A Rigorous Treatment of Maximum Likelihood for the Pentanomial Model

Isak Ellmer

November 2023

1 Introduction

We consider chess matches between two players with a fixed opening book and repeated openings. The openings are selected randomly from the opening book. A pair of matches is two matches where the player 1 plays white in a specific opening and then plays black in the same opening. We will sometimes refer to such a pair as a single match for simplicity.

1.1 Elo and the Pentanomial Model

The players are awarded points for each match in a pair. A loss gives 0 points, a draw $\frac{1}{2}$ and a win 1. If the probability of losing, drawing and winning are l, d and w respectively then the expected score of a match is $\frac{1}{2}d + w$.

When the openings are repeated this is no longer accurate. Consider the case where player 1 plays white in a specific opening and wins. There is a possibility that player 1 won, not because of superior play, but because white was heavily favored in this opening. When it is player 2's turn to play white in the same opening the expected score is no longer $\frac{1}{2}d + w$, but higher.

A partial solution to this problem is to consider the pentanomial model, as opposed to the old trinomial model. Instead of scoring a single match, we instead score the match pair. A loss still gives 0 points, a draw now gives $\frac{1}{4}$ and a win $\frac{1}{2}$. The sum of the individual match scores is then the score of the match pair. For example, winning one match and drawing one results in a score of $\frac{3}{4}$. The possible outcomes for a match pair are thus the scores 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ and 1 with respective probabilities p_0, $p_{\frac{1}{4}}$, $p_{\frac{1}{2}}$, $p_{\frac{3}{4}}$ and p_1, hence the name pentanomial.

The expected score is now given by $\frac{1}{4}p_{\frac{1}{4}} + \frac{1}{2}p_{\frac{1}{2}} + \frac{3}{4}p_{\frac{3}{4}} + p_1$. Letting $\theta = (p_0, ..., p_1)$ we take $\phi(\theta) = \frac{1}{4}p_{\frac{1}{4}} + \frac{1}{2}p_{\frac{1}{2}} + \frac{3}{4}p_{\frac{3}{4}} + p_1$ to be this expected score of a match. This is related to Elo in the following way

$$\phi(\theta) = \frac{1}{1 + 10^{-\Delta E/400}}$$

where ΔE is the Elo difference of the two players [1].

1.2 Hypothesis Testing

We denote by f_θ, the probability density of the outcome of a match with probabilities $\theta = (p_0, ..., p_1)$. Our goal is not to make any inference about θ, but to test hypotheses on the Elo rating difference between two players. Our hypotheses will thus be of the form

$$H : \phi(\theta) = C = \frac{1}{1 + 10^{-\Delta E/400}}$$

where we will always have $0 < C < 1$.

In order to test the hypothesis $H_0 : \phi(\theta) = C_0$ versus $H_1 : \phi(\theta) = C_1$ we perform a generalized sequential probability ratio test. That is for i.i.d observations $x_1, ..., x_N$ we watch the quantity

$$\frac{\sup_{\phi(\theta)=C_0} \prod_{j=1}^N f_\theta(x_j)}{\sup_{\phi(\theta)=C_1} \prod_{j=1}^N f_\theta(x_j)}$$

in some interval $[a, b]$. Or equivalently we watch

$$\sup_{\phi(\theta)=C_0} \sum_{j=1}^N \log f_\theta(x_j) - \sup_{\phi(\theta)=C_1} \sum_{j=1}^N \log f_\theta(x_j)$$

in some interval $[A, B] = [\log a, \log b]$.

2 Calculating the Supremum

We begin by noting that our outcome can only take 5 discrete values. Instead of the observations $x_1, ..., x_N$ we therefore consider the amount of times each outcome was observed. That is $N_0, N_{\frac{1}{4}}, N_{\frac{1}{2}}, N_{\frac{3}{4}}$ and N_1 where N_α is the amount of observations of outcome α. We will also normalize the problem in the sense that $n_\alpha = \frac{N_\alpha}{N}$ so that the sum of the n_α is 1.

Our expression now simplifies to

$$\sup_{\phi(\theta)=C} \sum_\alpha n_\alpha \log f_\theta(\alpha) = \sup_{\phi(\theta)=C} \sum_\alpha n_\alpha \log p_\alpha$$

where the sum is over $\alpha \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. We want to maximize $\sum n_\alpha \log p_\alpha$ subject to $\phi(\theta) = \sum \alpha p_\alpha = C$ and $\sum p_\alpha = 1$. Note that although we now maximize the normalized problem, our supremum value when testing hypotheses, will still be

$$\sup_{\phi(\theta)=C} \sum_\alpha N_\alpha \log p_\alpha.$$

This will however not cause any problems as

$$\sup_{\phi(\theta)=C} \sum_\alpha N_\alpha \log p_\alpha = N \sup_{\phi(\theta)=C} \sum_\alpha n_\alpha \log p_\alpha.$$

2.1 The Generalization

The problem will in fact be easier to solve if we generalize it. Let $\alpha_1 < ... < \alpha_N$ be real numbers and let $f : \{\alpha_1, ..., \alpha_N\} \to [0, 1]$ be a probability density. We denote $f(\alpha_j)$ by p_j for simplicity. Suppose that a sample $n_1, ..., n_N$ is taken from f. As before we want to maximize $g(\theta) = \sum_j n_j \log p_j$ subject to $\sum_j p_j = 1$ and $\sum_j a_j p_j = C$ where $\theta = (p_1, ..., p_N)$. We denote this region for θ, over which we try to maximize g, by R.

Note that we need to have $\alpha_1 \leq C \leq \alpha_N$, for otherwise there will be no solution. In the case where $C = \alpha_1$ or $C = \alpha_N$ the solution to the problem is trivial and is given by $p_1 = 1$ and $p_N = 1$ respectively. We can thus assume that $\alpha_1 < C < \alpha_N$. This will, as already mentioned, always be the case in our application to chess.

We start with the following very useful lemma.

Lemma 2.1. *Suppose that $\sum_j p_j = 1$ and $\sum_j \alpha_j p_j = C$. If $i < k < l$ and $p_k > 0$, then there exists arbitrarily close p'_j that still satisfies these equations, and such that $p'_i > p_i$, $p'_k < p_k$ and $p'_l > p_l$ and $p'_j = p_j$ for all other j. If instead $p_i > 0$ and $p_l > 0$ then the inequalities will be the other way.*

In particular we have $\Delta p_i + \Delta p_k + \Delta p_l = 0$, $\alpha_i \Delta p_i + \alpha_k \Delta p_k + \alpha_l \Delta p_l = 0$ and

$$\frac{\Delta p_i}{\Delta p_l} = \frac{\alpha_l - \alpha_k}{\alpha_k - \alpha_i}.$$

Proof. For small enough r let $p'_i = p_i + r(\alpha_l - \alpha_k)$, $p'_k = p_k - r(\alpha_l - \alpha_i)$ and $p'_l = p_l + r(\alpha_k - \alpha_i)$. We get

$$\sum_j p'_j - 1 = p'_i + p'_k + p'_l - (p_i + p_k + p_l) =$$

$$r(\alpha_l - \alpha_k) - r(\alpha_l - \alpha_i) + r(\alpha_k - \alpha_i) = 0.$$

Furthermore

$$\sum_j \alpha_j p'_j - C = \alpha_i p'_i + \alpha_k p'_k + \alpha_l p'_l - (\alpha_i p_i + \alpha_k p_k + \alpha_l p_l) =$$

$$r\alpha_i(\alpha_l - \alpha_k) - r\alpha_k(\alpha_l - \alpha_i) + r\alpha_l(\alpha_k - \alpha_i) = 0.$$

Choosing r positive or negative gives the different inequalities. Note that r must be chosen so that every p'_j is still positive. $\square$

Proposition 2.1. $-\infty < \sup_{\theta \in R} g(\theta) \leq 0$.

Proof. We begin by noting that $g \leq 0$ in R. The supremum is thus at most 0.

We will next show that there exists a point in R with all $p_j > 0$. We first solve $p_1 + p_N = 1$ and $\alpha_1 p_1 + \alpha_N p_N = C$. The solution is given by $p_1 = \frac{\alpha_N - C}{\alpha_N - \alpha_1} > 0$ and $p_N = \frac{C - \alpha_1}{\alpha_N - \alpha_1} > 0$. For any $1 < k < N$ we can now increase p_k by decreasing both p_1 and p_N as per Lemma 2.1.

This point θ_0 will be positive in all components, and we must then have $g(\theta_0) \leq \sup_{\theta \in R} g(\theta)$.

$\square$

Proposition 2.2. *There exists a global maximum of g in R.*

Proof. Let $\theta_n = (p_j^{(n)})$ be a sequence of points in R such that $g(\theta_n) \to M$.

For any j such that $n_j > 0$ we will eventually have $p_j^{(n)} > \delta_j$ for some $\delta_j > 0$. For otherwise we can choose a subsequence of θ_n such that $p_j^{(n)} < \frac{1}{n}$. Then $g(\theta_n) \leq n_j \log p_j^{(n)} < n_j \log \frac{1}{n}$ which tends to $-\infty$.

Let $\delta = \min \delta_j$. Then we will eventually have $\theta_n \in \prod I_j$ where $I_j = [\delta, 1]$ if $n_j > 0$ and otherwise $I_j = [0, 1]$. By compactness a subsequence of θ_n converges to θ in $\prod I_j$. Note that g is continuous on $\prod I_j$ by construction. We will thus have $g(\theta) = M$ which completes the proof. □

Now that we know a global maximum exists we aim to find it. We begin with the following lemma.

Proposition 2.3. *Let $\theta = (p_j)$ be a global maximum for g with $n_k = 0$ for some k. If $p_k > 0$ then $k = 1$ or $k = N$. Moreover if $n_1 = n_N = 0$, then one of p_1 and p_N must be 0.*

Proof. Suppose to the contrary that $p_k > 0$ with $1 < k < N$, and let l be such that $n_l > 0$. Assume without loss of generality that $k < l$. Then we can increase p_l and p_1 by decreasing p_k according to Lemma 2.1. Since g does not depend on p_k, but does depend on p_l this will increase the value of g.

For the second part, suppose once more for a contradiction that both p_1 and p_N are positive. Choose $1 < l < N$ such that $n_l > 0$. Now increase p_l by decreasing both p_1 and p_N. □

By Proposition 2.3 we can thus simply ignore any p_j with $1 < j < N$ that have $n_j = 0$, as they will all be 0 for our maximum. Going forward we will therefore assume that all n_j are nonzero except possibly for n_1 and n_N.

Example 2.1. Let $\alpha_1 = 1$, $\alpha_2 = 2$ and $\alpha_3 = 3$. And suppose we have the sample $n_1 = n_3 = 0$, $n_2 = 1$. We want to maximize $\log p_2$ subject to $p_1 + p_2 + p_3 = 1$ and $p_1 + 2p_2 + 3p_3 = C$ where $C > 2$. Subtracting the two equations gives $p_2 + 2p_3 = C - 1$ or $p_3 = \frac{C-1-p_2}{2}$. Substituting this in the first equation gives $p_1 + p_2 + \frac{C-1-p_2}{2} = 1$ or $p_1 = \frac{3-C-p_2}{2}$.

Recall that we need $0 \leq p_1, p_3$ which implies $0 \leq \frac{3-C-p_2}{2}$ and $0 \leq \frac{C-1-p_2}{2}$. This in turn gives $p_2 \leq C-1$ and $p_2 \leq 3-C$. Since $C > 2$ the second inequality is more restrictive, so our maximum is at $p_2 = 3 - C$. This gives $p_1 = 0$ and $p_3 = C - 2$. This is in line with what we expect according to Proposition 2.3. Our maximum value is moreover $1 \log p_2 = \log(3 - C)$.

2.2 Finding the Global Maximum

We now have everything we need to find the global maximum. We have reduced the problem to the following. All $n_j > 0$ except for n_1 and n_N. We begin with the case where all $n_j > 0$.

Lemma 2.2. *Suppose that all n_j are nonzero, then a global maximum of g is in the interior of the unit cube $I^N = [0,1]^N$.*

Proof. If it is not in the interior then $p_k = 1$ for some k. This can only happen if either $N = 1$ or $p_j = 0$ for all other j. If this is the case then $g(\theta) = -\infty$ since $n_j > 0$ for all j. This cannot be a global maximum of g by Proposition 2.1. $\square$

Lemma 2.2 lets us use Lagrange multipliers to find the maximum(s) of g, as we shall see in the following proposition.

Proposition 2.4. *Suppose that all n_j are nonzero, then*

$$h(\mu) = \sum_j \frac{(\alpha_j - C)n_j}{1 + (\alpha_j - C)\mu}$$

has a unique root μ_0 in the interval $\left(\frac{-1}{\alpha_N - C}, \frac{1}{C - \alpha_1}\right)$ and $\theta = (p_j)$ given by

$$p_j = \frac{n_j}{1 + (\alpha_j - C)\mu_0}$$

is the unique global maximum of g.

Proof. We proceed using Lagrange multipliers with

$$\mathcal{L}(\theta, \lambda, \mu) = \sum_j n_j \log p_j - \lambda \left(\sum_j p_j - 1 \right) - \mu \left(\sum_j \alpha_j p_j - C \right).$$

Differentiating with respect to p_j gives the equation

$$\frac{n_j}{p_j} - \lambda - \alpha_j \mu = 0.$$

Since $n_j > 0$ we know that $\lambda + \alpha_j \mu$ is nonzero, and we can thus solve for p_j, which gives

$$p_j = \frac{n_j}{\lambda + \alpha_j \mu}.$$

Substituting into the first two equations gives

$$\sum_j \frac{n_j}{\lambda + \alpha_j \mu} = 1$$

and

$$\sum_j \frac{\alpha_j n_j}{\lambda + \alpha_j \mu} = C.$$

It now follows that

$$\lambda + C\mu = \lambda \sum_j \frac{n_j}{\lambda + \alpha_j \mu} + \mu \sum_j \frac{\alpha_j n_j}{\lambda + \alpha_j \mu} = \sum_j n_j = 1.$$

Solving for λ and substituting gives

$$\sum_j \frac{\alpha_j n_j}{1+(\alpha_j - C)\mu} = C$$

and we finally deduce that

$$\sum_j \frac{(\alpha_j - C)n_j}{1+(\alpha_j - C)\mu} = \sum_j \frac{\alpha_j n_j}{1+(\alpha_j - C)\mu} - C\sum_j \frac{n_j}{1+(\alpha_j - C)\mu} = C - C = 0.$$

For any j and μ in the given interval we have $1 + (\alpha_j - C)\mu > 0$. For the case where $\mu \geq 0$ we get the inequalities $1 + (\alpha_j - C)\mu \geq 1 + (\alpha_1 - C)\mu > 0$. The case $\mu < 0$ is similar. This shows that h is smooth on the given interval.

The derivative is in fact given by

$$h'(\mu) = -\sum_j \frac{(\alpha_j - C)^2 n_j}{(1+(\alpha_j - C)\mu)^2} < 0$$

and h is thus strictly decreasing. It is not hard to see that $h \to \infty$ as $\mu \to \frac{-1}{\alpha_N - C}$. This is because $1 + (\alpha_N - C)\mu$ goes to 0 and for the numerator we have $(\alpha_N - C)n_N > 0$. Similarly we see that $h \to -\infty$ as $\mu \to \frac{1}{C-\alpha_1}$.

The intermediate value theorem implies that h has a root in the given interval. The fact that h is strictly decreasing implies that this root is unique. Moreover

$$p_j = \frac{n_j}{\lambda + \alpha_j \mu_0} = \frac{n_j}{1+(\alpha_j - C)\mu_0}$$

which completes the proof. □

The following example shows that the method of Proposition 2.4 fails if α_1 or α_N is 0.

Example 2.2. Consider the problem of maximizing $\frac{1}{2}\log p_2 + \frac{1}{2}\log p_3$ subject to $p_1 + p_2 + p_3 = 1$ and $p_1 + 2p_2 + 3p_3 = \frac{5}{2}$. We try to use Lagrange multipliers with

$$\mathcal{L} = \frac{1}{2}\log p_2 + \frac{1}{2}\log p_3 - \lambda(p_1 + p_2 + p_3 - 1) - \mu(p_1 + 2p_2 + 3p_3 - \frac{5}{2}).$$

Differentiating gives the equations $\lambda + \mu = 0$, $\frac{1}{2p_2} = \lambda + 2\mu = \mu$ and $\frac{1}{2p_3} = \lambda + 3\mu = 2\mu$. This gives $p_2 = \frac{1}{2\mu}$ and $p_3 = \frac{1}{4\mu}$. Furthermore we get

$$p_1 + \frac{1}{2\mu} + \frac{1}{4\mu} = 1$$

and

$$p_1 + \frac{2}{2\mu} + \frac{3}{4\mu} = \frac{5}{2}.$$

The solution to this system is $\mu = \frac{2}{3}$ and $p_1 = -\frac{1}{8} < 0$.

There is no obvious way to fix the problem that Lagrange multipliers fail if the maximum lies on the boundary. The following proposition instead gives a useful way of approximating the maximum.

Proposition 2.5. *Suppose that n_1 or n_N or both are 0 but all other n_j are nonzero. Let g_ε be a function of both $\varepsilon \in (0, \varepsilon_0)$ and θ of the form*

$$g_\varepsilon(\theta) = \sum_j n_j^{(\varepsilon)} \log p_j.$$

Suppose that $n_j^{(\varepsilon)} > 0$ and that $\lim_{\varepsilon\to 0} n_j^{(\varepsilon)} = n_j$. Let M be the maximum value of g. If θ_ε is the unique global maximum of g_ε, then $g_\varepsilon(\theta_\varepsilon) \to M$.

Proof. We consider only the case where $n_1 = 0$ here. Suppose that $g_\varepsilon(\theta_\varepsilon)$ does not approach M. There exists a subsequence such that $g_\varepsilon(\theta_\varepsilon) < M - \delta$ or $g_\varepsilon(\theta_\varepsilon) > M + \delta$ for some positive δ. We will refer to this subsequence in ε as the same sequence since it does not affect the argument.

For the first case let θ_0 be the maximum of g. If $p_1 > 0$ for θ_0, then $g_\varepsilon(\theta_0)$ tends to $g(\theta_0)$. However $g_\varepsilon(\theta_0) \le g_\varepsilon(\theta_\varepsilon) < M - \delta$ which is of course impossible. If $p_1 = 0$ we choose $p_1^{(\varepsilon)} = n_1^{(\varepsilon)}$, $p_k^{(\varepsilon)} = p_k - (1+r)n_1^{(\varepsilon)}$ and $p_l^{(\varepsilon)} = p_l + rn_1^{(\varepsilon)}$ for suitable $r > 0$ and $1 < k < l$ as in Lemma 2.1.

Now for $\tilde\theta_\varepsilon = (p_j^{(\varepsilon)})$ we get

$$|g_\varepsilon(\tilde\theta_\varepsilon) - g(\tilde\theta_\varepsilon)| \le \sum_j |(n_j^{(\varepsilon)} - n_j)\log p_j^{(\varepsilon)}| =$$

$$|n_1^{(\varepsilon)}||\log p_1^{(\varepsilon)}| + \sum_{j\ne 1} |n_j^{(\varepsilon)} - n_j||\log p_j^{(\varepsilon)}| =$$

$$-n_1^{(\varepsilon)} \log n_1^{(\varepsilon)} + \sum_{j\ne 1} |n_j^{(\varepsilon)} - n_j||\log p_j^{(\varepsilon)}|.$$

The term $-n_1^{(\varepsilon)} \log n_1^{(\varepsilon)}$ goes to 0 because $n_1^{(\varepsilon)}$ does and the sum because $p_j^{(\varepsilon)}$ goes to $p_j > 0$ for $j \ne 1$. Note that by continuity of g, $g(\tilde\theta_\varepsilon) \to g(\theta_0)$. This in turn implies that $g_\varepsilon(\tilde\theta_\varepsilon) \to g(\theta_0) = M$. However we also have $g_\varepsilon(\tilde\theta_\varepsilon) \le g_\varepsilon(\theta_\varepsilon) < M - \delta$ a contradiction.

For the second case consider $g_\varepsilon(\theta_\varepsilon) - g(\theta_\varepsilon) > M + \delta - M = \delta$. We will show once more that $g_\varepsilon(\theta_\varepsilon) - g(\theta_\varepsilon)$ tends to 0. Let μ_ε be the root of h_ε from Proposition 2.4. We know that

$$p_j^{(\varepsilon)} = \frac{n_j^{(\varepsilon)}}{1 + (\alpha_j - C)\mu_\varepsilon}.$$

Note that $p_j^{(\varepsilon)}$ is bounded below by

$$p_j^{(\varepsilon)} = \frac{n_j^{(\varepsilon)}}{1 + (\alpha_j - C)\mu_\varepsilon} > \frac{n_j^{(\varepsilon)}}{M}$$

for some M. It now once more follows that

$$|g_\varepsilon(\theta_\varepsilon) - g(\theta_\varepsilon)| \leq \sum_j |(n_j^{(\varepsilon)} - n_j)\log p_j^{(\varepsilon)}| =$$

$$|n_1^{(\varepsilon)}||\log p_1^{(\varepsilon)}| + \sum_{j\neq 1} |n_j^{(\varepsilon)} - n_j||\log p_j^{(\varepsilon)}| =$$

$$-n_1^{(\varepsilon)} \log p_1^{(\varepsilon)} + \sum_{j\neq 1} |n_j^{(\varepsilon)} - n_j||\log p_j^{(\varepsilon)}| <$$

$$-n_1^{(\varepsilon)} \log \frac{n_1^{(\varepsilon)}}{M} + \sum_{j\neq 1} |n_j^{(\varepsilon)} - n_j||\log p_j^{(\varepsilon)}| \to 0.$$

□

Remark 2.1. For the proof of Proposition 2.5 to work in the case where both n_1 and n_N are zero we need that either $n_1^{(\varepsilon)} \log n_N^{(\varepsilon)}$ or $n_N^{(\varepsilon)} \log n_1^{(\varepsilon)}$ goes to 0. Note that we need to choose the $n_k^{(\varepsilon)}$ that is the largest. And then $n_l^{(\varepsilon)} \log n_k^{(\varepsilon)}$ goes to 0. Looking at

$$\frac{n_1^{(\varepsilon)}}{n_N^{(\varepsilon)}}$$

we see that it is either bounded or we can pass to another subsequence where $n_1^{(\varepsilon)} > n_N^{(\varepsilon)}$, and we can complete the proof.

3 Generalized Hypotheses

Most of the time we are not even interested in what the exact Elo difference between two players are. Suppose for example that we make an adjustment to a chess program. We do not care if the engine improved by 10 or 100 Elo. If it improves the chess program, the change is good.

For example the null hypothesis will be that the program has not improved. That is

$$H_0 : \text{The Elo of player 1 does not exceed that of player 2.}$$

The alternative hypothesis is that the chess program has in fact improved. Suppose that we take the complement as our alternative hypothesis.

$$H_1 : \text{The Elo of player 1 exceeds that of player 2.}$$

In both of these cases we want to examine hypotheses $H : \phi(\theta) \leq C$ or $H : \phi(\theta) > C$. In fact, we generalize even more and calculate any hypothesis of the form $H : \phi(\theta) \in \Phi$ where $\Phi \subset (\alpha_1, \alpha_N)$ is connected. If Φ is not connected we can still calculate the supremum by two cases as stated in Corollary 3.4. The supremum we want to calculate is then

$$\sup_{\phi(\theta)\in\Phi} \sum_j n_j \log p_j .$$

We now let $n_j > 0$ be fixed. Let $m : (\alpha_1, \alpha_N) \to (-\infty, 0]$ be the function, of the variable $x = C$, given by

$$m(x) = \sup_{\phi(\theta)=x} \sum_j n_j \log p_j .$$

Lemma 3.1. *m has a global maximum at $x_0 = \sum_j \alpha_j n_j$.*

Proof. We want to calculate $\sup \sum_j n_j \log p_j$. We once more proceed using Lagrange multipliers but without any restriction on $\phi(\theta)$. We thus get

$$\mathcal{L} = \sum_j n_j \log p_j - \lambda \left(\sum_j p_j - 1 \right) .$$

We now have

$$\frac{n_j}{p_j} = \lambda$$

or equivalently

$$p_j = \frac{n_j}{\lambda} .$$

This implies

$$\sum_j \frac{n_j}{\lambda} = 1$$

and we get $\lambda = 1$. Now $p_j = n_j$ for all j and the maximum of m must be at $\phi(\theta) = \sum_j \alpha_j p_j = \sum_j \alpha_j n_j = x_0$. □

The following is the main result of this section.

Proposition 3.1. *m is C^2 and has a global maximum at $x_0 = \sum_j \alpha_j n_j$. Moreover m is strictly increasing on the interval $(\alpha_1, x_0]$ and strictly decreasing on $[x_0, \alpha_N)$.*

Proof. Recall the function

$$h(\mu, x) = \sum_j \frac{(\alpha_j - x) n_j}{1 + (\alpha_j - x)\mu} .$$

As stated in Proposition 2.4, we can solve $h(\mu, x) = 0$ for μ as a function of x. We also proved that

$$\frac{\partial h}{\partial \mu} < 0$$

for all μ and x in their respective intervals. The Implicit Function Theorem [2] now implies that the function $\mu(x)$ is at least C^1.

We now turn to the derivative of m. Note that the variables p_j depend on x as

$$p_j(x) = \frac{n_j}{1 + (\alpha_j - x)\mu(x)}.$$

It follows that m is C^1 since μ is. The derivative is given by

$$m'(x) = \sum_j \frac{n_j}{p_j} p_j'(x) = \sum_j \frac{n_j}{p_j} \frac{n_j}{(1 + (\alpha_j - x)\mu(x))^2}(\mu(x) - (\alpha_j - x)\mu'(x)) =$$

$$\sum_j p_j(x)(\mu(x) - (\alpha_j - x)\mu'(x)) = \mu(x) \sum_j p_j(x) - \mu'(x) \sum_j (\alpha_j - x) p_j(x) =$$

$$\mu(x)1 - \mu'(x)h(\mu(x), x) = \mu(x).$$

We now see that m is C^2 since μ is C^1.

Suppose now that $m'(x) = 0$, then $\mu(x) = 0$ and

$$p_j = \frac{n_j}{1 + (\alpha_j - x)0} = n_j.$$

We thus have $x = \sum_j \alpha_j p_j = \sum_j \alpha_j n_j = x_0$. Together with Lemma 3.1 we see that x_0 is the only root of m'.

If x tends to α_1 then the condition $\sum_j \alpha_j p_j = x$ ensures that the maximum θ tends to $(1, 0, ..., 0)$. $m(x)$ must then tend to $-\infty$ for all n_j are positive. A similar argument shows that m tends to $-\infty$ as $x \to \alpha_N$.

By the continuity of the derivative m' and the single root x_0, we know that m' cannot change sign on (α_1, x_0) nor (x_0, α_N). Since m tends to $-\infty$ on both endpoints we conclude that m is strictly increasing on $(\alpha_1, x_0]$, and strictly decreasing on $[x_0, \alpha_N)$. □

Corollary 3.1. *Let $\Phi \subseteq (\alpha_1, \alpha_N)$ be a connected nonempty set. Then*

$$\sup_{\phi(\theta) \in \Phi} \sum_j n_j \log p_j = \sup_{\phi(\theta) = C} \sum_j n_j \log p_j$$

where C is the unique point in $\overline{\Phi}$ closest to x_0.

Corollary 3.2. *Let $\Phi \subseteq (\alpha_1, x_0]$ be a nonempty set. Then*

$$\sup_{\phi(\theta) \in \Phi} \sum_j n_j \log p_j = \sup_{\phi(\theta) = C} \sum_j n_j \log p_j$$

where C is the greatest point in $\overline{\Phi}$.

Corollary 3.3. *Let $\Phi \subseteq [x_0, \alpha_N)$ be a nonempty set. Then*

$$\sup_{\phi(\theta) \in \Phi} \sum_j n_j \log p_j = \sup_{\phi(\theta) = C} \sum_j n_j \log p_j$$

where C is the smallest point in $\overline{\Phi}$.

Corollary 3.4. *Let $\Phi \subseteq (\alpha_1, \alpha_N)$ be a nonempty set. If Φ intersects both $(\alpha_1, x_0]$ and $[x_0, \alpha_N)$ then*

$$\sup_{\phi(\theta)\in\Phi} \sum_j n_j \log p_j = \max\left\{ \sup_{\phi(\theta)=C_1} \sum_j n_j \log p_j, \sup_{\phi(\theta)=C_2} \sum_j n_j \log p_j \right\}$$

where C_1 is the greatest point in $\overline{\Phi} \cap (\alpha_1, x_0]$ and C_2 the smallest in $\overline{\Phi} \cap [x_0, \alpha_N)$.

References

[1] Wikipedia contributors. Elo rating system — Wikipedia, the free encyclopedia, 2023. `https://en.wikipedia.org/w/index.php?title=Elo_rating_system&oldid=1185663969`.

[2] Wikipedia contributors. Implicit function theorem — Wikipedia, the free encyclopedia, 2023. `https://en.wikipedia.org/w/index.php?title=Implicit_function_theorem&oldid=1181247577`.