of these choices brings very different consequences when applied on a regular basis.
- There was no serious attempt to estimate the probability of false positive and false negative. The confrontation with four other games played live showed that a similarity value of 98% is higher than usual, but in fact this only implies that the probability of a false positive is not much larger than 25% (one fourth).
- To give a reasonable estimate of the probability of false positives, one should compute extensive statistics of the similarity probability of the moves of many different players in many different games (as a statistician, I would say at the very least 10 players and 10 matches each). These statistics should have 2-3 repetition on the same hardware settings and moreover should be replicated on 2-3 different hardware settings. Nothing of the above was done.

A.4 Multiple tests fallacy

One of the most important lessons of statistics is that *multiple tests are dangerous.*

Say that a new wonderful anti-cheater test has a very small probability of false positive of just 0.1%. Then one honest player has a small chance of being wrongly condemned... in *one* game.

But if in one year he plays 12 important online games, the probability becomes 1.19%. If he plays in this way for 10 years, the probability becomes 11.3%. So each honest player has more than 10% probability of being wrongly condemned in a short career.

In other words, among 100 amateur players, in ten years 11 will face this unjust situation and have their career ruined.

In fact, when dealing with multiple tests, the probability of false positive must be taken much smaller accordingly. Since we don't want unjust condemnations to happen too often, the probability of false positive should be at most of the order of $2.5 \cdot 10^{-5}$. This way, among 1000 innocent amateur players, there would be no more than 2-3 wrongly convicted in 10 years, so the test would be reasonably solid.

Two objections come to mind.

(a) In this way it is impossible to get a cheater, because a test with so few false positive would have almost 100% of false negative. Correct, but:

  (a.1) the cheater can be caught next time: multiple tests apply to him also;

  (a.2) the social cost of an online-only cheater getting away with it is much much smaller than the cost of many honest players being disqualified;

  (a.3) there is always the possibility to mitigate the cost of false positives, by taking lesser measures in case of a positive: for example one could ask for a replay of the single match, or put the offender under special attention. In fact this is the common practice in all those clinical tests in which you need to accept many false positive, because false negative are unacceptable, as for the HIV test: whenever you have a positive, you perform a different independent test on the same person, without immediately raising alarm.

(b) We are not really performing so many tests: after all there was only one protest on that one match, so only one test was performed.

  This is actually false. The protest was that Carlo played similarly to Leela in that game, and this is *not independent* of the test itself.

  A fictional example will explain why this is important. Suppose the test becomes common and one wants to discredit nation A. Then he can check all the games of that nation in one (or many) tournaments and find the one match in which a player from nation A has the maximum similarity index. Then he needs only make one protest, against that game, which will then turn out to be the most suspect. The referee needs not to do may tests if the accuser did them already to choose the game.

  The above problem is not only important in the presence of such malevolent and fraudulent behavior. It is fundamental that the accusation is based on some evidence or clue independent of the test statistics. Examples of these clues are: (in live play) seeing the player go to the toilet often and in critical moments of the game, or (online) noticing a peculiar pattern in the time of the moves, or (both) checking a match where the winner was supposed to be much weaker than his opponent.

  Otherwise you are implicitly doing multiple tests, with the aggravation that it is often impossible to estimate their number.

In view of all the above arguments, the fact that no estimate of the probability of false positives was given by the expert witness, voids the validity of the decision from a statistical point of view.

We stress that it is not possible to get this estimate without much work on many games. Nevertheless as an interesting exercise, we propose some examples in which we obtain very rough estimates using binomial distribution, with threshold 98% and 94%. Let p denote the probability that a single move by an honest player marks as similar, and α the probability that the test on the match results in a false positive.

- If there are 50% forced moves with $p = 1$ and 50% open moves with $p = 0.7$, then we get (depending on threshold) $\alpha_{98} = 0.001571$ and $\alpha_{94} = 0.0332$
- If there are 30% moves with $p = 1$ and 70% moves with $p = 0.8$, then we get $\alpha_{98} = 0.003955$, $\alpha_{94} = 0.0605$
- If there are 100% moves with $p = 0.85$, then we get $\alpha_{98} = 0.002905$, $\alpha_{94} = 0.0460$

Since all the above numbers are (much) greater than 0.1%, in 10 years this test would wrongly convict (much) more than 10% of all the honest players in the Go community. In particular, more than one innocent player would be wrongly convicted every year. On the other hand, if the similarity that Metta truly had in the contested game (94% or less) was considered enough to rule a player guilty, and since all the above numbers for 94% threshold are (much) greater than 3%, then every year out of 100 innocent regular players, there would be an expected number of (much) more than 30 wrongly convicted.

A.5 Bottom line on numbers in this case

While all the details of how Leela works are not public, it is known that it is inspired by AlphaGo, but without reinforcement learning. From a practical point of view, this means that when Leela starts thinking, the initial moves it likes are what it predicts a strong human would play "by shape", without reading. Then computation begins, and Leela explores a tree of possibilities, which is nevertheless strongly based on that initial prediction.

Hence if a player's moves match Leela's quite well it may simply mean that those moves were natural to play. This is why in chess it was proposed [19] that the metric to be used is not matching strong bots or engines, *per se*. The proposed criterion is playing better than you usually do, plus independent evidence that you are playing better because you are cheating. If you play very well, you will often choose moves that the top bots do. But if you choose obvious moves, you may be playing well, but not better than usual.

In this sense it is very reasonable that the similarity index used in the test is often high. It is in fact much higher than people would usually guess, even for live games. It is then again not surprising that in the lack of a professional statistical analysis of the test, 98% similarity was simply rejected as unacceptable.

In fact statisticians recommend that when counting small numbers, percentages should never be used, as they may deceive the reader. In this case, where the most probable outcome for a move is to be similar, it would have been customary to express the statistics as a *number of mismatches*. Like, there was 1 mismatch out of 50 moves in the game in exam.

The *expected* (or average) number of mismatches in a game of the same level as the one contested, will be a small number, like 10-15. The *observed* number of mismatches in many games, on the other hand, will feature random oscillations, and for an integer number in this range, those oscillations will typically be wild. For example in the case of the simplest possible model, of Poisson random variable with mean from 10 to 15, the range of frequently observed values will be 5-22.

With a more complex model it would not be surprising at all if a value of 1 mismatch could happen not too rarely. For example, it would be incorrect to model the mismatch number of each player and each game with the same distribution: there will always be some players with lower or higher numbers, according to personal style, and moreover there will always be games with fewer move choices and games with many more, depending on the game being calm or fierce.

So if a player with a Leela-like style of play does a calm match, maybe the average number of mismatches can be lower, and given the random oscillations, a value of 1 mismatch would not be so incredible as the astonishing 98% correspondence of moves stated be the referee appears.

B About the author of this document

Francesco Morandin. Professor of mathematics in the Italian university. He is an expert in probability and Deep Learning, and he is the PhD advisor of Carlo Metta in Deep Learning.