1. Exact results

Assume given real numbers
\[ a_1 < a_2 < \cdots < a_N \]
and a probability distribution
\[ P : \{a_1, \ldots, a_N\} \to \mathbb{R} : a_i \mapsto p_i \]
Assume a sample taken from \{a_1, \ldots, a_N\} according to \( P \) has sample distribution \((\hat{p}_i)_{i=1,\ldots,N}\). We want to compute the corresponding MLE for the true distribution \((p_i)_{i=1,\ldots,N}\), 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
\[ a_1 < s < a_N, \forall i : \hat{p}_i \neq 0 \]

**Proposition 1.1.** The ML distribution is unique. It is given by
\[ p_i = \frac{\hat{p}_i}{1 + \theta(a_i - s)} \]
where \( \theta \) is the unique root of the equation
\[ \sum_i \frac{\hat{p}_i(a_i - s)}{1 + \theta(a_i - s)} = 0 \]
in the interval \([-1/(a_N - s), 1/(s - a_1)]\).

**Proof.** We have to maximize the objective function
\[ \text{LLR}(p_i) = \sum_i \hat{p}_i \log p_i \]
subject to the constraints
\[ \sum_i p_i = 1 \]
\[ \sum_i a_i p_i = s \]
\[ p_i > 0 \]

The objective function is continuous on (1.4) and approaches \(-\infty\) on the boundary. So it has at least one maximum. To prove that it has a unique maximum it suffices to prove that there is a unique extremal value.
Using Lagrange multipliers we have to determine the extremal values of
\[ \sum_i \hat{p}_i \log p_i - \lambda \left( \sum_i p_i - 1 \right) - \theta \left( \sum_i p_i a_i - s \right) \]

We obtain
\[ (\lambda + \theta a_i) p_i = \hat{p}_i \]
and hence by (1.1) \( \lambda + \theta a_i \neq 0 \) so that

(1.5) \[ p_i = \frac{\hat{p}_i}{\lambda + \theta a_i} \]

where \( \lambda, \theta \) must satisfy
(1.6) \[ \sum_i \frac{\hat{p}_i}{\lambda + \theta a_i} = 1 \]
(1.7) \[ \sum_i \frac{\hat{p}_i a_i}{\lambda + \theta a_i} = s \]

Evaluating \( \lambda(1.6) + \theta(1.7) \) we find
\[ \lambda + \theta s = 1 \]
and hence \( \lambda = 1 - \theta s \) and we immediately obtain (1.2) from (1.5).

If is clear that (1.6)(1.7) imply (1.3). Assume (1.3) holds. Then
\[ 1 = \sum_i \hat{p}_i \]
\[ = \sum_i \frac{\hat{p}_i (1 + (a_i - s)\theta)}{1 + (a_i - s)\theta} \]
\[ = \sum_i \frac{\hat{p}_i}{1 + (a_i - s)\theta} + \theta \sum_i \frac{\hat{p}_i (a_i - s)}{1 + (a_i - s)\theta} \]
so that (1.6) holds. On the other hand we also have
\[ \sum_i \frac{\hat{p}_i (1 + (a_i - s)\theta)}{1 + (a_i - s)\theta} = (1 - s\theta) \sum_i \frac{\hat{p}_i}{1 + (a_i - s)\theta} + \theta \sum_i \frac{\hat{p}_i a_i}{1 + (a_i - s)\theta} \]
We conclude that (1.7) holds, unless perhaps if \( \theta = 0 \). If \( \theta = 0 \) then \( \lambda = 1 \) and (1.7) is equivalent to
\[ \hat{\mu} := \sum_i \hat{p}_i a_i = s \]
which also follows from (1.3).

Hence we have to solve (1.3) for \( \theta \). Moreover the fact that \( p \geq 0 \) leads to the additional constraint
\[ \hat{p}_i > 0 \Rightarrow 1 + \theta (a_i - s) > 0 \]
So we should have
\[ \theta > -\frac{1}{a_i - s} \quad \text{if} \ s < a_i \text{ and } \hat{p}_i > 0 \]
\[ \theta < \frac{1}{s - a_i} \quad \text{if} \ s > a_i \text{ and } \hat{p}_i > 0 \]
By (1.1) this is equivalent to
\[ \theta \in \left[ -\frac{1}{a_N - s}, \frac{1}{s - a_1} \right] \]

One verifies that on this interval the left hand side of (1.3) is strictly descending
and goes from \(+\infty\) to \(-\infty\). Hence (1.3) has a unique solution. □

**Remark 1.2.** It is easy to see that Proposition 1.1 is still true under the weaker
hypothesis \( \hat{p}_1 > 0, \hat{p}_N > 0 \). Moreover if there are \( i, j \) such that \( a_i < s < a_j \)
and \( \hat{p}_i \neq 0, \hat{p}_j \neq 0 \) then a suitable analogue of Proposition 1.1 still holds (\( \theta \) must be
in the interval between the poles of (1.3) which contains zero). If such \( i, j \) do not
exist then the description of the ML distribution is different. When (1.1) does not
hold it is easier in practice to deform \( \hat{p}_i \) a little bit so that it becomes true. One
may think of this as introducing a very weak prior.

**Remark 1.3.** (1.3) can be trivially solved numerically. For example using Newton’s
method.

## 2. Approximate results

**Proposition 2.1.** Let \( LLR \) be the generalized log-likelihood ratio for \( \mu = \mu_0 \) versus
\( \mu = \mu_1 \), divided by the sample size. Then we have

\[ LLR \simeq \frac{1}{2} \log \left( \frac{\sum \hat{p}_i (\mu_0 - a_i)^2}{\sum \hat{p}_i (\mu_1 - a_i)^2} \right) \]

**Proof.** Let \( \theta = \theta(s) \) be the solution to (1.3). By (1.2) the corresponding maximal
log-likelihood value (divided by the sample size) is given by

\[ LL(s) := -\sum \hat{p}_i \log(1 + \theta(s)(a_i - s)) \]

Developing the left hand side of (1.3) in a Taylor series in \( \theta \) and keeping only the
first order term we get

\[ \sum \hat{p}_i (a_i - s) - \theta \sum \hat{p}_i (a_i - s)^2 = \hat{\mu} - s - \theta \sum \hat{p}_i (a_i - s)^2 \]

so that we get

\[ \theta(s) \simeq \frac{\hat{\mu} - s}{\sum \hat{p}_i (a_i - s)^2} \]

This is the only approximation we make in the proof. From (2.2) we obtain

\[ LLR = -\sum \hat{p}_i \int_{\mu_0}^{\mu_1} \frac{d}{ds} \log(1 + \theta(s)(a_i - s)) \, ds \]

\[ = -\sum \hat{p}_i \int_{\mu_0}^{\mu_1} \theta'(s)(a_i - s) - \theta(s) \frac{1}{1 + \theta(s)(a_i - s)} \, ds \]

\[ = \int_{\mu_0}^{\mu_1} \theta(s) \, ds \]

\[ \simeq \int_{\mu_0}^{\mu_1} \frac{\hat{\mu} - s}{\sum \hat{p}_i (a_i - s)^2} \, ds \quad \text{by (1.3)(1.6)} \]

\[ \simeq \int_{\mu_0}^{\mu_1} \frac{\hat{\mu} - s}{\sum \hat{p}_i (a_i - s)^2} \, ds \quad \text{by (2.3)} \]
On the other hand
\[
\frac{d}{ds} \log \left( \sum_i \hat{p}_i(s - a_i)^2 \right) = 2 \frac{\sum \hat{p}_i(s - a_i)}{\sum \hat{p}_i(s - a_i)^2} \\
= 2 \frac{s - \mu}{\sum \hat{p}_i(s - a_i)^2}
\]
so that we find
\[
\begin{align*}
\text{LLR} & \equiv -\frac{1}{2} \int_{\mu_0}^{\mu_1} \frac{d}{ds} \log \left( \sum_i \hat{p}_i(s - a_i)^2 \right) ds \\
& = \frac{1}{2} \log \left( \sum_i \hat{p}_i(\mu_0 - a_i)^2 \right) - \frac{1}{2} \log \left( \sum_i \hat{p}_i(\mu_1 - a_i)^2 \right)
\end{align*}
\]
finishing the proof. \(\Box\)

Remark 2.2. Experiments show that the approximation (2.1) is very accurate.