

CSE702 Week 7: Main Modeling Issues

Here is an overview:

1. Is the loglog-linear model "naturally ordained" or can it be improved at minimal "Occam's Razor" cost? (I regard instilling specific chess knowledge beyond the notions of move values and depth of thinking as a high cost.)
2. To what extent do the main cross-checking quantities behave like i.i.d. (normal) random variables?
3. How should be bias in other category quantities be identified and corrected? Should it be? (A category like **KnightMove** can be corrected by fiat by post-processing the model's output probabilities. The correction is oblivious to the specific chess knowledge of knights; the category **AdvancingMove** etc. would be corrected the same way.)
4. Why does maximum likelihood expectation (MLE) give biased results? Can we correct it? Could the correction behave better regarding the chess-specific quantities?
5. Are Elo rating level and time-to-think fully fungible? Or would the model trained on, say, 5-minute Blitz chess behave differently from translating the Standard-chess model down by 575 Elo points?
6. Is the "EWN" notion of difficulty effective?

| Selection Test | ProjVal | St.Dev | Actual; | Proj% | Actual% | 2sigma range | z-score | | BrierSc | LikelySc |
|----------------|---------|--------|---------|---------|---------|-----------------|------------|---------------|---------|----------|
| Delta01-10 | 1745.99 | 31.65: | 1728.00 | 32.76%: | 32.42% | 31.57%--33.95%, | z = +0.57, | engm% = 0.00 | 1.547 | 1.810 |
| Delta11-30 | 2223.05 | 37.44: | 2222.00 | 27.84%: | 27.82% | 26.90%--28.77%, | z = +0.03, | engm% = 0.00 | 2.170 | 2.671 |
| Delta31-70 | 1655.54 | 34.83: | 1674.00 | 16.45%: | 16.63% | 15.76%--17.14%, | z = -0.53, | engm% = 0.00 | 3.371 | 6.083 |
| Delta71-150 | 754.40 | 24.48: | 760.00 | 6.89%: | 6.94% | 6.44%-- 7.34%, | z = -0.23, | engm% = 0.00 | 2.410 | 6.674 |
| Error025 | 3326.69 | 46.20: | 3395.00 | 23.54%: | 24.03% | 22.89%--24.20%, | z = -1.48, | engm% = 0.00 | 4.114 | 8.710 |
| Error050 | 1768.41 | 35.79: | 1789.00 | 12.54%: | 12.69% | 12.04%--13.05%, | z = -0.58, | engm% = 0.00 | 3.335 | 9.248 |
| Error100 | 743.54 | 24.27: | 762.00 | 5.30%: | 5.43% | 4.95%-- 5.64%, | z = -0.76, | engm% = 0.00 | 3.727 | 11.869 |
| Error200 | 274.19 | 15.22: | 278.00 | 1.97%: | 1.99% | 1.75%-- 2.18%, | z = -0.25, | engm% = 0.00 | 2.532 | 14.319 |
| Error400 | 106.61 | 9.68: | 79.00 | 0.79%: | 0.58% | 0.64%-- 0.93%, | z = +2.85, | engm% = 0.00 | -1.788 | 5.008 |
| EvalGoesToZero | 3338.01 | 32.96: | 3236.00 | 26.62%: | 25.81% | 26.10%--27.15%, | z = -3.09, | engm% = 25.75 | 6.985 | 12.659 |

| Selection Test | ProjVal | St.Dev | Actual; | Proj% | Actual% | 2sigma range | z-score | | BrierSc | LikelySc |
|----------------|----------|--------|----------|---------|---------|-----------------|-------------|---------------|---------|----------|
| PawnMove | 6445.34 | 55.43: | 6487.00 | 23.92%: | 24.07% | 23.50%--24.33%, | z = +0.75, | engm% = 25.63 | 1.633 | 4.565 |
| KnightMove | 4108.08 | 43.15: | 4740.00 | 21.39%: | 24.68% | 20.94%--21.84%, | z = +14.64, | engm% = 23.76 | 14.548 | 16.885 |
| BishopMove | 4230.98 | 43.98: | 4421.00 | 20.67%: | 21.60% | 20.24%--21.10%, | z = +4.32, | engm% = 20.70 | 3.614 | 5.403 |
| RookMove | 6095.77 | 51.15: | 5595.00 | 24.69%: | 22.66% | 24.28%--25.11%, | z = -9.79, | engm% = 22.97 | -6.084 | -4.582 |
| QueenMove | 4112.25 | 40.86: | 3905.00 | 22.86%: | 21.71% | 22.41%--23.32%, | z = -5.07, | engm% = 21.48 | -1.417 | 0.964 |
| KingMove | 2745.59 | 38.52: | 2590.00 | 10.10%: | 9.52% | 9.81%--10.38%, | z = -4.04, | engm% = 9.17 | -3.914 | -2.790 |
| Castling | 302.11 | 13.80: | 435.00 | 14.97%: | 21.56% | 13.60%--16.34%, | z = +9.63, | engm% = 20.27 | 9.561 | 9.227 |
| Capture | 5323.93 | 38.21: | 6673.00 | 22.03%: | 27.61% | 21.71%--22.34%, | z = +35.31, | engm% = 26.54 | 14.638 | 23.120 |
| NonCapture | 18847.07 | 38.21: | 17498.00 | 77.97%: | 72.39% | 77.66%--78.29%, | z = -35.31, | engm% = 73.46 | 14.638 | 23.177 |
| Promotion | 18.74 | 2.40: | 12.00 | 23.42%: | 15.00% | 17.42%--29.42%, | z = -2.81, | engm% = 15.00 | -1.794 | -1.670 |
| AdvancingMove | 16844.82 | 63.70: | 18241.00 | 60.28%: | 65.27% | 59.82%--60.73%, | z = +21.92, | engm% = 64.86 | -3.438 | -2.302 |
| RetreatingMove | 5351.92 | 50.60: | 4542.00 | 19.51%: | 16.55% | 19.14%--19.88%, | z = -16.01, | engm% = 16.80 | -7.556 | -5.440 |
| SidewaysMove | 5865.26 | 54.68: | 5279.00 | 21.44%: | 19.30% | 21.04%--21.84%, | z = -10.72, | engm% = 19.48 | -6.063 | -4.640 |

| | | | | | | | | | | | |
|--------------------|----------|--------|----------|---------|---------|----------|---------|--------------|----------------|---------|--------|
| CheckingMove | 1094.41 | 21.93: | 1305.00 | 8.84%: | 10.54% | 8.49%-- | 9.20%, | z = +9.61, | engm% = 10.42 | 8.433 | 12.806 |
| EngineMove | 13387.01 | 75.13: | 13387.00 | 47.51%: | 47.51% | 46.98%-- | 48.05%, | z = -0.00, | engm% = 100.00 | 5.625 | -5.775 |
| PlayedMove | 9819.18 | 64.65: | 28176.00 | 34.85%: | 100.00% | 34.39%-- | 35.31%, | z = +283.94, | engm% = 47.51 | 321.835 | 440.82 |
| SamePieceAsPrevMov | 666.41 | 15.57: | 819.00 | 12.03%: | 14.78% | 11.47%-- | 12.59%, | z = +9.80, | engm% = 14.17 | 2.312 | 4.655 |
| EqualTopMove | 14698.62 | 73.88: | 14573.00 | 52.31%: | 51.86% | 51.78%-- | 52.84%, | z = -1.70, | engm% = 100.00 | -4.115 | -0.851 |

Should the following anomalies be fixed *post-hoc* (i.e., by fudging the probabilities at the very end to match distributional results) or left as-is?

- Knight moves
- Capturing moves
- Advancing moves
- Castling---maybe to a lesser extent.

Scripted Runs

These can be saved in files or simply copied and pasted. Here is a current example:

```
./ir Kom13 EWN
JospemTTKom13
addOutputFile JospemResults.txt JospemResults
clearTurns addTurns /projects/regan/Chess/CSE702/AIF/MartinezAlcantaraTTG4BlitzJan-
Mar2024_Kom133d20-30pv64.aif
newFilters
PlayerToMove Alcantara2m Alcantara
done
attach tleq60 n
showTrial n

perfTest useRating 1975 goTest
runIPR 2100 MartinezAlcantaraTT2024Kom13IPR

newFilters
OnMoveFacing vKramnik Kramnik
done
showTrial y
perfTest goTest
perfTest useRating 1975 goTest slideScale fromRating 2000 goScaleSpec
runIPR 2200 JospemVKramnikKom13IPR

detach vKramnik n
newFilters
```

```
OnMoveFacingOneOf JospemVElite Kramnik Carlsen Gormally Niemann Caruana Bartel
Shankland Firouzja Guijarro Shirov Bacrot Sargsyan Sjugirov Tabatabaei .
done
showTrial y
perfTest goTest
loadTrialSpec MartinezAlcantaraTT2024Kom13IPR perfTest goTest
perfTest useRating 1975 goTest
runIPR 2100 JospemCherryPickIPRKom13EWN
```

For fast chess using Stockfish 11, I use a 5-pawn cap rather than a 4-pawn cap on advantage. So after the "attach tleq60 n" command, also do

```
attach pnew5norm n detach pnew4norm n
```

Of course this also involves changing "Kom133" and then "Kom13" to "SF11" everywhere. We can also do the experiment in EWN mode rather than UW.

"Five-Finger Exercises" for Over Spring Break

1. How correlated is zM_1 to zM_2 in performance tests based on ratings (as opposed to after fitting) of players in large tournaments?
2. How close is zM_2 to standard Gaussian? (again, in results files of rating-based tests)
3. How close is zM_3 to standard Gaussian?
4. How close is zM_4 to standard Gaussian?
5. How close is zM_5 to standard Gaussian?
6. How close is $zDelta01-10$ to standard Gaussian?
7. How close is $zDelta11-30$ to standard Gaussian?
8. How close is $zDelta31-70$ to standard Gaussian?
9. How close is $zDelta71-150$ to standard Gaussian?
10. How correlated are the $zDelta$ quantities to each other?