Charting the Best Play: Introducing the XG Analysis Table

by Gus Contos

Backgammon analysis has changed over the last fifteen years, reflecting the increasing power and influence of computers. Once upon a time, an author examining a checker play would furnish a diagram of the position, and then write a few paragraphs discussing the pros and cons of the likely candidates. The diagram and the discussion were the two standard components of any analysis. Here's a simple example, constructed to examine a common situation:

The obvious choice is between playing it safe with 6/1, 5/4, and ripping a checker off the six point, with the risk of leaving a shot after 5-5 and 6-6. In the old days, the author would now try to calculate the chances of being hit, estimate the chances of winning after being hit, and guess the gammon percentages. Our author might even support those estimates by —rolling out the two candidates—rolling dice and playing out the positions resulting from each move, taking many hours to gather the data from a few hundred trials.

All this changed in the 1990s, when the wonders of neural-network technology made it possible to offer backgammon programs (bots) of world-class strength for sale at reasonable prices. These programs can, among other things, carry out very efficient simulations—rollouts performed at lightning electronic speed, as opposed to rollouts performed at a glacial human speed. Instead of spending those tedious hours doing hundreds of manual rollouts, an author can now take a few minutes to put a position (or many positions) into the computer, give the bot some instructions, and then walk away while the computer does the work. The bot will roll out the candidates, record the results, perform some statistical analyses to test the validity of the data, and present all the relevant figures neatly in a table or chart. Authors started including these analysis tables in their write-ups, and the tables now appear as the standard third component in any serious analysis of a backgammon position. Here's the analysis table for our example:

1.	Rollout[1]	6/0ff	eq: +1.076
	Player:	96.19% (G:16.20% B:0.09%)	Conf: ± 0.004 (+1.072...+1.080)
	Opponent:	3.81% (G:0.00% B:0.00%)	Duration: 2 minutes 14 seconds
2.	Rollout[1]	6/1 5/4	eq: +1.068 (-0.007)
	Player:	98.17% (G:10.73% B:0.04%)	Conf: ± 0.002 (+1.066...+1.070)
	Opponent:	1.83% (G:0.00% B:0.00%)	Duration: 2 minutes 13 seconds
3.	3 ply	6/1 4/3	eq: +1.019 (-0.057)
	Player:	96.18% (G:10.32% B:0.03%)	
	Opponent:	3.82% (G:0.00% B:0.00%)	
4.	3 ply	5/0ff 4/3	eq: +0.971 (-0.104)
	Player:	92.35% (G:14.26% B:0.06%)	
	Opponent:	7.65% (G:0.00% B:0.00%)	
5.	1 ply	5/0ff 1/0ff	eq: +0.487 (-0.588)
	Player:	71.37% (G:15.36% B:0.43%)	
	Opponent:	28.63% (G:0.00% B:0.00%)	

[1] 1296 Games rolled with Variance Reduction.
Moves and cube decisions: 3 ply

For this analysis, I used eXtreme Gammon (XG), a commercial program endorsed by the USBGF. Other backgammon programs may use a somewhat different format, but will yield the same kind of information. Serious students of the game will be familiar with this layout, and can take in its meaning with a glance. Newcomers and casual players, however, may find this mass of numbers cryptic and confusing. For their benefit, this article will explain what it all means.

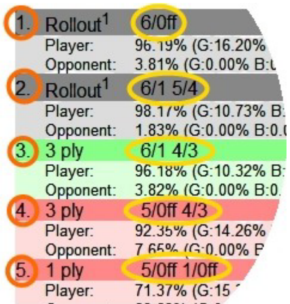

Figure 1

Notice first that XG has listed five candidate moves, with the strongest at the top, as indicated by the ranking numbers in the far left column (circled in orange in Figure 1). There are more possible plays, of course, but XG saves space and time by considering only the five strongest. The description of each specific play is listed a few spaces to the right. (circled in yellow.)

Ranking numbers and move descriptions are pretty much self-explanatory, but what are those items we see listed between them? (Figure 2)

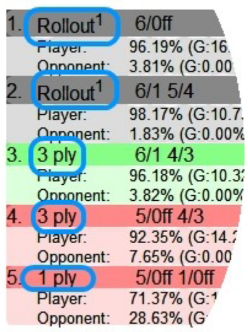

Figure 2

These terms tell us how XG evaluated each of the candidates. To understand them, we must examine the several methods XG can use to assess the merits of any given play.

Plies, More Plies, and Rollouts

The simplest method is 1 ply, as listed for our fifth candidate. We can think of plies as an indication of how far ahead the computer looks in making its judgment. For a 1-ply analysis, XG inspects the position resulting from the play, notes all the relevant features, and estimates the likelihood of the various outcomes - wins, gammons, and backgammons for both sides.

To make a 2-ply analysis, XG makes the candidate play, then considers all 21 distinct dice rolls for the opponent. XG makes what it thinks is the best play for each roll, carries out a 1-ply analysis for each resulting position, and then com-bines the figures to get an overall average for the group. A 2-ply analysis does not appear in our example. A 3-ply analysis carries the look-ahead one step further: XG makes the candidate play, considers the 21 possible responses by the opponent, then for each of those 21, considers the 21 replies by the player. This yields 441 positions, for each of which XG performs a 1-ply analysis, then computes a single average for the whole group. In our example, we see that XG used the 3-ply evaluation for candidates 3 and 4.

This pattern of looking ahead another move can be extended to four plies and higher. The tradeoff here is clear: each ply adds one more roll and move, making the assessment more accurate, but each ply also multiplies the number of positions to be examined by 21, increasing the time needed for the evaluation by the same factor. There is also the possibility that our bot's assessments may be flawed; if this were true, adding more plies would not help. For these reasons, ply evaluations are useful for screening all legal moves and finding the likely candidates—but to make the most accurate evaluations and choose between close alter-natives, it's normally best to use rollouts.

We can see how all this worked in our example. XG dismissed candidate 5 after a 1-ply evaluation. Candidates 3 and 4 merited 3-ply examinations, but were not serious contenders. The top two candidates were very close, so I instructed XG to roll them out.

Let's look more closely at the chart for candidate move 1:

1. Rollout[1]	6/Off	eq: +1.076
Player:	96.19% (G:16.20% B:0.09%)	Conf: ± 0.004 (+1.072...+1.080)
Opponent:	3.81% (G:0.00% B:0.00%)	Duration: 2 minutes 14 seconds

XG tells us here it performed a rollout: It played 6/Off, rolled the dice, played both sides until the game was decided one way or another, recorded the result, then repeated the whole process many times. XG also tells us, on the third line, that it took 2 minutes and 14 seconds to perform the rollout. But how did the bot pick the best moves during the rollouts, or decide about cube actions? And how many games are we talking about—ten, a hundred, a thousand ? Those questions are answered in the footnote at the bottom of the chart:

[1] 1296 Games rolled with Variance Reduction.
Moves and cube decisions: 3 ply

The second line of the footnote answers our first question: all moves and cube *(Continued from page 7)*
decisions were made with the benefit of 3-ply analysis. The footnote also answers our other question: XG rolled out 1296 games—with something called Variance Reduction. H'm, what is this Variance Reduction?

Briefly, Variance Reduction is a method of adjusting for lucky rolls that might otherwise skew our picture of what happens, on average, when playing out a position. Another way of reducing the distorting effect of luck would be to roll out a larger sample of games. In most cases, though, using Variance Reduction is far more efficient than increasing the number of roll-outs, even with the extra work needed to judge the —luckiness‖ of each roll.

Now that we know how XG came up with the numbers in the table, it's time to see what those numbers mean.

Numbers By the Half Dozen

Looking again at the table entry for the first candidate, 6/Off, we find this listing in the second line:

Player: 96.19% (G:16.20% B:0.09%)

This tells us that playing 6/Off won 96.19% of the games, and of this number, 16.20% ended in gammons or backgammons, with 0.09% being specifically backgammons. Thus, the player won 79.99% single games, 16.11% gammons, and 0.09% back-gammons. The third line: Opponent: 3.81% (G:0.00% B:0.00%) gives us the corresponding numbers for the opponent, who won only 3.81% of the games after White plays 6/Off.

Looking back at the full analysis table, we see the same six-number listing for each of the five candidates. In our example, we know that the Opponent gammons and backgammons will always be zero, leaving four percentages to consider for each move. Short races would involve only two. But most contact positions will generate six different percentages for each move—six numbers to compare across all candidate plays. It would be simpler if we had these results combined into a single number for each move, one number to measure that play's overall strength, This is exactly what XG does for us: for each move, it computes a net value for the resulting position—the ***equity*** (eq: in the analysis table).

Keeping in mind that higher is better, we (and XG) are looking for the play with the highest equity (6/Off in this example). Inferior plays lose equity, and the exact amount each loses appears in parentheses. In our example, we see that the third best play, 6/1 4/3, loses 0.057 in equity compared to the top play. This gives us a way of measuring the magnitude of an error—how costly it is.

A Tape Measure for Errors

The ability to gauge the cost of errors is important; it shows us where we need to improve. There is a problem, however, in trying to compare the magnitude of errors from various positions: the cost of any given error can vary widely depending on the situation in which it appears. Our example is from early in a match. Another time, we may want to examine a similar position from late in a match, or from a money game, or maybe with the cube still centered. For a meaningful comparison of errors, we need a standard scale that can provide a consistent measurement across all situations. For this reason, XG has adjusted the equities in our chart to provide the standard scale we need. These are *normalized* equities.

In this article, we won't examine how XG computes normalized equities, or whether they are truly accurate across all situations. We'll simply point out that all widely used backgammon programs provide normalized equities, and virtually all serious players accept them as valid and useful tools. After all, we can use the same tape measure for carpets, picture frames, or drain pipes. Think of normalized equities as a tape measure for errors.

Confidence Game

Let's look once more at the top two entries in our chart:

1. Rollout[1]	6/0ff		eq: +1.076
Player:	96.19% (G:16.20% B:0.09%)	Conf: ± 0.004 (+1.072...+1.080)	
Opponent:	3.81% (G:0.00% B:0.00%)	Duration: 2 minutes 14 seconds	
2. Rollout[1]	6/1 5/4		eq: +1.068 (-0.007)
Player:	98.17% (G:10.73% B:0.04%)	Conf: ± 0.002 (+1.066...+1.070)	
Opponent:	1.83% (G:0.00% B:0.00%)	Duration: 2 minutes 13 seconds	

Those numbers labeled Conf:± are *confidence intervals*; they remind us that any equities computed after rolling out a position will always carry some degree of uncertainty. There is always the possibility that one side benefited from luckier dice, even after collecting a sample of thousands of games, and even after applying the techniques of variance reduction. For every rollout, XG uses standard methods to assess this uncertainty, and lists it as a confidence interval—the range in which the true value lies. The larger the confidence interval, the greater the uncertainty.

Reading from the table, we see that the true value of the equity for the first play (6/Off) could differ from the listed value (+1.076) by 0.004 in either direction; it could be as low as 1.072 or as high as 1.080. For the second play (6/1 5/4), the confidence interval was smaller: plus or minus 0.002, so the true value could be as low as 1.066 or as high as 1.070.

Once we have a basic understanding of confidence intervals, we have a way of judging the accuracy of the rollout data: an objective measure of how reliable the numbers are. If we aren't satisfied, we can have XG do more work to improve our accuracy by, for example, extending the number of rollouts. In our example, we can be reasonably confident that 6/Off is best.

This has been a brief survey of the XG analysis table. There is more to say about all the topics we've mentioned—much more. Yet even this short introduction should make it clear why the analysis table has become an indispensable part of all serious backgammon discussions. By providing a wealth of information in a compact form, it provides one of our most valuable tools for charting the best play.