PAIRING QUALITY - Background and plan

Geoff Kaniuk *geoff@kaniuk.co.uk*

August 2018

Contents

1 INTRODUCTION

The pairing algorithms used in most Go tournaments are non-deterministic. They are based on rules which attempt to provide pairings satisfying a set of conditions dependent on game results. For each condition c, we count the number of games N_c which meet the condition. If N_{G} is the number of games in the tournament, the ratio

$$Q_{\mathrm{c}} = \frac{N_{\mathrm{c}}}{N_{\mathrm{G}}} \tag{1}$$

provides a measure of the pairing quality for the condition c.

The pairing quality for any particular tournament is affected by a number of factors.

- number of rounds
- number of players
- player grade distribution
- player area (country and club)
- pairing rules
- probability of win

The pairing rules determine which conditions are relevant to the particular tournament in question. Swiss and McMahon are the most common non-deterministic pairing rules. Random pairing is a useful tool for gaining initial insights and testing, and can even be useful for specialised applications.

Deterministic pairing includes Round Robin and its other cyclic variants such as team tournaments. These may also provide insights as to how the tournament quality is affected by the pairing strategy.

For each set of pairing rules, pairing algorithms have been developed employing a variety of techniques designed to meet the conditions laid down by the rules. The purpose of this study is to explore the relation betwen the pairing algorithms and the pairing quality.

The purpose of this document is to outline a plan whereby this exploration can be achieved.

2 PAIRING RULES AND QUALITY

The pairing rules for the tournaments mentioned form a hierarchy of increasing complexity. This section exposes the measures of quality induced by the rules.

Random Pairing

For random pairing there are just two rules: players are paired at random and no pairings are repeated. Randomness can be achieved in the maximum cardinality matching algorithm `cmatch`[1] by shuffling the edges prior to pairing. The *no-repeat* rule can be enforced by removing edges of matched pairs prior to pairing the next round. This of course means there is a danger that at some point the matching is not perfect, and some players who met in an earlier round will not be paired. In this case the pairing is *flawed*, and for the purposes of this study the tournament is abandonded.

It is quite easy to construct examples of flawed pairings for small tournaments. For example 6 players in a 4 round event can become flawed in round 4 if a poor pairing is chosen in round 3. So for random pairing, a quality measure is:

$$Q_{\text{flawed}} = 1 - \frac{\frac{1}{2}N_{\text{unpaired}}}{N_{\text{G}}} \tag{2}$$

N_{unpaired} is the number of unpaired players.

Swiss Pairing

In a Swiss tournament with R rounds, we require no repeat games and try to pair players on the same score. In many Swiss tournaments pairing of players in a *same-score* group are paired with various seeding and splitting strategies.

Quality for same-score pairings - *even* games is:

$$Q_{\text{even}} = \frac{N_{\text{even}}}{N_{\text{G}}} \tag{3}$$

N_{even} is the number of even games in the tournament.

Quality for seeding within a same-score group is given by:

$$Q_{\text{seed}} = \frac{N_{\text{seed}}}{N_{\text{G}}} \tag{4}$$

N_{seed} is the total number of games where the seeding condition is *exactly* satisfied.

It may be that this defintion is too strict, as in later rounds the splitting strategy can fail, forcing players in the same seed group to play each other. A possibly more forgiving definition for N_{seed} is the number of games in which the players are chosen from the *separate* subgroups.

Quality for colour alternation is defined as the ratio:

$$Q_{\text{colour}} = \frac{N_{\text{cswap}}}{N_{\text{G}}} \tag{5}$$

N_{cswap} is the number of games where *both* players are allocated colours opposite to their colours at the previous round.

McMahon Pairing

McMahon pairing inherits all the rules from Swiss pairing with the addition of a bar setting and an *area-rule*, which tries to avoid pairing players from the same country or club. One of the aims of the bar setting is to ensure that players above the bar play each other. So a quality measure associated with the bar setting is:

$$Q_{\text{peer}} = \frac{N_{\text{peer}}}{N_{\overline{\text{bar}}}} \tag{6}$$

$N_{\overline{\text{bar}}}$ is the number of potential games above the bar and N_{peer} is the number of games where *both* players are above the bar.

The quality for the area-rule is defined by:

$$Q_{\text{area}} = \frac{N_{\text{mixed-country}} + N_{\text{same-country mixed-club}}}{N_{\underline{\text{bar}}}} \tag{7}$$

The number of games between different countries is $N_{\text{mixed-country}}$. The number of games between players from the same country, but different clubs is $N_{\text{same-country mixed-club}}$. The number of games below the bar is $N_{\underline{\text{bar}}}$.

3 IMPLIED QUALITIES

So far we have considered the qualities arising directly from the pairing rules. Others arise from *desired* tournament characteristics. There are some that emerge from the probabilistic behaviour of the system comprised of (tournament entry, pairing rules, and probability of win). The main ones are:

- **Unique winner**. In most tournaments a unique winner is desirable, but can often only be enforced by appropriate additional tie-breaks applied to the final ranking.
- **Uneven games**. These are games between players with different scores but no handicap. There are a host of strategies for pairing such games, but these games are supposed to be low in number.
- **Handicap games**. These are inevitable at the bottom end of the entry and poor weight settings may result in too large a number.
- **Average opponent grade**. This is a critical parameter in the algorithm for setting the bar [2] and may play a rôle in the pairing quality for players at or below the bar grade.
- **Average wins**. It was found in [2] that the average number of wins for players well below the bar is $R/2$ for tournaments with R rounds and reasonably well populated grade groups. It will be useful to monitor this quantity for any variation with change of weight.
- **Ranking**. Assume all players enter a McMahon tournament with grades correctly reflecting their rating. Then we can expect the final player ranking to be well correlated with the players' grades.

4 MAXIMUM WEIGHTED MATCHING

For Swiss and McMahon tournaments, the maximum weighted matching algorithm `wmatch` [1] produces the pairing. A model for the allocation of weights to each potential pair has been discused in [3]. The general principle is to establish a rule-deviation D_{ij}^k for each rule k governing the pairing between any two players i and j. The weight associated with the rule deviation has the form:

$$\omega_{ij}^k = W_k \psi(D_{ij}^k)$$

Here W_k is the *weight coefficient*for the rule k, and $\psi(D)$ is the *weight factor* associated with the given rule-deviation. The total weight allocated to the pair is the sum:

$$W_{ij} = \sum_k \omega_{ij}^k$$

The maximum weighted matching algorithm finds a sequence of ij pairs for which the sum over all pairs in the sequence is a maximum. Thus it maximises:

$$W = \sum_{ij} W_{ij}$$

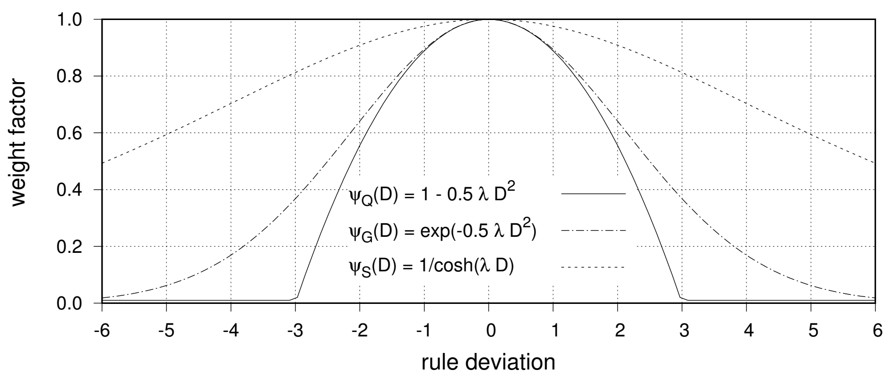

Figure 1: Weight factor functions showing examples for quadratic: ψ_Q; gaussian ψ_G; and sech ψ_S. λ is chosen so that $\psi_Q(D) = 0$ at $D = 3$.

The weight factor function ψ fulfills the same rôle as Open Gotha's *concavity function* [4] in preventing degeneracy when combining the weight components in the above sum. Figure 1 shows possible weight factors compared to the quadratic one used in Open Gotha (and effectively in MacMahon [5]). The essential requirement is that the component contributes maximum weight when the rule deviation is near zero and much less weight when the rule deviation is large.

It is evident from the the shape of ψ that if λ is very small then the weight component ω_{ij}^k provide very little selection: i.e. the same total weight can be achieved for pairs with very different (or large) rule deviations. So the pairings will be fairly random. On the other hand if λ is very large then only rule deviations of exactly zero will lead to pairs with non-zero weights. This may result in a flawed pairing.

The weight coefficients prioritise the relative importance of the rules, and it is the aim of this study to determine which qualities are sensitive to changes in values of the weight parameters (W_k, λ_k).

5 MONTE CARLO DATA COLLECTION

We can represent a particular choice of weight parameters by a single point in a multi-dimensional weight space $\mathcal{W} = \mathbb{R}^{N_W}$. For a Swiss pairing N_W is 8 (two parameters for each of score, colour, even and uneven rules). For McMahon this increases to 10 for the area-rule, and maybe more if there are special handicap or banding rules.

For Swiss pairing we have identified 4 qualities of interest, extending to 6 for McMahon pairing. There are a further 6 implied qualities (Section 3) potentially of interest in all pairing rules. Again we can collect all the qualities to form a point in *quality space* $\mathcal{Q} = \mathbb{R}^{N_Q}$.

We can expose the relationship between weight parameters and tournament quality via a Monte Carlo process. We simulate a fixed sequence of tournaments for a given set of weights $\mathbf{W} \in \mathcal{W}$. Each tournament τ in the fixed sequence $\mathcal{T}$ gives rise to a tournament quality $\mathbf{Q}_\tau \in \mathcal{Q}$.

We can accumulate the qualities for all the tournaments in $\mathcal{T}$ to give an overall quality $\mathbf{Q_W}$ for the weight $\mathbf{W}$. Thus define the total component Q_c as the ratio of the count sums over all tournaments. This easily leads via (1) to Q_c expressed as a weighted sum of the individual component qualities $q_c(\tau)$:

$$Q_c = \sum_\tau n_c(\tau)/N_G = \sum_\tau \Gamma_G(\tau) q_c(\tau) \tag{8}$$

$\Gamma_G(\tau)$ is proportional to the number of games in the tournament τ.

The following algorithm illustrates the essentials of the process using three independent random number generators for weights, tournament creation and result simulation:

Algorithm 1. `Quality Weight Probe`

step 1 : Set up a range for each weight parameter.

step 2 : Choose a weight point $W \in \mathcal{W}$ at random within the range.

step 3 : Initialise the tournament generator[1].

step 4 : Generate a sample tournament $\tau \in \mathcal{T}$ from a known distribution.

step 5 : Pair the tournament M_P times accumulating a quality point $\mathbf{q}(\tau)$.

step 6 : Use $\mathbf{q}(\tau)$ to update the quality point $\mathbf{Q}(\mathbf{W})$.

step 7 : Continue at step 4 M_T times.

step 8 : Record $\mathbf{W}$ and $\mathbf{Q}(\mathbf{W})$.

step 9 : Continue at step 2 M_W times.

step 10 : Halt.

The outcome of this simulation is a set of pairs $\mathcal{M} = \{(\mathbf{W}_s, \mathbf{Q}_s)\}$. This is stored as a single file.

The software and data will be made available as an open source project.

[1] This ensures that the *same* sequence $\mathcal{T}$ is generated for each new weight W

6 THE PLAN

At this stage it is unknown what the characteristics of the relation between weight and quality are. There are two immediate questions that need to be answered:

- **Continuity**. If two weights are similar are the associated qualities also similar?
- **Sensitivity**. It is quite likely that 2 very different weight settings say $\mathbf{W}_a$ and $\mathbf{W}_b$ might produce very similar qualities, so $\mathbf{Q}_a \simeq \mathbf{Q}_b \simeq \mathbf{Q}$. Consider the set of all weights along the line joining $PW_a, \mathbf{W}_b$. Are the qualities for these weights also similar to PQ?

The answers to these questions will tell us how well behaved (or otherwise) the relation between quality and weight is. The ultimate hope of this study is to identify a region of weight space giving good quality for a large range of tournaments and every weight in the region.

Simulations will be carried out for the three pairing methods discussed earlier.

- **Random pairing**. This should provide suitable values for the sample counts M_P, M_T, M_W identified in the above algorithm. Random pairing uses `cmatch`, and if a tournament is flawed under `cmatch` it is likely that the entry is also unsuitable for running as a Swiss or McMahon tournament.

  There are no rules governing colour alternation or even games. Nevertheless we record all qualities which act as controls to validate the simulation.

  There are no weights to change, instead we record quality details for each tournament in the sequence $\mathcal{T}$. In this way we can identify flawed tournaments and any outliers with other very poor quality components.

- **Swiss pairing**. We start with just two rules: try to minimise the score difference and try to alternate colours. This means we have just 4 weight parameters to vary and the data set $\mathcal{M}$ can easily be managed on a spreadsheet. Initial analysis will examine the extent of clustering in the quality components.
- **McMahon pairing**. Repeat the simulation for the same Swiss conditions. Compare the analysis results.

The above process can then be repeated including more weight components at each stage. A report will be issued as each stage is completed.

References

[1] Harold.N Gabow, Implementation of Algorithms for maximum matching on nonbipartite graphs(chapters 2 and 4),PhD Thesis,Stanford University,1974

[2] Geoff Kaniuk,Determination of the McMahon Bar, August 2015 www.kaniuk.co.uk/articles/pairing/mcmahon-bar.pdf

[3] Geoff Kaniuk,Weights in McMahon Pairing - Revised, May 2015 www.kaniuk.co.uk/pairing/mcmahon-weights-revised.pdf

[4] Luc Vannier,Open Gotha user's guide, vannier.info/jeux/download/GothaHelp_en.pdf

[5] Christoph Gerlach,Thesis,Section 2.3.1 www.cgerlach.de/go/diplom.pdf