

# COMPARING THE APPROXIMATIONS FOR THE GENERALIZED LOG LIKELIHOOD RATIO OF A MULTINOMIAL DISTRIBUTION

MICHEL VAN DEN BERGH

## 1. INTRODUCTION AND STATEMENT OF THE RESULTS

We recall some results from [2]. Assume given real numbers

$$a_1 < a_2 < \cdots < a_N$$

and a probability distribution

$$P : \{a_1, \dots, a_N\} \rightarrow \mathbb{R} : a_i \mapsto p_i$$

Assume a sample taken from  $\{a_1, \dots, a_N\}$  according to  $P$  has sample distribution  $(\hat{p}_i)_{i=1, \dots, N}$ . We want to compute the corresponding MLE for the true distribution  $(p_i)_{i=1, \dots, N}$ , subject to the condition that the latter's expectation value is  $s$ . I.e.  $\sum_i p_i a_i = s$ .

For simplicity we will assume

$$(1.1) \quad a_1 < s < a_N, \forall i : \hat{p}_i \neq 0$$

**Proposition 1.1.** *The ML distribution is unique. It is given by*

$$(1.2) \quad p_i = \frac{\hat{p}_i}{1 + \theta(a_i - s)}$$

where  $\theta$  is the unique root of the equation

$$(1.3) \quad \sum_i \frac{\hat{p}_i(a_i - s)}{1 + \theta(a_i - s)} = 0$$

in the interval  $[-1/(a_N - s), 1/(s - a_1)]$ .

Let  $\mu = \sum_i p_i a_i$  and let  $\text{LLR}_{\text{exact}}$  be the generalized log-likelihood ratio [4] for  $\mu = s_0$  versus  $\mu = s_1$ , divided by the sample size. If  $(\theta_i)_{i=1,2}$  are the solutions to (2.2) for  $s = \mu_i$  then by (1.2) we have

$$\text{LLR}_{\text{exact}} = \sum_i \hat{p}_i \log \frac{1 + \theta_0(a_i - s_0)}{1 + \theta_1(a_i - s_1)}$$

Computing  $\text{LLR}_{\text{exact}}$  requires numerically solving the rational equation (1.2) *twice*. This is trivial to do numerically and indeed this is how it is done in Fishtest [5]. Nonetheless to fortify our intuition it is useful to have more manageable approximations to  $\text{LLR}_{\text{exact}}$ . One such approximation was given in [2, Proposition 2.1].

$$\text{LLR}_{\text{alt}} = \frac{1}{2} \log \left( \frac{\sum_i \hat{p}_i (s_0 - a_i)^2}{\sum_i \hat{p}_i (s_1 - a_i)^2} \right)$$

Another even simpler approximation was given in [1, (2.1)].

$$\text{LLR}_{\text{alt2}} = \frac{1}{2} \frac{(s_1 - s_0)(2\hat{\mu} - s_0 - s_1)}{\hat{\sigma}^2}$$

where  $\hat{\mu}$  and  $\hat{\sigma}^2$  are respectively the sample mean and variance. In other words

$$\hat{\mu} = \sum_i \hat{p}_i a_i, \quad \hat{\sigma}^2 = \sum_i \hat{p}_i (a_i - \hat{\mu})^2$$

*Remark 1.2.*  $\text{LLR}_{\text{alt}}$  and  $\text{LLR}_{\text{alt2}}$  also make sense for continuous distributions. In contrast  $\text{LLR}_{\text{exact}}$  only seems to make sense for continuous distributions with bounded support.

Put  $\Delta := (s_1 - s_0)/\hat{\sigma}$ . It can be seen from the data in Example 1.3 below that  $\Delta$  is typically quite small. Moreover it follows from [3] that the expected duration  $d$  of an SPRT test with reasonable resolution is  $\sim 1/\Delta^2$ . Or  $\Delta \sim 1/\sqrt{d}$ .

In this note we relate  $\text{LLR}_{\text{exact}}$ ,  $\text{LLR}_{\text{alt}}$  and  $\text{LLR}_{\text{alt2}}$  by providing a power series expression in  $\Delta$  for them, up to  $\Delta^4$ . In other words we are developing  $\text{LLR}_{\text{exact}}$ ,  $\text{LLR}_{\text{alt}}$  and  $\text{LLR}_{\text{alt2}}$  in terms of  $1/\sqrt{d}$ .

Let  $\nu$ ,  $\kappa$  be respectively the skewness and the excess kurtosis for the sample distribution. Thus

$$\nu = \frac{\hat{\mu}_3}{\hat{\sigma}^3} \quad \kappa = \frac{\hat{\mu}_4}{\hat{\sigma}^4} - 3$$

where  $\hat{\mu}_3$ ,  $\hat{\mu}_4$  are the third and fourth central sample moments. In other words

$$\hat{\mu}_3 = \sum_i \hat{p}_i (a_i - \hat{\mu})^3 \quad \hat{\mu}_4 = \sum_i \hat{p}_i (a_i - \hat{\mu})^4$$

Finally let  $h$  be the relative position of  $\hat{\mu}$  with respect to the interval  $[s_0, s_1]$ . Formally

$$h = \frac{2\hat{\mu} - s_0 - s_1}{s_1 - s_0}$$

so that  $h = 1$  corresponds to  $\hat{\mu} = s_1$  and  $h = -1$  corresponds to  $\hat{\mu} = s_0$ .

With these definitions we have the following formulas

$$(1.4) \quad \text{LLR}_{\text{alt2}} = \frac{1}{2} h \Delta^2$$

$$(1.5) \quad \text{LLR}_{\text{alt}} = \frac{1}{2} h \Delta^2 - \frac{1}{8} (h^3 + h) \Delta^4 + \dots$$

$$(1.6) \quad \text{LLR}_{\text{exact}} = \frac{1}{2} h \Delta^2 + \frac{1}{12} \nu (3h^2 + 1) \Delta^3 + \frac{1}{8} (2\nu^2 - \kappa - 1) (h^3 + h) \Delta^4 + \dots$$

Or

$$(1.7) \quad \text{LLR}_{\text{alt}} - \text{LLR}_{\text{alt2}} \cong -\frac{1}{8} (h^3 + h) \Delta^4$$

$$(1.8) \quad \text{LLR}_{\text{exact}} - \text{LLR}_{\text{alt}} \cong \frac{1}{12} \nu (3h^2 + 1) \Delta^3 + \frac{1}{8} (2\nu^2 - \kappa) (h^3 + h) \Delta^4$$

**Example 1.3.** We illustrate the above formulas with some pentanomial data from Fishtest [5]. We have  $s_i = 1/(1 + 10^{-e_i/400})$  where  $e_0, e_1$  are the Elo bounds, given in the first two columns. The column labeled “approx” gives the approximation to  $\text{LLR}_{\text{exact}}$  obtained via the formula (1.8).

$e_0$	$e_1$	pent. freqs	size	$\Delta$	$h$	$\nu$	$\kappa$	exact	alt	rel. err.	approx	rel. err.	err. ratio
-0.50	1.50	[729, 5234, 10174, 5154, 898]	22189	0.0132	1.520	0.0463	-0.1717	2.93696	2.93533	-5.6e-04	2.93696	2.8e-07	-1990.0
-0.50	1.50	[441, 3213, 7170, 3164, 399]	14387	0.0140	-2.106	-0.0144	0.0065	-2.96711	-2.96644	-2.3e-04	-2.96711	3.0e-07	-762.5
-0.50	1.50	[1746, 11875, 23034, 11679, 1820]	50154	0.0133	-0.666	0.0154	-0.1731	-2.93641	-2.93673	1.1e-04	-2.93641	-1.7e-08	-6266.1
-0.50	1.50	[550, 3237, 6268, 3224, 478]	13757	0.0131	-2.483	-0.0245	-0.1821	-2.94598	-2.94457	-4.8e-04	-2.94598	5.4e-07	-888.4
0.25	1.75	[504, 6686, 20525, 6609, 557]	34881	0.0122	-1.141	0.0244	0.4324	-2.94903	-2.94977	2.5e-04	-2.94903	-5.6e-08	-4487.4
-1.50	0.50	[271, 2151, 4827, 2241, 299]	9789	0.0139	3.091	0.0028	-0.0289	2.94174	2.94151	-7.7e-05	2.94174	5.6e-07	-137.5

## 2. DERIVATIONS

2.1. **The expression for  $\text{LLR}_{\text{alt2}}$ .** This is obvious.

2.2. **The expression for  $\text{LLR}_{\text{exact}}$ .** We will use a formula which was derived in the proof of [2, Proposition 2.1]

$$(2.1) \quad \text{LLR} = \int_{s_0}^{s_1} \theta(s) ds$$

where  $\theta = \theta(s)$  is the root of

$$(2.2) \quad \sum_i \frac{\hat{p}_i(a_i - s)}{1 + \theta(a_i - s)} = 0$$

in the interval  $[-1/(a_N - s), 1/(s - a_1)]$ . It will be convenient to write

$$\hat{o}_n(s) = \sum_i \hat{p}_i(a_i - s)^n$$

Note

$$\begin{aligned} \hat{o}_1(s) &= \hat{\mu} - s \\ \hat{o}_2(s) &= \hat{\sigma}^2 + (\hat{\mu} - s)^2 \\ \hat{o}_3(s) &= \hat{\mu}_3 + 3\hat{\sigma}^2(\hat{\mu} - s) + (\hat{\mu} - s)^3 \\ \hat{o}_4(s) &= \hat{\mu}_4 + 4\hat{\mu}_3(\hat{\mu} - s) + 6\hat{\sigma}^2(\hat{\mu} - s)^2 + (\hat{\mu} - s)^4 \end{aligned}$$

We obtain

$$\hat{o}_1(s) - \theta \hat{o}_2(s) + \theta^2 \hat{o}_3(s) - \theta^3 \hat{o}_4(s) + \dots = 0$$

Or

$$\theta = \frac{\hat{o}_1(s)}{\hat{o}_2(s)} + \theta^2 \frac{\hat{o}_3(s)}{\hat{o}_2(s)} - \theta^3 \frac{\hat{o}_4(s)}{\hat{o}_2(s)} + \dots$$

This equation can be solved by repeated self substitution. First step:

$$\theta \cong \frac{\hat{o}_1(s)}{\hat{o}_2(s)}$$

Second step:

$$\theta \cong \frac{\hat{o}_1(s)}{\hat{o}_2(s)} + \frac{\hat{o}_3(s)\hat{o}_1(s)^2}{\hat{o}_2(s)^3} - \frac{\hat{o}_4(s)\hat{o}_1(s)^3}{\hat{o}_2(s)^4}$$

Third step (truncating at  $\hat{o}_1(s)^3$ ):

$$\begin{aligned}
\theta &\cong \frac{\hat{o}_1(s)}{\hat{o}_2(s)} + \frac{\hat{o}_3(s)\hat{o}_1(s)^2}{\hat{o}_2(s)^3} - \frac{\hat{o}_4(s)\hat{o}_1(s)^3}{\hat{o}_2(s)^4} + 2\frac{\hat{o}_3(s)^2\hat{o}_1(s)^3}{\hat{o}_2(s)^5} \\
&\cong \frac{\hat{\mu} - s}{\hat{\sigma}^2 + (\hat{\mu} - s)^2} + \frac{(\hat{\mu}_3 + 3\hat{\sigma}^2(\hat{\mu} - s))(\hat{\mu} - s)^2}{\hat{\sigma}^6} - \frac{\hat{\mu}_4(\hat{\mu} - s)^3}{\hat{\sigma}^8} + 2\frac{\hat{\mu}_3^2(\hat{\mu} - s)^3}{\hat{\sigma}^{10}} \\
&\cong \frac{\hat{\mu} - s}{\hat{\sigma}^2} - \frac{(\hat{\mu} - s)^3}{\hat{\sigma}^4} + \frac{(\hat{\mu}_3 + 3\hat{\sigma}^2(\hat{\mu} - s))(\hat{\mu} - s)^2}{\hat{\sigma}^6} - \frac{\hat{\mu}_4(\hat{\mu} - s)^3}{\hat{\sigma}^8} + 2\frac{\hat{\mu}_3^2(\hat{\mu} - s)^3}{\hat{\sigma}^{10}} \\
&= \frac{1}{\hat{\sigma}^2}(\hat{\mu} - s) + \frac{\hat{\mu}_3}{\hat{\sigma}^6}(\hat{\mu} - s)^2 + \left( \frac{2}{\hat{\sigma}^4} - \frac{\hat{\mu}_4}{\hat{\sigma}^8} + \frac{2\hat{\mu}_3^2}{\hat{\sigma}^{10}} \right) (\hat{\mu} - s)^3
\end{aligned}$$

The integral is

$$\int_{s_0}^{s_1} \theta(s) ds = -\frac{1}{2\hat{\sigma}^2}(\hat{\mu} - s)^2 - \frac{\hat{\mu}_3}{3\hat{\sigma}^6}(\hat{\mu} - s)^3 - \frac{1}{4} \left( \frac{2}{\hat{\sigma}^4} - \frac{\hat{\mu}_4}{\hat{\sigma}^8} + \frac{2\hat{\mu}_3^2}{\hat{\sigma}^{10}} \right) (\hat{\mu} - s)^4 \Big|_{s_0}^{s_1}$$

Put

$$\begin{aligned}
\delta &= s_1 - s_0 \\
m &= (s_1 + s_0)/2
\end{aligned}$$

Then

$$\hat{\mu} = m + h\delta/2$$

We have

$$\begin{aligned}
\hat{\mu} - s_0 &= \delta/2 + h\delta/2 = \frac{1}{2}\delta(h+1) \\
\hat{\mu} - s_1 &= -\delta/2 + h\delta/2 = \frac{1}{2}\delta(h-1)
\end{aligned}$$

Hence

$$\begin{aligned}
(\hat{\mu} - s_0)^2 - (\hat{\mu} - s_1)^2 &= \delta^2 h \\
(\hat{\mu} - s_0)^3 - (\hat{\mu} - s_1)^3 &= \frac{1}{4}\delta^3(3h^2 + 1) \\
(\hat{\mu} - s_0)^4 - (\hat{\mu} - s_1)^4 &= \frac{1}{2}\delta^4(h^3 + h)
\end{aligned} \tag{2.3}$$

Substituting yields

$$\int_{s_0}^{s_1} \theta(s) ds = \frac{h\delta^2}{2\hat{\sigma}^2} + \frac{\hat{\mu}_3\delta^3}{12\hat{\sigma}^6}(3h^2 + 1) + \frac{1}{8} \left( \frac{2}{\hat{\sigma}^4} - \frac{\hat{\mu}_4}{\hat{\sigma}^8} + \frac{2\hat{\mu}_3^2}{\hat{\sigma}^{10}} \right) \delta^4(h^3 + h)$$

which is (1.6).

2.3. **The expression for  $\text{LLR}_{\text{alt1}}$ .** We have

$$\begin{aligned}
\frac{1}{2} \log \left( \frac{\sum_i \hat{p}_i (s_0 - a_i)^2}{\sum_i \hat{p}_i (s_1 - a_i)^2} \right) &= \frac{1}{2} \log \left( \frac{\hat{\sigma}^2 + (\hat{\mu} - s_0)^2}{\hat{\sigma}^2 + (\hat{\mu} - s_1)^2} \right) \\
&= \frac{1}{2} \log \left( \frac{1 + \frac{(\hat{\mu} - s_0)^2}{\sigma^2}}{1 + \frac{(\hat{\mu} - s_1)^2}{\hat{\sigma}^2}} \right) \\
&\cong \frac{1}{2} \left( \frac{(\hat{\mu} - s_0)^2}{\hat{\sigma}^2} - \frac{1}{2} \frac{(\hat{\mu} - s_0)^4}{\hat{\sigma}^4} - \frac{(\hat{\mu} - s_1)^2}{\hat{\sigma}^2} + \frac{1}{2} \frac{(\hat{\mu} - s_1)^4}{\hat{\sigma}^4} \right) \\
&= \frac{h\delta^2}{2\hat{\sigma}^2} - \frac{1}{8} \frac{\delta^4}{\hat{\sigma}^4} (h^3 + h) \quad (\text{using (2.3)})
\end{aligned}$$

This is (1.5).

#### REFERENCES

1. Michel Van den Bergh, *A practical introduction to the GSPRT*, [http://hardy.uhasselt.be/Fishtest/GSPRT\\_approximation.pdf](http://hardy.uhasselt.be/Fishtest/GSPRT_approximation.pdf).
2. ———, *The generalized maximum likelihood ratio for the expectation value of a distribution*, [http://hardy.uhasselt.be/Fishtest/support\\_MLE\\_multinomial.pdf](http://hardy.uhasselt.be/Fishtest/support_MLE_multinomial.pdf).
3. ———, *The SPRT for Brownian motion*, <http://hardy.uhasselt.be/Fishtest/sprta.pdf>.
4. Xiaou Li, Jingchen Liu, and Zhiliang Ying, *Generalized Sequential Probability Ratio Test for Separate Families of Hypotheses*, [http://stat.columbia.edu/~jcliu/paper/GSPRT\\_SQA3.pdf](http://stat.columbia.edu/~jcliu/paper/GSPRT_SQA3.pdf).
5. Gary Linscott (original author), *The Fishtest framework*, <https://github.com/glinscott/fishtest>.