The Cross-entropy of Piecewise Linear Probability Density Functions

Anonymous authors
Paper under double-blind review

Abstract

The cross-entropy and its related terms from information theory (e.g. entropy, Kullback-Leibler divergence) are found throughout artificial intelligence/machine learning. This includes many of the major successes, both current and historic, where they commonly appear as the natural objective of an optimisation procedure for learning model parameters, or their distributions. This paper presents a derivation of the differential cross-entropy between two 1D probability density functions represented as piecewise linear functions. Previously, this would need to have been approximated via numerical integration, or equivalent, for which calculating gradients is impractical. Implementation details are explored and experimental validation is presented. This paper contributes the necessary theory for the practical optimisation of information theoretic objectives when dealing with piecewise linear distributions directly; applications will be explored in future papers.

1 Introduction

Information theory (Shannon, 1948) provides a mathematical toolbox, that, while originally motivated by the problem of communication over noisy channels, has proved essential to artificial intelligence (AI). It often appears as a training objective for machine learning (ML), e.g. information gain for random forests (Sethi & Sarvarayudu, 1982). It is integral to the evidence lower bound (ELBO) in variational inference (Jordan et al., 1999), as used by techniques such as variational autoencoders (Kingma & Welling, 2014) and diffusion models (Sohl-Dickstein et al., 2015). Cross entropy loss (Hinton et al., 1995) is often preferred over the classical mean squared error (Prince, 2023) for neural networks, e.g. in transformers (Devlin et al., 2019). It also plays a part in theory, e.g. the maximum entropy principle for selecting distribution of Jaynes (1957).

In practise the uses of information theory are relatively simple. Most algorithms use Monte Carlo integration (Metropolis et al., 1953) to calculate the mismatch between data and model, with the training set acting as a fixed sample from the target distribution. It is regularly used as a loss function for discrete classification, but in the context of continuous regression alternate objectives are common, as differential cross entropy is hard to calculate in general. Variational methods often limit themselves to the exponential family (Wainwright & Jordan, 2008), for which analytic expressions are known. This motivates this paper: adding to the list of probabilistic objects with a known cross entropy increases the design space available for practical algorithms.

This papers novel contribution is an analytic expression for the cross entropy between two 1D piecewise linear probability density functions (PDFs),

$$H(P,Q) = -\sum_i \delta_i \left[\frac{p_i \log(q_i) + p_{i+1}\log(q_{i+1})}{2} - \frac{p_{i+1}q_i^2 - p_i q_{i+1}^2}{2(q_{i+1}-q_i)^2}\left(\log(q_{i+1}) - \log(q_i)\right) - \frac{(3p_i + p_{i+1})q_{i+1} - (p_i + 3p_{i+1})q_i}{4(q_{i+1}-q_i)} \right]. \quad (1)$$

For the purpose of open access the author has applied a Creative Commons Attribution (CC-BY) licence to any Author Accepted Manuscript version arising. A complete implementation, including code to generate the included figures, is in the supplementary material and also available from ******. There is no data.

It sums over every linear segment, indexed by i, with p_i the PDF P evaluated at the start of the segment and p_{i+1} evaluated at the end. The same relationship holds for q and Q while δ_i is the width of segment i. Entropy ($= H(P, P)$) and the Kullback-Leibler divergence (Kullback & Leibler, 1951) ($= H(P, Q) - H(P, P)$) follow immediately and are also a novel contribution.

Practically, using numerical integration is simpler and will work for many applied problems. But in an AI/ML context, where cross entropy is often an optimisation objective, the need for an analytical (and differentiable) result is clear. Numerical integration is also slow, requiring many samples for an accurate result, though in some cases this can be avoided with stochastic gradient descent (Robbins & Monro, 1951). Related work follows, with the derivation of the result in Section 3. This is followed by numerical validation in Section 4 and, finally, a conclusion.

2 Related work

A history of piecewise linear PDFs is presented, building up to various uses within AI/ML. Quadrature based approaches are discussed briefly. The triangular distribution (Simpson, 1755; Johnson, 1997), arguably the simplest, is the only piecewise linear density function for which an information theoretic calculation appears to be available: differential entropy (Lazo & Rathie, 1978). This result adapts to the trapezoidal distribution (René van Dorp & Kotz, 2003).

A histogram (Pearson, 1895) can be a piecewise linear PDF if it represents binned continuous data and area has been normalised. Its cross entropy is trivial. The frequency polygon (Scott, 1985b) connects the centres of a regular histogram, and is probably the earliest example of constructing a general piecewise linear PDF[1]. This is only consistent for a regular histogram however, because otherwise the maximum likelihood solutions differ between the two representations. History appears to have omitted or forgotten the obvious correction for this.

A variant is the edge frequency polygon (Jones et al., 1998), which connects the midpoints between bins (half way between bin heights at the edges between them) with linear segments. It confers no advantage over the frequency polygon, having the same convergence in a mean squared error sense. This convergence is matched by the kernel density estimate (Rosenblatt, 1956; Parzen, 1962), which generates piecewise linear PDFs if a piecewise linear kernel is used, such as a triangle. Alternatively, averaging many histograms with different origins (Scott, 1985a), such that the bins are misaligned, achieves a comparable effect with computational advantage. It is piecewise linear though has an excessive number of change points. Another approach is to take a kernel density estimate with any kernel and then fit a linear representation to it (Lin et al., 2006); this is of course approximate and requires a finite or truncated kernel.

Some approaches go directly to a piecewise linear representation. Beirlant et al. (1999) fit a linear segment to each bin of a histogram; this does result in discontinuities. Alternatively, Karlis & Xekalaki (2008) fit a mixture of triangular distributions, which generates an arbitrary polygon without discontinuities; they refer to it as the "polygonal distribution". A least squares approach with the segments fixed but their heights allowed to vary has been proposed (Wielen & Wielen, 2015), as has a maximum likelihood estimator (Nguyen & McLachlan, 2016). The segments remain connected for both. Nguyen & McLachlan (2018) have shown that the maximum likelihood estimator is consistent.

Perron & Mengersen (2001) take a non-parametric Bayesian approach, estimating a non-decreasing function as the cumulative distribution function of a mixture of triangular distributions. This approach models an explicit Poisson draw of the mixture count followed by a Dirichlet over membership, using reversible jump MCMC for inference. Alternatively, Ho et al. (2017) use a Dirichlet process prior (Ferguson, 1973) for doing a Bayesian density estimate with a mixture of triangular distributions. They are motivated by the problem of making a Bayesian estimate of an unknown distributions mode and utilise a Gibbs sampler with the stick breaking construction (Sethuraman, 1994).

Beyond explicit models built around piecewise linear PDFs there are also incidental uses within AI. Numerous models output histograms as discrete representations of continuous values. As an example stereo

[1]Earlier mentions exist but lack detail, e.g. Pearson in 1925 (Tarter & Kronmal, 1976).

algorithms for rectified images output disparity, for which examples utilising dynamic programming (Cox et al., 1996), belief propagation (Felzenszwalb & Huttenlocher, 2006), graph cuts (Boykov et al., 2001) and convolutional neural networks (Žbontar & LeCun, 2015) exist. Finally, as commonly taught when introducing neural networks (Prince, 2023), a network that only uses regularised linear units (ReLU) (Fukushima, 1975) generates a multivariate piecewise linear function.

For the purpose of verification numerical integration (Gibb, 1916) and Monte Carlo integration (Metropolis et al., 1953) are used in Section 4. Other choices exist, primarily those based on quadrature (Gonnet, 2012). This requires a family of functions for which the relevant information theoretic terms can be calculated. For PDFs the Gram-Charlier/Edgeworth series (Wallace, 1958) are commonly chosen, but you can also evaluate the integrand directly with a more general technique (Place & Stach, 1999). Hyvärinen (1997) has proposed a specific quadrature scheme designed for calculating differential entropy. While not the focus of this paper, the equation contributed does enable 1D quadrature with piecewise linear approximations for cross entropy. Examples of these approximations being used within machine learning include independent component analysis (Jutten & Herault, 1991) and projection pursuit (Huber, 1985).

3 Derivation

The cross entropy will now be derived. Note that, much like entropy has to define $0 \log(0) = 0$, similar issues occur. In the interest of clarity these are considered after the initial derivation. Differential cross entropy is defined as (Jaynes, 1963)

$$H(P, Q) = -\int P(x) \log Q(x) dx, \tag{2}$$

where P and Q are two PDFs and x is integrated over their domain. Taking P and Q to be piecewise linear and 1D we can consider the integral to be the sum of many linear sections,

$$-\sum_i \delta_i \int_0^1 ((1-t)p_i + tp_{i+1}) \log((1-t)q_i + tq_{i+1}) dt, \tag{3}$$

where p_i is P evaluated at the start of section i and p_{i+1} is P evaluated at the end; likewise for q and Q. δ_i is the section width, $x_{i+1} - x_i$. If the linear segments of P and Q are not in alignment then extra splits can be added.

Consider a single section, i, and the series (Olver et al., 2010, Equation. 4.6.4.),

$$\log(a) = 2 \sum_{n \in \{1,3,5,\ldots\}} \frac{1}{n} \left(\frac{a-1}{a+1} \right)^n, \tag{4}$$

as defined for $a > 0$. Combine to obtain

$$-\delta_i \int_0^1 ((1-t)p_i + tp_{i+1}) \left[2 \sum_{n \in \{1,3,5,\ldots\}} \frac{1}{n} \left(\frac{(1-t)q_i + tq_{i+1} - 1}{(1-t)q_i + tq_{i+1} + 1} \right)^n \right] dt, \tag{5}$$

which rearranges to

$$-2\delta_i \sum_{n \in \{1,3,5,\ldots\}} \frac{1}{n} \int_0^1 (p_i + (p_{i+1} - p_i)t) \left(\frac{q_i - 1 + (q_{i+1} - q_i)t}{q_i + 1 + (q_{i+1} - q_i)t} \right)^n dt. \tag{6}$$

Apply integration by parts and clean up,

$$-\delta_i \frac{p_i + p_{i+1}}{2} \log(q_{i+1}) - 2\delta_i (q_{i+1} - q_i) \sum_{n=1}^{\infty} (-1)^n \left\{ \int_0^1 \left(p_i t + \frac{(p_{i+1} - p_i)t^2}{2} \right) \frac{(q_i - 1 + (q_{i+1} - q_i)t)^{n-1}}{(q_i + 1 + (q_{i+1} - q_i)t)^n} dt \right\}, \tag{7}$$

then repeat,

$$-\delta_i \frac{p_i + p_{i+1}}{2} \log(q_{i+1}) - 2\delta_i(q_{i+1} - q_i) \sum_{n=1}^{\infty} (-1)^n \left\{ \left(\frac{p_i}{2} + \frac{(p_{i+1} - p_i)}{6} \right) \frac{(q_i - 1 + (q_{i+1} - q_i))^{n-1}}{(q_i + 1 + (q_{i+1} - q_i))^n} \right\} +$$
$$2\delta_i(q_{i+1} - q_i)^2 \left[\sum_{n=1}^{\infty} (-1)^n \left\{ \int_0^1 \left(\frac{p_i t^2}{2} + \frac{(p_{i+1} - p_i)t^3}{6} \right) \left\{ \frac{(n-1)(q_i - 1 + (q_{i+1} - q_i)t)^{n-2}}{(q_i + 1 + (q_{i+1} - q_i)t)^n} \right\} dt \right\} - \right.$$
$$\left. \sum_{n=1}^{\infty} (-1)^n \left\{ \int_0^1 \left(\frac{p_i t^2}{2} + \frac{(p_{i+1} - p_i)t^3}{6} \right) \left\{ \frac{n(q_i - 1 + (q_{i+1} - q_i)t)^{n-1}}{(q_i + 1 + (q_{i+1} - q_i)t)^{n+1}} \right\} dt \right\} \right]. \quad (8)$$

The first term of the second infinite sequence evaluates to zero due to the $n - 1$ term — remove by adding one to n throughout. This matches up the powers of the terms within the second and third infinite series, which may now be combined,

$$-\delta_i \frac{p_i + p_{i+1}}{2} \log(q_{i+1}) - \delta_i \frac{q_{i+1} - q_i}{q_{i+1} + 1} \left(\frac{2p_i + p_{i+1}}{3} \right) \sum_{n=1}^{\infty} (-1)^n \left\{ \frac{(q_{i+1} - 1)^{n-1}}{(q_{i+1} + 1)^{n-1}} \right\} -$$
$$2\delta_i(q_{i+1} - q_i)^2 \sum_{n=1}^{\infty} (-1)^n \left\{ \int_0^1 \left(p_i t^2 + \frac{(p_{i+1} - p_i)t^3}{3} \right) \left\{ \frac{n(q_i - 1 + (q_{i+1} - q_i)t)^{n-1}}{(q_i + 1 + (q_{i+1} - q_i)t)^{n+1}} \right\} dt \right\}. \quad (9)$$

The first infinite sequence is a geometric series, for which the closed form equivalence $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$ applies because, by construction, $|r| < 1$. This gives

$$-\delta_i \frac{p_i + p_{i+1}}{2} \log(q_{i+1}) + \delta_i \frac{(2p_i + p_{i+1})(q_{i+1} - q_i)}{6q_{i+1}} -$$
$$2\delta_i(q_{i+1} - q_i)^2 \sum_{n=1}^{\infty} n(-1)^n \left\{ \int_0^1 \left(p_i t^2 + \frac{(p_{i+1} - p_i)t^3}{3} \right) \left\{ \frac{(q_i - 1 + (q_{i+1} - q_i)t)^{n-1}}{(q_i + 1 + (q_{i+1} - q_i)t)^{n+1}} \right\} dt \right\} \quad (10)$$

This process, of integration by parts, merging two sequences and then collapsing a third, can be repeated until a pattern emerges. After the first round the sequence that collapses matches the definition of the polylogarithm (Jonquière, 1889) (the geometric series is a special case), for which an analytic expression exists (Wood, 1992). This gives

$$-\delta_i \frac{p_i + p_{i+1}}{2} \log(q_{i+1}) + \delta_i \sum_{n=1}^{\infty} \frac{(n+1)p_i + p_{i+1}}{n(n+1)(n+2)} \left(\frac{q_{i+1} - q_i}{q_{i+1}} \right)^n. \quad (11)$$

Next, use the substitution

$$\frac{1}{n(n+1)(n+2)} = \frac{(n-1)^2(n+2)}{4n} - \frac{n^2(n+3) + (n-1)n(n+2)}{4(n+1)} + \frac{n(n+1)(n+3)}{4(n+2)} \quad (12)$$

to generate many infinite series, each of which is identifiable as a polylogarithm of the form

$$\mathrm{Li}_s(\cdot) = \sum_{n=1}^{\infty} \frac{1}{n^s} \left(\frac{q_{i+1} - q_i}{q_{i+1}} \right)^n, \quad (13)$$

where the argument has been prefilled. This gives

$$
\begin{aligned}
-\delta_i \frac{p_i + p_{i+1}}{2} \log(q_{i+1}) + \delta_i \left(\frac{p_i}{4} - \frac{p_i q_{i+1}}{2(q_{i+1} - q_i)} + \frac{p_i q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{-3}(\cdot) + \\
\delta_i \left(\frac{p_i + p_{i+1}}{4} - \frac{(p_{i+1} - p_i) q_{i+1}}{2(q_{i+1} - q_i)} + \frac{(p_{i+1} - 3p_i) q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{-2}(\cdot) + \\
\delta_i \left(\frac{-3p_i}{4} - \frac{(-2p_i - p_{i+1}) q_{i+1}}{2(q_{i+1} - q_i)} + \frac{(p_i - 2p_{i+1}) q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{-1}(\cdot) + \\
\delta_i \left(\frac{-p_i - 3p_{i+1}}{4} - \frac{(p_i - p_{i+1}) q_{i+1}}{(q_{i+1} - q_i)} + \frac{(3p_i - p_{i+1}) q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{0}(\cdot) + \\
\delta_i \left(\frac{p_i + p_{i+1}}{2} - \frac{p_{i+1} q_{i+1}}{(q_{i+1} - q_i)} + \frac{(p_{i+1} - p_i) q_{i+1}^2}{2(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{1}(\cdot).
\end{aligned} \tag{14}
$$

Substituting the closed form definitions of the various polylogarithms (Wood, 1992) and substantial cancelling of terms obtains

$$
-\delta_i \frac{p_i \log(q_i) + p_{i+1} \log(q_{i+1})}{2} + \delta_i \frac{p_{i+1} q_i^2 - p_i q_{i+1}^2}{2(q_{i+1} - q_i)^2} \left(\log(q_{i+1}) - \log(q_i) \right) + \delta_i \frac{(3p_i + p_{i+1}) q_{i+1} - (p_i + 3p_{i+1}) q_i}{4(q_{i+1} - q_i)}, \tag{15}
$$

which completes the derivation needed to state Equation 1.

3.1 Singularity

As expected, $0 \log(0) = 0$ has to be specified for Equation 1 to work, i.e. this is Lebesgue integration. However, there is also a singularity when $q_i = q_{i+1}$. To ignore the second and third term when $q_{i+1} = q_i$ it must be that

$$
\lim_{q_{i+1} \to q_i} \left[\delta_i \frac{p_{i+1} q_i^2 - p_i q_{i+1}^2}{2(q_{i+1} - q_i)^2} \left(\log(q_{i+1}) - \log(q_i) \right) + \delta_i \frac{(3p_i + p_{i+1}) q_{i+1} - (p_i + 3p_{i+1}) q_i}{4(q_{i+1} - q_i)} \right] = 0. \tag{16}
$$

The two terms have to be considered simultaneously for this to be the case. Convert the first term of the limit back into an infinite sequence with Equation 4,

$$
\delta_i (p_{i+1} q_i^2 - p_i q_{i+1}^2) \sum_{n \in \{1,3,5,\ldots\}} \frac{1}{n} \frac{(q_{i+1} - q_i)^{n-2}}{(q_{i+1} + q_i)^n}, \tag{17}
$$

and note that the limit is trivially true for $n = 3$ onwards. Put the $n = 1$ term only back into the limit,

$$
\lim_{q_{i+1} \to q_i} \left[\delta_i (p_{i+1} q_i^2 - p_i q_{i+1}^2) \frac{(q_{i+1} - q_i)^{-1}}{(q_{i+1} + q_i)^1} + \delta_i \frac{(3p_i + p_{i+1}) q_{i+1} - (p_i + 3p_{i+1}) q_i}{4(q_{i+1} - q_i)} \right], \tag{18}
$$

and rearrange to obtain

$$
\lim_{q_{i+1} \to q_i} \left[\delta_i \frac{(p_{i+1} - p_i)(q_{i+1} - q_i)}{4(q_{i+1} + q_i)} \right] \tag{19}
$$

which is simply zero.

3.2 Implementation

Computing with Equation 15 is stable when a safe distance from the singularity, but becomes unstable due to the limits of floating point maths (IEEE, 2019). A stable everywhere alternative is

$$
-\delta_i \frac{p_i \log(q_i) + p_{i+1} \log(q_{i+1})}{2} + \delta_i \frac{(p_{i+1} - p_i)(q_{i+1} - q_i)}{4(q_{i+1} + q_i)} + \delta_i \frac{p_{i+1} q_i^2 - p_i q_{i+1}^2}{(q_{i+1} + q_i)^2} \sum_{n \in \{1,3,5,\ldots\}} \frac{1}{n+2} \left(\frac{q_{i+1} - q_i}{q_{i+1} + q_i} \right)^n, \tag{20}
$$

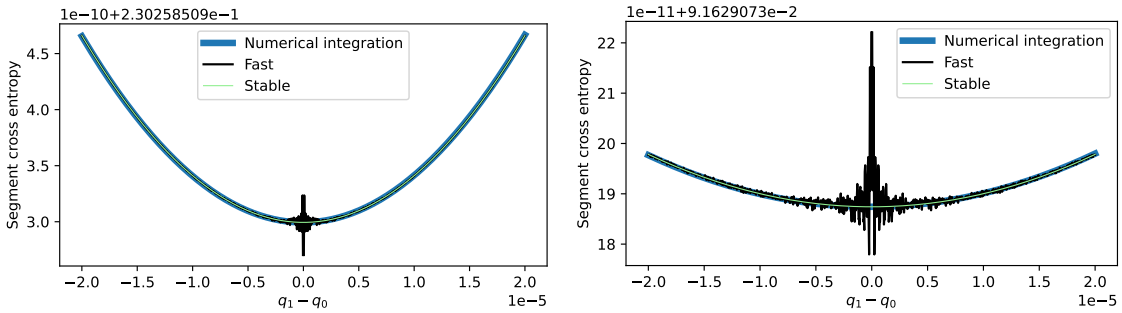

Figure 1: Zoomed in plots of the cross entropy of a linear segment, showing the error as the singularity is approached. Number top left is the scale of the y axis, bottom right the scale of the x axis. Calculated with `double` precision. For both plots $p_0 = p_1 = 0.1$ and $\delta_0 = 1$; for the left plot $\frac{q_0+q_1}{2} = 0.1$ while for the right $\frac{q_0+q_1}{2} = 0.4$. The x axis shows the difference between the two q values, i.e. they cross over such that the singularity is in the middle and the mean remains constant. This demonstrates how switching from Equation 15 to 20 only needs to occur when within $1e^{-5}$ of the singularity, with this range increasing with q. To obtain this precision with numerical integration required 2^{24} samples with averages organised over three levels to avoid underflow; it is over 90000× slower than the stable equation.

which is obtained from the limit calculation above. The infinite series converges quite slowly; use Equation 4 and assume without loss of generality, that $q_{i+1} \geq q_i$, hence

$$0 \leq \sum_{n \in \{1,3,5,\ldots\}} \frac{1}{n+r} \left(\frac{q_{i+1} - q_i}{q_{i+1} + q_i} \right)^n < \frac{1}{2} \log \left(\frac{q_{i+1}}{q_i} \right), \tag{21}$$

where $r > 0$, and $r = 2$ gets a bound on Equation 20. If ϵ is the error after running to $n = N - 2$, inclusive, then

$$\epsilon = \sum_{n \in \{N, N+2, N+4, \ldots\}} \frac{1}{n+2} \left(\frac{q_{i+1} - q_i}{q_{i+1} + q_i} \right)^n = \left(\frac{q_{i+1} - q_i}{q_{i+1} + q_i} \right)^{N-1} \sum_{n \in \{1,3,5,\ldots\}} \frac{1}{n+N+1} \left(\frac{q_{i+1} - q_i}{q_{i+1} + q_i} \right)^n \tag{22}$$

such that

$$0 \leq \epsilon < \frac{1}{2} \left(\frac{q_{i+1} - q_i}{q_{i+1} + q_i} \right)^{N-1} \log \left(\frac{q_{i+1}}{q_i} \right). \tag{23}$$

In practise $\frac{q_{i+1}-q_i}{q_{i+1}+q_i}$ is almost always small and the bound is quite loose[2], so you obtain multiple bits of precision with each iteration.

Using a basic Python/`numpy` implementation Equation 20 is 40× slower than Equation 15 — the preference is to use the fast equation where possible. Figure 1 explores where the transition between approaches should occur: within $|q_0 - q_1| \leq 1e^{-5}$ is the rule selected for the implementation in the supplementary material. A more complex rule may make sense depending on the hardware/likelihood of being close to the singularity.

4 Numerical validation

Three demonstrations have been selected, primarily to provide numerical validation. In all cases numerical integration (Gibb, 1916) and Monte Carlo integration (Metropolis et al., 1953) have been used to verify the result. Operations have been done with `float` (32 bit) precision for the presented approach, consistent with the precision commonly used on GPUs. For the first two demonstrations there is also a direct solution ("*True*"), because they use Gaussian distributions, but note that the Gaussians have to be converted to piecewise linear PDFs, introducing some error that is then reflected by the proposed approach ("*Analytic*")

[2]It can be made tighter with polylogarithms

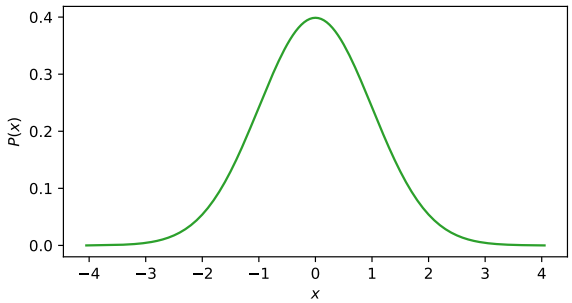

Figure 2: The standard Gaussian ($\mu = 0, \sigma^2 = 1$), represented as a piecewise linear PDF and hence truncated. Which, due to how curve rendering works in the digital realm, is identical to rendering a standard Gaussian without explicitly constructing a piecewise linear PDF. The piecewise linear approximation was made by matching the area under the linear segments with the area under the Gaussian's CDF, with a regular spacing.

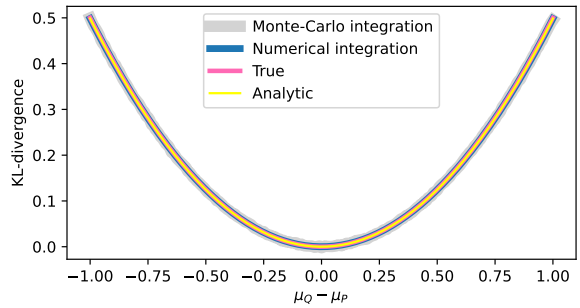

Figure 3: The Kullback-Leibler divergence, calculated via cross entropy with $H(P, Q) - H(P, P)$, of two standard width Gaussian distributions as their means are varied. The x axis shows the difference between the means. Four approaches have been used to calculate this result: the known solution ("*True*"), and then three approaches based on a piecewise linear approximation with "*Analytic*" the presented approach.

as well as both of the sampling integrators. The conclusion throughout is that the presented approach is accurate.

First, Figure 2 shows a truncated standard Gaussian distribution (Gauss, 1823). It has been represented as a piecewise linear PDF: the area under each linear section was matched to the Gaussian and then it was renormalised, to account for the truncation. Entropy of the standard Gaussian is known to be ~ 1.4189385. The presented approach gives 1.4189711, noting that some variation is expected due to the truncation and linear approximation. Numerical integration gives 1.4189712, identical to `float` precision. Monte Carlo integration gives 1.4191646, but is using the same number of samples as numerical integration (2^{24}) which is not enough to match the precision — it's a poor choice of integrator for a 1D function.

Figure 3 shows a sweep of the Kullback-Leibler divergence between two standard width Gaussians as the delta between their means varies. As before there is a known equation for this ("*True*"), which is again expected to not match exactly due to the distributions being represented with linear sections. The maximum difference between numerical integration with 2^{24} samples and the analytic solution is 0.000594; increasing the numerical integration sample count made no further difference. Based on the results elsewhere it is reasonable to believe that numerical integration is converging poorly.

Finally, Figure 4 shows the cross entropy directly, calculated for a sweep as two parts of a mixture distribution swap position. This reflects the more arbitrary distributions that justify the use of a piecewise linear distribution. There is no known equation for this, but the maximum difference observed between numerical integration with 2^{24} samples and the presented analytic approach is $5.8e^{-6}$.

5 Conclusion

This is an enabling paper: no immediately useful algorithm has been presented. Instead, a valuable operation for a flexible distribution representation has been presented. This will enable the use of piecewise linear PDF representations for constructing AI/ML algorithms where, previously, it was computationally impractical or impossible. Potential applications are found throughout the introduction, but this paper was motivated by a potential approach to variational inference. While not the primary motivation, this approach can be used to approximate cross entropy for any 1D PDF for which a piecewise linear approximation can be made. An evaluation of doing this against alternatives is future work.

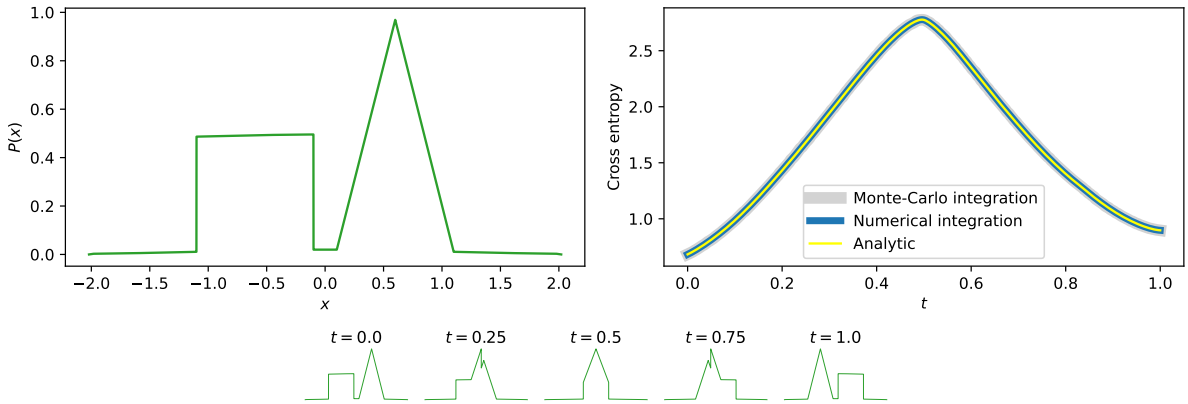

Figure 4: The top left graph shows a mixture distribution: a uniform distribution on the left and a triangular distribution on the right, with equal probability, plus a (low probability) wide Gaussian to avoid infinities in the following, converted into a piecewise linear PDF. As shown by the bottom row, the top left graph is showing $t = 0$: at $t = 1$ the cube and triangle have switched places, with them moving linearly for the transition, e.g. at $t = 0.5$ the cube is wearing a hat. The top right graph plots the cross entropy, $H(P, Q)$, where P is always at $t = 0$ while Q varies from $t = 0$ to $t = 1$. At the right we have the entropy of the top graph, in the centre the shapes are poorly aligned, giving the highest cross entropy, then, when they have swapped places, it's a better match, if imperfect. The analytic solution is shown alongside numerical and Monte Carlo integration, for verification.

A substantially extended version of the derivation is presented in the appendix. The supplementary material is a generic library for calculating the differential cross entropy, KL-divergence and entropy of piecewise linear PDFs, plus the code used to generate the figures[3].

References

Jan Beirlant, Alain Berlinet, and László Györfi. On piecewise linear density estimators. *Statistica Neerlandica*, 53(3):287–308, 1999.

Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

Ingemar J. Cox, Sunita L. Hingorani, Satish B. Rao, and Bruce M. Maggs. A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3):542–567, 1996.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *North American Chapter of the Association for Computational Linguistics*, 2019.

Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70:41–54, 2006.

Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pp. 209–230, 1973.

Kunihiko Fukushima. Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20 (3–4):121–136, 1975.

Carl F. Gauss. *Theoria combinationis observationum erroribus minimis obnoxiae*. Henricus Dieterich, 1823.

[3]This library is named *orogram*; same construction as *histogram* except the wooden posts ("*histo*") have been switched for mountains ("*oro*") to reflect the triangular nature of a piecewise linear PDF.

David Gibb. *A Course in Interpolation and Numerical Integration for the Mathematical Laboratory*. G. Bell & Sons, Limited, 1916.

Pedro Gonnet. A review of error estimation in adaptive quadrature. *Computing Surveys*, 44(4), 2012.

Geoffrey E. Hinton, Peter Dayan, Brendan J. Frey, and Radford M. Neal. The "wake-sleep" algorithm for unsupervised neural networks. *Science*, 268(5214):1158–1161, 1995.

Chi-san Ho, Paul Damien, and Stephen Walker. Bayesian mode regression using mixtures of triangular densities. *Journal of Econometrics*, 197(2):273–283, 2017.

Peter J. Huber. Projection pursuit. *The Annals of Statistics*, 13(2):435–475, 1985.

Aapo Hyvärinen. New approximations of differential entropy for independent component analysis and projection pursuit. *Advances in Neural Information Processing Systems*, 10:273–279, 1997.

IEEE. 754-2019 — IEEE standard for floating-point arithmetic, 2019.

Edwin T. Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620–630, 1957.

Edwin T. Jaynes. Information theory and statistical mechanics (lecture). *Brandeis University Summer Institute Lectures in Theoretical Physics*, pp. 181–218, 1963.

David Johnson. The triangular distribution as a proxy for the beta distribution in risk analysis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(3):387–398, 1997.

M. C. Jones, M. Samiuddin, A. H. Al-Harbey, and T. A. H. Maatouk. The edge frequency polygon. *Biometrika*, 85(1):235–239, 1998.

Alfred Jonquière. Note sur la série $\sum_{n=1}^{\infty} \frac{x^n}{n^s}$. *Bulletin de la Société Mathématique de France*, 17:142–152, 1889.

Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37:183–233, 1999.

Christian Jutten and Jeanny Herault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, 1991.

Dimitris Karlis and Evdokia Xekalaki. The polygonal distribution. *Advances in Mathematical and Statistical Modeling*, pp. 21–33, 2008.

Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. *International Conference on Learning Representations*, 2014.

Solomon Kullback and Richard A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

Aida C. G. Verdugo Lazo and Pushpa N. Rathie. On the entropy of continuous probability distributions. *Transactions of Information Theory*, 24(1):120–122, 1978.

Chien-Tai Lin, Jyh-Shyang Wu, and Chia-Hung Yen. A note on kernel polygons. *Biometrika*, 93(1):228–234, 2006.

Nicholas C. Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6): 1087–1092, 1953.

Hien D. Nguyen and Geoff J. McLachlan. Maximum likelihood estimation of triangular and polygonal distributions. *Communications in Statistics — Theory and Methods*, 102:23–36, 2016.

Hien D. Nguyen and Geoff J. McLachlan. Some theoretical results regarding the polygonal distribution. *Computational Statistics & Data Analysis*, 47(20):5083–5095, 2018.

Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark. *The NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

Emanuel Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

Karl Pearson. X. Contributions to the mathematical theory of evolution. — ii. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society A*, 186:343–414, 1895.

Francois Perron and Kerrie Mengersen. Bayesian nonparametric modeling using mixtures of triangular distributions. *Biometrics*, 57(2):518–528, 2001.

Jerry Place and Jerry Stach. Efficient numerical integration using Gaussian quadrature. *Simulation*, 73(4): 232–238, 1999.

Simon J. D. Prince. *Understanding Deep Learning*. MIT Press, 2023.

Johan René van Dorp and Samuel Kotz. Generalized trapezoidal distributions. *Metrika*, 58:85–97, 2003.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.

David W. Scott. Averaged shifted histograms: effective nonparametric density estimators in several dimensions. *The Annals of Statistics*, 13:1024–1040, 1985a.

David W. Scott. Frequency polygons: theory and application. *Journal of the American Statistical Association*, 80(390):348–354, 1985b.

Ishwar K. Sethi and G. P. R. Sarvarayudu. Hierarchical classifier design using mutual information. *Pattern Analysis and Machine Intelligence*, 4:441–445, 1982.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4(2):639–650, 1994.

Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3): 379–423, 1948.

Thomas Simpson. On the advantage of taking the mean of a number of observations, in practical astronomy. *Philosophical Transactions of the Royal Society*, 49:82–93, 1755.

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. *International Conference on Machine Learning*, 32:2256–2265, 2015.

Michael E. Tarter and Richard A. Kronmal. An introduction to the implementation and theory of nonparametric density estimation. *The American Statistician*, 30(3):105–112, 1976.

Jure Žbontar and Yann LeCun. Computing the stereo matching cost with a convolutional neural network. *Computer Vision and Pattern Recognition*, pp. 1592–1599, 2015.

Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008.

David L. Wallace. Asymptotic approximations to distributions. *Annals of Mathematical Statistics*, 29(3): 635–654, 1958.

Michael Vander Wielen and Ryan Vander Wielen. The general segmented distribution. *Communications in Statistics — Theory and Methods*, 44(10):1994–2009, 2015.

David C. Wood. The computation of polylogarithms. Technical Report 15–92*, University of Kent, 1992.

A Full derivation

The process of doing the derivation was digital. Specifically, Atom[4] was used to edit Markdown with inline Latex equations using MathJAX. Consequentially, it is excessively detailed: each step was done by copying and pasting the previous line then editing it. The below is the raw proof; it probably doubles as a statement about mathematical paranoia. While inclusion is necessary to ensure complete rigour, it's probably more valuable as an example of the difference between what goes into the body of a paper and what's actually done, to inform and/or terrify future researchers. It does still serve the purpose of clarifying the steps of the derivation. Some editing of the surrounding text has been done for clarity, mostly adding details that were originally omitted and adapting to formatting changes, but the mathematical steps have not been edited at all.

Differential cross entropy is defined as,

$$H(P,Q) = -\int P(x)\log Q(x)dx, \tag{24}$$

where the integral is over the (shared) domain of the two distributions, P and Q. In the linear case this becomes

$$H(P,Q) = -\sum_i \delta_i \int_0^1 ((1-t)p_i + tp_{i+1})\log((1-t)q_i + tq_{i+1})dt, \tag{25}$$

where the sum is over each piecewise linear section, with δ_i that sections width (i to $i+1$). Lowercase letters indicate an evaluation of the respective uppercase PDF, at the edge of a linear segment: subscript i at the start of segment i, subscript $i+1$ at the end. Hence, we need to be able to evaluate

$$-\delta_i \int_0^1 ((1-t)p_i + tp_{i+1})\log((1-t)q_i + tq_{i+1})dt. \tag{26}$$

Start with the series[5]

$$\log(a) = 2\sum_{n\in\{1,3,5,\ldots\}} \frac{1}{n}\left(\frac{a-1}{a+1}\right)^n, \tag{27}$$

defined for $a > 0$. Use it to rewrite the objective as

$$-\delta_i \int_0^1 ((1-t)p_i + tp_{i+1})\left[2\sum_{n\in\{1,3,5,\ldots\}} \frac{1}{n}\left(\frac{(1-t)q_i + tq_{i+1} - 1}{(1-t)q_i + tq_{i+1} + 1}\right)^n\right] dt = \tag{28}$$

$$-2\delta_i \sum_{n\in\{1,3,5,\ldots\}} \int_0^1 \frac{((1-t)p_i + tp_{i+1})}{n}\left(\frac{(1-t)q_i + tq_{i+1} - 1}{(1-t)q_i + tq_{i+1} + 1}\right)^n dt = \tag{29}$$

$$-2\delta_i \sum_{n\in\{1,3,5,\ldots\}} \frac{1}{n}\int_0^1 (p_i + (p_{i+1} - p_i)t)\left(\frac{q_i - 1 + (q_{i+1} - q_i)t}{q_i + 1 + (q_{i+1} - q_i)t}\right)^n dt. \tag{30}$$

Use

$$\frac{d}{dt}\left(p_i t + \frac{(p_{i+1} - p_i)t^2}{2}\right) = p_i + (p_{i+1} - p_i)t, \tag{31}$$

plus

$$\frac{d}{dt}\left(\frac{(q_i - 1 + (q_{i+1} - q_i)t)^m}{(q_i + 1 + (q_{i+1} - q_i)t)^n}\right) = \tag{32}$$

$$\frac{\frac{d}{dt}(q_i - 1 + (q_{i+1} - q_i)t)^m}{(q_i + 1 + (q_{i+1} - q_i)t)^n} + (q_i - 1 + (q_{i+1} - q_i)t)^m \frac{d}{dt}\frac{1}{(q_i + 1 + (q_{i+1} - q_i)t)^n} = \tag{33}$$

[4] A now defunct text editor.
[5] Equation 4.6.4, P.108 of NIST Handbook of Mathematical Functions, 2010

$$\frac{m(q_i-1+(q_{i+1}-q_i)t)^{m-1}(q_{i+1}-q_i)}{(q_i+1+(q_{i+1}-q_i)t)^n}-(q_i-1+(q_{i+1}-q_i)t)^m\frac{n(q_i+1+(q_{i+1}-q_i)t)^{n-1}(q_{i+1}-q_i)}{(q_i+1+(q_{i+1}-q_i)t)^{2n}}= \tag{34}$$

$$(q_{i+1}-q_i)\left\{\frac{m(q_i-1+(q_{i+1}-q_i)t)^{m-1}}{(q_i+1+(q_{i+1}-q_i)t)^n}-\frac{n(q_i-1+(q_{i+1}-q_i)t)^m}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right\}, \tag{35}$$

where some generalising has been done pre-emptively by introducing $m \leq n$ in anticipation of later, though for this first use $n = m$, Apply integration by parts to get

$$-2\delta_i\sum_{n\in\{1,3,5,\ldots\}}\frac{1}{n}\left\{\left[\left(p_it+\frac{(p_{i+1}-p_i)t^2}{2}\right)\left(\frac{q_i-1+(q_{i+1}-q_i)t}{q_i+1+(q_{i+1}-q_i)t}\right)^n\right]_0^1-\right.$$
$$\left.\int_0^1\left(p_it+\frac{(p_{i+1}-p_i)t^2}{2}\right)n(q_{i+1}-q_i)\left[\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^n}-\frac{(q_i-1+(q_{i+1}-q_i)t)^n}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right]dt\right\}= \tag{36}$$

$$-2\delta_i\left\{\frac{p_i+p_{i+1}}{2}\sum_{n\in\{1,3,5,\ldots\}}\frac{1}{n}\left\{\left(\frac{q_i-1+(q_{i+1}-q_i)}{q_i+1+(q_{i+1}-q_i)}\right)^n\right\}-\right.$$
$$(q_{i+1}-q_i)\sum_{n\in\{1,3,5,\ldots\}}\left\{\int_0^1\left(p_it+\frac{(p_{i+1}-p_i)t^2}{2}\right)\times\right.$$
$$\left.\left.\left[\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^n}-\frac{(q_i-1+(q_{i+1}-q_i)t)^n}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right]dt\right\}\right\}= \tag{37}$$

$$-\delta_i\frac{p_i+p_{i+1}}{2}\log(q_{i+1})+2\delta_i(q_{i+1}-q_i)\sum_{n\in\{1,3,5,\ldots\}}\left\{\int_0^1\left(p_it+\frac{(p_{i+1}-p_i)t^2}{2}\right)\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^n}dt\right\}$$
$$-2\delta_i(q_{i+1}-q_i)\sum_{n\in\{1,3,5,\ldots\}}\left\{\int_0^1\left(p_it+\frac{(p_{i+1}-p_i)t^2}{2}\right)\frac{(q_i-1+(q_{i+1}-q_i)t)^n}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}dt\right\}= \tag{38}$$

$$-\delta_i\frac{p_i+p_{i+1}}{2}\log(q_{i+1})-2\delta_i(q_{i+1}-q_i)\sum_{n=1}^{\infty}(-1)^n\left\{\int_0^1\left(p_it+\frac{(p_{i+1}-p_i)t^2}{2}\right)\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^n}dt\right\}, \tag{39}$$

which is moderately horrifying. Consider the remaining integration, poking up like an enormous boil,

$$\int_0^1\left(p_it+\frac{(p_{i+1}-p_i)t^2}{2}\right)\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^n}dt, \tag{40}$$

and prepare for integration by parts again. Need

$$\frac{d}{dt}\left(\frac{p_it^2}{2}+\frac{(p_{i+1}-p_i)t^3}{6}\right)=p_it+\frac{(p_{i+1}-p_i)t^2}{2} \tag{41}$$

in addition to the above, where this time the generalisation is used. Stepping through,

$$-\delta_i\frac{p_i+p_{i+1}}{2}\log(q_{i+1})-2\delta_i(q_{i+1}-q_i)\sum_{n=1}^{\infty}\left\{(-1)^n\int_0^1\left(p_it+\frac{(p_{i+1}-p_i)t^2}{2}\right)\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^n}dt\right\}= \tag{42}$$

$$-\delta_i\frac{p_i+p_{i+1}}{2}\log(q_{i+1})-$$
$$2\delta_i(q_{i+1}-q_i)\sum_{n=1}^{\infty}(-1)^n\left\{\left[\left(\frac{p_it^2}{2}+\frac{(p_{i+1}-p_i)t^3}{6}\right)\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^n}\right]_0^1-(q_{i+1}-q_i)\times\right.$$
$$\left.\int_0^1\left(\frac{p_it^2}{2}+\frac{(p_{i+1}-p_i)t^3}{6}\right)\left\{\frac{(n-1)(q_i-1+(q_{i+1}-q_i)t)^{n-2}}{(q_i+1+(q_{i+1}-q_i)t)^n}-\frac{n(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right\}dt\right\}= \tag{43}$$

$$
\begin{aligned}
&-\delta_i \frac{p_i+p_{i+1}}{2}\log(q_{i+1}) - 2\delta_i(q_{i+1}-q_i)\sum_{n=1}^{\infty}(-1)^n\left\{\left(\frac{p_i}{2}+\frac{(p_{i+1}-p_i)}{6}\right)\frac{(q_i-1+(q_{i+1}-q_i))^{n-1}}{(q_i+1+(q_{i+1}-q_i))^n}\right\}+\\
&2\delta_i(q_{i+1}-q_i)^2\left[\sum_{n=1}^{\infty}(-1)^n\left\{\int_0^1\left(\frac{p_it^2}{2}+\frac{(p_{i+1}-p_i)t^3}{6}\right)\left\{\frac{(n-1)(q_i-1+(q_{i+1}-q_i)t)^{n-2}}{(q_i+1+(q_{i+1}-q_i)t)^n}\right\}dt\right\}-\right.\\
&\left.\sum_{n=1}^{\infty}(-1)^n\left\{\int_0^1\left(\frac{p_it^2}{2}+\frac{(p_{i+1}-p_i)t^3}{6}\right)\left\{\frac{n(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right\}dt\right\}\right] = \quad (44)
\end{aligned}
$$

Note first entry of sequence of second line is zero, and offset to align powers with sequence of third line,

$$
\begin{aligned}
&-\delta_i \frac{p_i+p_{i+1}}{2}\log(q_{i+1}) - \delta_i(q_{i+1}-q_i)\left(p_i+\frac{(p_{i+1}-p_i)}{3}\right)\sum_{n=1}^{\infty}(-1)^n\left\{\frac{(q_{i+1}-1)^{n-1}}{(q_{i+1}+1)^n}\right\}-\\
&2\delta_i(q_{i+1}-q_i)^2\left[\sum_{n=1}^{\infty}(-1)^n\left\{\int_0^1\left(\frac{p_it^2}{2}+\frac{(p_{i+1}-p_i)t^3}{6}\right)\left\{\frac{n(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right\}dt\right\}+\right.\\
&\left.\sum_{n=1}^{\infty}(-1)^n\left\{\int_0^1\left(\frac{p_it^2}{2}+\frac{(p_{i+1}-p_i)t^3}{6}\right)\left\{\frac{n(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right\}dt\right\}\right] = \quad (45)
\end{aligned}
$$

$$
\begin{aligned}
&-\delta_i \frac{p_i+p_{i+1}}{2}\log(q_{i+1}) - \delta_i\frac{q_{i+1}-q_i}{q_{i+1}+1}\left(\frac{2p_i+p_{i+1}}{3}\right)\sum_{n=1}^{\infty}(-1)^n\left\{\frac{(q_{i+1}-1)^{n-1}}{(q_{i+1}+1)^{n-1}}\right\}-\\
&2\delta_i(q_{i+1}-q_i)^2\sum_{n=1}^{\infty}(-1)^n\left\{\int_0^1\left(p_it^2+\frac{(p_{i+1}-p_i)t^3}{3}\right)\left\{\frac{n(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right\}dt\right\} = \quad (46)
\end{aligned}
$$

Focus on second term on first line, which is an infinite geometric sequence,

$$
-\delta_i\frac{q_{i+1}-q_i}{q_{i+1}+1}\left(\frac{2p_i+p_{i+1}}{3}\right)\sum_{n=1}^{\infty}(-1)^n\left\{\frac{(q_{i+1}-1)^{n-1}}{(q_{i+1}+1)^{n-1}}\right\} = \quad (47)
$$

$$
+\delta_i\frac{q_{i+1}-q_i}{q_{i+1}+1}\left(\frac{2p_i+p_{i+1}}{3}\right)\sum_{n=0}^{\infty}(-1)^n\left\{\frac{(q_{i+1}-1)^{n}}{(q_{i+1}+1)^{n}}\right\} = \quad (48)
$$

$$
+\delta_i\frac{q_{i+1}-q_i}{q_{i+1}+1}\left(\frac{2p_i+p_{i+1}}{3}\right)\sum_{n=0}^{\infty}\left(-\frac{q_{i+1}-1}{q_{i+1}+1}\right)^n = \quad (49)
$$

and use $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$ as it is the case that $|r| < 1$ by construction,

$$
+\delta_i\frac{q_{i+1}-q_i}{q_{i+1}+1}\left(\frac{2p_i+p_{i+1}}{3}\right)\frac{1}{1+\frac{q_{i+1}-1}{q_{i+1}+1}} = \quad (50)
$$

$$
+\delta_i\frac{q_{i+1}-q_i}{q_{i+1}+1}\left(\frac{2p_i+p_{i+1}}{3}\right)\frac{q_{i+1}+1}{q_{i+1}+1+q_{i+1}-1} = \quad (51)
$$

$$
+\delta_i\frac{(2p_i+p_{i+1})(q_{i+1}-q_i)}{6q_{i+1}} \quad (52)
$$

Now put it back into the original equation,

$$
\begin{aligned}
&-\delta_i \frac{p_i+p_{i+1}}{2}\log(q_{i+1}) + \delta_i\frac{(2p_i+p_{i+1})(q_{i+1}-q_i)}{6q_{i+1}}-\\
&2\delta_i(q_{i+1}-q_i)^2\sum_{n=1}^{\infty}n(-1)^n\left\{\int_0^1\left(p_it^2+\frac{(p_{i+1}-p_i)t^3}{3}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right\}dt\right\}. \quad (53)
\end{aligned}
$$

Time for yet another integration of parts, which needs

$$\frac{d}{dt}\left(\frac{p_i t^3}{3}+\frac{(p_{i+1}-p_i)t^4}{12}\right)=p_i t^2+\frac{(p_{i+1}-p_i)t^3}{3} \tag{54}$$

plus the generalised derivative from above; consider just the final term,

$$-2\delta_i(q_{i+1}-q_i)^2\sum_{n=1}^{\infty}n(-1)^n\left\{\int_0^1\left(p_i t^2+\frac{(p_{i+1}-p_i)t^3}{3}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right\}dt\right\}= \tag{55}$$

$$\begin{aligned}&-2\delta_i(q_{i+1}-q_i)^2\sum_{n=1}^{\infty}n(-1)^n\left\{\left[\left(\frac{p_i t^3}{3}+\frac{(p_{i+1}-p_i)t^4}{12}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right)\right]_0^1-\right.\\&\int_0^1\left(\frac{p_i t^3}{3}+\frac{(p_{i+1}-p_i)t^4}{12}\right)(q_{i+1}-q_i)\times\\&\left.\left\{\frac{(n-1)(q_i-1+(q_{i+1}-q_i)t)^{n-2}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}-\frac{(n+1)(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}=\end{aligned} \tag{56}$$

$$\begin{aligned}&-2\delta_i(q_{i+1}-q_i)^2\sum_{n=1}^{\infty}n(-1)^n\left\{\left(\frac{p_i}{3}+\frac{(p_{i+1}-p_i)}{12}\right)\left(\frac{(q_{i+1}-1)^{n-1}}{(q_{i+1}+1)^{n+1}}\right)\right\}+\\&2\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(-1)^n\left\{\int_0^1\left(\frac{p_i t^3}{3}+\frac{(p_{i+1}-p_i)t^4}{12}\right)\left\{\frac{(n-1)(q_i-1+(q_{i+1}-q_i)t)^{n-2}}{(q_i+1+(q_{i+1}-q_i)t)^{n+1}}\right\}dt\right\}-\\&2\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(-1)^n\left\{\int_0^1\left(\frac{p_i t^3}{3}+\frac{(p_{i+1}-p_i)t^4}{12}\right)\left\{\frac{(n+1)(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}=\end{aligned} \tag{57}$$

Second line is zero for first entry in sequence, so shift along to align with last line,

$$\begin{aligned}&-2\delta_i(q_{i+1}-q_i)^2\sum_{n=1}^{\infty}n(-1)^n\left\{\left(\frac{3p_i+p_{i+1}}{12}\right)\left(\frac{(q_{i+1}-1)^{n-1}}{(q_{i+1}+1)^{n+1}}\right)\right\}-\\&2\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(\frac{p_i t^3}{3}+\frac{(p_{i+1}-p_i)t^4}{12}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}-\\&2\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(\frac{p_i t^3}{3}+\frac{(p_{i+1}-p_i)t^4}{12}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}=\end{aligned} \tag{58}$$

$$\begin{aligned}&+2\delta_i(q_{i+1}-q_i)^2\left(\frac{3p_i+p_{i+1}}{12}\right)\sum_{n=0}^{\infty}(n+1)(-1)^n\left(\frac{(q_{i+1}-1)^n}{(q_{i+1}+1)^{n+2}}\right)-\\&\frac{4}{3}\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(p_i t^3+\frac{(p_{i+1}-p_i)t^4}{4}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}=\end{aligned} \tag{59}$$

$$\begin{aligned}&+\frac{2}{3}\delta_i\left(\frac{q_{i+1}-q_i}{q_{i+1}+1}\right)^2\left(\frac{3p_i+p_{i+1}}{4}\right)\sum_{n=0}^{\infty}(n+1)\left(-\frac{q_{i+1}-1}{q_{i+1}+1}\right)^n-\\&\frac{4}{3}\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(p_i t^3+\frac{(p_{i+1}-p_i)t^4}{4}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}=\end{aligned} \tag{60}$$

The infinite sequence in the first line is an example of a polylogarithm, well, equivalent to two evaluations of, so

$$+\frac{2}{3}\delta_i\left(\frac{q_{i+1}-q_i}{q_{i+1}+1}\right)^2\left(\frac{3p_i+p_{i+1}}{4}\right)\left[\frac{1}{1+\frac{q_{i+1}-1}{q_{i+1}+1}}-\frac{\frac{q_{i+1}-1}{q_{i+1}+1}}{\left(1+\frac{q_{i+1}-1}{q_{i+1}+1}\right)^2}\right]-$$
$$\frac{4}{3}\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(p_it^3+\frac{(p_{i+1}-p_i)t^4}{4}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}= \quad (61)$$

$$+\frac{2}{3}\delta_i\left(\frac{3p_i+p_{i+1}}{4}\right)\left(\frac{q_{i+1}-q_i}{q_{i+1}+1}\right)^2\left[\frac{q_{i+1}+1}{2q_{i+1}}-\frac{(q_{i+1}+1)(q_{i+1}-1)}{4q_{i+1}^2}\right]-$$
$$\frac{4}{3}\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(p_it^3+\frac{(p_{i+1}-p_i)t^4}{4}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}= \quad (62)$$

$$+\delta_i\frac{(3p_i+p_{i+1})(q_{i+1}-q_i)^2}{24q_{i+1}^2}-$$
$$\frac{4}{3}\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(p_it^3+\frac{(p_{i+1}-p_i)t^4}{4}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}. \quad (63)$$

Now bring back the rest of the equation,

$$-\delta_i\frac{p_i+p_{i+1}}{2}\log(q_{i+1})+\delta_i\frac{(2p_i+p_{i+1})(q_{i+1}-q_i)}{6q_{i+1}}+\delta_i\frac{(3p_i+p_{i+1})(q_{i+1}-q_i)^2}{24q_{i+1}^2}-$$
$$\frac{4}{3}\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(p_it^3+\frac{(p_{i+1}-p_i)t^4}{4}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}. \quad (64)$$

This is suggestive, but not enough terms to be sure, so it's time to scream into the void again, with your integration by parts voice. Going to need

$$\frac{d}{dt}\left(\frac{p_it^4}{4}+\frac{(p_{i+1}-p_i)t^5}{20}\right)=p_it^3+\frac{(p_{i+1}-p_i)t^4}{4} \quad (65)$$

and the generalised derivative above. As before, consider only the final line,

$$-\frac{4}{3}\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(p_it^3+\frac{(p_{i+1}-p_i)t^4}{4}\right)\left\{\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right\}dt\right\}= \quad (66)$$

$$-\frac{4}{3}\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\left[\left(\frac{p_it^4}{4}+\frac{(p_{i+1}-p_i)t^5}{20}\right)\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right]_0^1-\right.$$
$$\int_0^1\left(\frac{p_it^4}{4}+\frac{(p_{i+1}-p_i)t^5}{20}\right)(q_{i+1}-q_i)\times$$
$$\left.\left(\frac{(n-1)(q_i-1+(q_{i+1}-q_i)t)^{n-2}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}-\frac{(n+2)(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}= \quad (67)$$

$$
\begin{aligned}
&-\frac{4}{3}\delta_i(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\left(\frac{p_i}{4}+\frac{(p_{i+1}-p_i)}{20}\right)\frac{(q_i-1+(q_{i+1}-q_i))^{n-1}}{(q_i+1+(q_{i+1}-q_i))^{n+2}}\right\}+\\
&\frac{4}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(\frac{p_it^4}{4}+\frac{(p_{i+1}-p_i)t^5}{20}\right)\left(\frac{(n-1)(q_i-1+(q_{i+1}-q_i)t)^{n-2}}{(q_i+1+(q_{i+1}-q_i)t)^{n+2}}\right)dt\right\}-\\
&\frac{4}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\int_0^1\left(\frac{p_it^4}{4}+\frac{(p_{i+1}-p_i)t^5}{20}\right)\left(\frac{(n+2)(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}=
\end{aligned}
\tag{68}
$$

Note that first entry of the second line's summation is zero, so shift to align with third line,

$$
\begin{aligned}
&-\frac{4}{12}\delta_i\left(p_i+\frac{(p_{i+1}-p_i)}{5}\right)(q_{i+1}-q_i)^3\sum_{n=1}^{\infty}n(n+1)(-1)^n\left\{\frac{(q_{i+1}-1)^{n-1}}{(q_{i+1}+1)^{n+2}}\right\}-\\
&\frac{4}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(\frac{p_it^4}{4}+\frac{(p_{i+1}-p_i)t^5}{20}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}-\\
&\frac{4}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(\frac{p_it^4}{4}+\frac{(p_{i+1}-p_i)t^5}{20}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}=
\end{aligned}
\tag{69}
$$

$$
\begin{aligned}
&+\frac{4}{12}\delta_i\left(\frac{4p_i+p_{i+1}}{5}\right)(q_{i+1}-q_i)^3\sum_{n=0}^{\infty}(n+1)(n+2)(-1)^n\left\{\frac{(q_{i+1}-1)^{n}}{(q_{i+1}+1)^{n+3}}\right\}-\\
&\frac{8}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(\frac{p_it^4}{4}+\frac{(p_{i+1}-p_i)t^5}{20}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}=
\end{aligned}
\tag{70}
$$

$$
\begin{aligned}
&+\frac{4}{12}\delta_i\left(\frac{4p_i+p_{i+1}}{5}\right)\frac{(q_{i+1}-q_i)^3}{(q_{i+1}+1)^3}\sum_{n=0}^{\infty}(2+3n+n^2)\left(-\frac{q_{i+1}-1}{q_{i+1}+1}\right)^n-\\
&\frac{2}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(p_it^4+\frac{(p_{i+1}-p_i)t^5}{5}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}=
\end{aligned}
\tag{71}
$$

The infinite series of the first line is equivalent to three polylogarithms of different order,

$$
\begin{aligned}
&+\frac{1}{3}\delta_i\left(\frac{4p_i+p_{i+1}}{5}\right)\frac{(q_{i+1}-q_i)^3}{(q_{i+1}+1)^3}\left[\frac{2}{1+\frac{q_{i+1}-1}{q_{i+1}+1}}-\frac{3\frac{q_{i+1}-1}{q_{i+1}+1}}{\left(1+\frac{q_{i+1}-1}{q_{i+1}+1}\right)^2}-\frac{\frac{q_{i+1}-1}{q_{i+1}+1}\left(1-\frac{q_{i+1}-1}{q_{i+1}+1}\right)}{\left(1+\frac{q_{i+1}-1}{q_{i+1}+1}\right)^3}\right]-\\
&\frac{2}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(p_it^4+\frac{(p_{i+1}-p_i)t^5}{5}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}=
\end{aligned}
\tag{72}
$$

$$
\begin{aligned}
&+\frac{1}{3}\delta_i\left(\frac{4p_i+p_{i+1}}{5}\right)\frac{(q_{i+1}-q_i)^3}{(q_{i+1}+1)^3}\left[\frac{q_{i+1}+1}{q_{i+1}}-\frac{3(q_{i+1}-1)(q_{i+1}+1)^2}{4(q_{i+1}+1)q_{i+1}^2}-\frac{2(q_{i+1}-1)(q_{i+1}+1)^3}{8(q_{i+1}+1)^2q_{i+1}^3}\right]-\\
&\frac{2}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(p_it^4+\frac{(p_{i+1}-p_i)t^5}{5}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}=
\end{aligned}
\tag{73}
$$

$$+\frac{1}{3}\delta_i\left(\frac{4p_i+p_{i+1}}{5}\right)\frac{(q_{i+1}-q_i)^3}{(q_{i+1}+1)^3}\left[\frac{4q_{i+1}(q_{i+1}+1)^2-3(q_{i+1}-1)(q_{i+1}+1)^2}{4(q_{i+1}+1)q_{i+1}^2}-\frac{2(q_{i+1}-1)(q_{i+1}+1)^3}{8(q_{i+1}+1)^2q_{i+1}^3}\right]-$$
$$\frac{2}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(p_it^4+\frac{(p_{i+1}-p_i)t^5}{5}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}= \tag{74}$$

$$+\frac{1}{3}\delta_i\left(\frac{4p_i+p_{i+1}}{5}\right)\frac{(q_{i+1}-q_i)^3}{(q_{i+1}+1)^3}\left[\frac{(q_{i+1}+3)(q_{i+1}+1)^2}{4(q_{i+1}+1)q_{i+1}^2}-\frac{2(q_{i+1}-1)(q_{i+1}+1)^3}{8(q_{i+1}+1)^2q_{i+1}^3}\right]-$$
$$\frac{2}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(p_it^4+\frac{(p_{i+1}-p_i)t^5}{5}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}= \tag{75}$$

$$+\frac{1}{3}\delta_i\left(\frac{4p_i+p_{i+1}}{5}\right)\frac{(q_{i+1}-q_i)^3}{(q_{i+1}+1)^3}\left[\frac{2q_{i+1}(q_{i+1}+3)(q_{i+1}+1)^3-2(q_{i+1}-1)(q_{i+1}+1)^3}{8(q_{i+1}+1)^2q_{i+1}^3}\right]-$$
$$\frac{2}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(p_it^4+\frac{(p_{i+1}-p_i)t^5}{5}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}= \tag{76}$$

$$+\frac{1}{3}\delta_i\left(\frac{4p_i+p_{i+1}}{5}\right)\frac{(q_{i+1}-q_i)^3}{(q_{i+1}+1)^3}\left[\frac{(q_{i+1}+1)^5}{4(q_{i+1}+1)^2q_{i+1}^3}\right]-$$
$$\frac{2}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(p_it^4+\frac{(p_{i+1}-p_i)t^5}{5}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}= \tag{77}$$

$$+\delta_i\frac{(4p_i+p_{i+1})(q_{i+1}-q_i)^3}{60q_{i+1}^3}-$$
$$\frac{2}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(p_it^4+\frac{(p_{i+1}-p_i)t^5}{5}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\}. \tag{78}$$

Now bring back the rest of the terms to get

$$-\delta_i\frac{p_i+p_{i+1}}{2}\log(q_{i+1})+\delta_i\frac{(2p_i+p_{i+1})(q_{i+1}-q_i)}{6q_{i+1}}+\delta_i\frac{(3p_i+p_{i+1})(q_{i+1}-q_i)^2}{24q_{i+1}^2}+\delta_i\frac{(4p_i+p_{i+1})(q_{i+1}-q_i)^3}{60q_{i+1}^3}-$$
$$\frac{2}{3}\delta_i(q_{i+1}-q_i)^4\sum_{n=1}^{\infty}n(n+1)(n+2)(-1)^n\left\{\int_0^1\left(p_it^4+\frac{(p_{i+1}-p_i)t^5}{5}\right)\left(\frac{(q_i-1+(q_{i+1}-q_i)t)^{n-1}}{(q_i+1+(q_{i+1}-q_i)t)^{n+3}}\right)dt\right\} \tag{79}$$

From this we can see the sequence is[6]

$$-\delta_i\frac{p_i+p_{i+1}}{2}\log(q_{i+1})+\delta_i\sum_{n=1}^{\infty}\frac{(n+1)p_i+p_{i+1}}{n(n+1)(n+2)}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^n \tag{80}$$

This is a pretty good solution, at least computationally speaking, and it has been verified by comparison to numerical integration. It is a little numerically unstable however and it looks like that last infinite sequence can be removed — it's time for the march of the polylogarithms.

[6]Divisor is https://oeis.org/A007531

Start with

$$\frac{1}{n(n+1)(n+2)} = \frac{(n-1)^2(n+2)}{4n} - \frac{n^2(n+3)+(n-1)n(n+2)}{4(n+1)} + \frac{n(n+1)(n+3)}{4(n+2)} \tag{81}$$

as derived in Subsection A.2. This allows the goal to be rewritten as

$$\begin{aligned}
&-\delta_i \frac{p_i+p_{i+1}}{2}\log(q_{i+1}) + \\
&\qquad \delta_i \sum_{n=1}^{\infty} \frac{(n-1)^2(n+2)((n+1)p_i+p_{i+1})}{4n}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^n - \\
&\delta_i \sum_{n=1}^{\infty} \frac{n^2(n+3)((n+1)p_i+p_{i+1})+(n-1)n(n+2)((n+1)p_i+p_{i+1})}{4(n+1)}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^n + \\
&\qquad \delta_i \sum_{n=1}^{\infty} \frac{n(n+1)(n+3)((n+1)p_i+p_{i+1})}{4(n+2)}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^n =
\end{aligned} \tag{82}$$

and then offset the sequences so the denominator is just $4n$,

$$\begin{aligned}
&-\delta_i \frac{p_i+p_{i+1}}{2}\log(q_{i+1}) + \\
&\qquad \delta_i \sum_{n=1}^{\infty} \frac{(n-1)^2(n+2)((n+1)p_i+p_{i+1})}{4n}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^n - \\
&\delta_i \sum_{n=2}^{\infty} \frac{(n-1)^2(n+2)(np_i+p_{i+1})+(n-2)(n-1)(n+1)(np_i+p_{i+1})}{4n}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^{n-1} + \\
&\qquad \delta_i \sum_{n=3}^{\infty} \frac{(n-2)(n-1)(n+1)((n-1)p_i+p_{i+1})}{4n}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^{n-2} =
\end{aligned} \tag{83}$$

Now correct the sequences to start from one by noting that the early terms are zero and correct for the exponents by pre-multiplication. It should be noted that there is a dependency on $0\log(0) = 0$ being the case here, as it usually is for terms like cross entropy.

$$\begin{aligned}
&-\delta_i \frac{p_i+p_{i+1}}{2}\log(q_{i+1}) + \\
&\qquad \delta_i \sum_{n=1}^{\infty} \frac{(n-1)^2(n+2)((n+1)p_i+p_{i+1})}{4n}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^n - \\
&\delta_i \left(\frac{q_{i+1}}{q_{i+1}-q_i}\right) \sum_{n=1}^{\infty} \frac{(n-1)^2(n+2)(np_i+p_{i+1})+(n-2)(n-1)(n+1)(np_i+p_{i+1})}{4n}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^n + \\
&\qquad \delta_i \left(\frac{q_{i+1}}{q_{i+1}-q_i}\right)^2 \sum_{n=1}^{\infty} \frac{(n-2)(n-1)(n+1)((n-1)p_i+p_{i+1})}{4n}\left(\frac{q_{i+1}-q_i}{q_{i+1}}\right)^n
\end{aligned} \tag{84}$$

Then explode each summed fraction in turn. First:

$$\frac{(n-1)^2(n+2)((n+1)p_i+p_{i+1})}{4n} = \frac{(n-1)^2(n+1)(n+2)p_i+(n-1)^2(n+2)p_{i+1}}{4n} = \tag{85}$$

$$\frac{(n^2-2n+1)(n^2+3n+2)p_i+(n^2-2n+1)(n+2)p_{i+1}}{4n} = \tag{86}$$

$$\frac{(n^4+n^3-3n^2-n+2)p_i+(n^3-3n+2)p_{i+1}}{4n} = \tag{87}$$

$$\frac{n^3p_i}{4}+\frac{n^2p_i}{4}-\frac{3np_i}{4}-\frac{p_i}{4}+\frac{p_i}{2n}+\frac{n^2p_{i+1}}{4}-\frac{3p_{i+1}}{4}+\frac{p_{i+1}}{2n} = \tag{88}$$

$$\frac{n^3 p_i}{4} + \frac{n^2(p_i + p_{i+1})}{4} - \frac{3np_i}{4} - \frac{p_i + 3p_{i+1}}{4} + \frac{p_i + p_{i+1}}{2n}. \tag{89}$$

Second:

$$\frac{(n-1)^2(n+2)(np_i + p_{i+1}) + (n-2)(n-1)(n+1)(np_i + p_{i+1})}{4n} = \tag{90}$$

$$\frac{(n^2 - 2n + 1)n(n+2)p_i + (n^2 - 2n + 1)(n+2)p_{i+1}}{4n} + \frac{(n^2 - 3n + 2)n(n+1)p_i + (n^2 - 3n + 2)(n+1)p_{i+1}}{4n} = \tag{91}$$

$$\frac{(n^4 - 3n^2 + 2n)p_i + (n^3 - 3n + 2)p_{i+1} + (n^4 - 2n^3 - n^2 + 2n)p_i + (n^3 - 2n^2 - n + 2)p_{i+1}}{4n} = \tag{92}$$

$$\frac{(2n^4 - 2n^3 - 4n^2 + 4n)p_i + (2n^3 - 2n^2 - 4n + 4)p_{i+1}}{4n} = \tag{93}$$

$$\frac{n^3 p_i}{2} + \frac{n^2(p_{i+1} - p_i)}{2} - \frac{n(2p_i + p_{i+1})}{2} + (p_i - p_{i+1}) + \frac{p_{i+1}}{n}. \tag{94}$$

Third:

$$\frac{(n-2)(n-1)(n+1)((n-1)p_i + p_{i+1})}{4n} = \frac{(n^2 - 3n + 2)(n+1)(n-1)p_i + (n^2 - 3n + 2)(n+1)p_{i+1}}{4n} = \tag{95}$$

$$\frac{(n^4 - 3n^3 + n^2 + 3n - 2)p_i + (n^3 - 2n^2 - n + 2)p_{i+1}}{4n} = \tag{96}$$

$$\frac{n^3 p_i}{4} + \frac{n^2(p_{i+1} - 3p_i)}{4} + \frac{n(p_i - 2p_{i+1})}{4} + \frac{(3p_i - p_{i+1})}{4} + \frac{p_{i+1} - p_i}{2n}. \tag{97}$$

For notational convenience define a pre-filled in polylogarithm of order s as

$$\mathrm{Li}_s(\cdot) = \sum_{n=1}^{\infty} \frac{1}{n^s} \left(\frac{q_{i+1} - q_i}{q_{i+1}} \right)^n \tag{98}$$

then put it all together and rewrite the equation as

$$\begin{aligned} -\delta_i \frac{p_i + p_{i+1}}{2} \log(q_{i+1}) + \delta_i \left(\frac{p_i}{4} - \frac{p_i q_{i+1}}{2(q_{i+1} - q_i)} + \frac{p_i q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{-3}(\cdot) + \\ \delta_i \left(\frac{p_i + p_{i+1}}{4} - \frac{(p_{i+1} - p_i) q_{i+1}}{2(q_{i+1} - q_i)} + \frac{(p_{i+1} - 3p_i) q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{-2}(\cdot) + \\ \delta_i \left(\frac{-3p_i}{4} - \frac{(-2p_i - p_{i+1}) q_{i+1}}{2(q_{i+1} - q_i)} + \frac{(p_i - 2p_{i+1}) q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{-1}(\cdot) + \\ \delta_i \left(\frac{-p_i - 3p_{i+1}}{4} - \frac{(p_i - p_{i+1}) q_{i+1}}{(q_{i+1} - q_i)} + \frac{(3p_i - p_{i+1}) q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{0}(\cdot) + \\ \delta_i \left(\frac{p_i + p_{i+1}}{2} - \frac{p_{i+1} q_{i+1}}{(q_{i+1} - q_i)} + \frac{(p_{i+1} - p_i) q_{i+1}^2}{2(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{1}(\cdot) \end{aligned} \tag{99}$$

This has been coded and verified, and unlike the earlier equation is numerically stable (yay!). But it's a mess, so keep going and simplify as much as possible.

Consider each term in turn, starting by dropping closed form definitions for the polylogarithms in, starting with order $s = -3$

$$\delta_i \left(\frac{p_i}{4} - \frac{p_i q_{i+1}}{2(q_{i+1} - q_i)} + \frac{p_i q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{-3}(\cdot) = \tag{100}$$

$$\delta_i\left(\frac{p_i(q_{i+1}-q_i)^2-2p_iq_{i+1}(q_{i+1}-q_i)+p_iq_{i+1}^2}{4(q_{i+1}-q_i)^2}\right)\frac{\frac{q_{i+1}-q_i}{q_{i+1}}\left(1+4\frac{q_{i+1}-q_i}{q_{i+1}}+\frac{(q_{i+1}-q_i)^2}{q_{i+1}^2}\right)}{\left(1-\frac{q_{i+1}-q_i}{q_{i+1}}\right)^4}= \tag{101}$$

$$\delta_i\left(\frac{p_iq_{i+1}^2-2p_iq_iq_{i+1}+p_iq_i^2-2p_iq_{i+1}^2+2p_iq_iq_{i+1}+p_iq_{i+1}^2}{4(q_{i+1}-q_i)^2}\right)\frac{(q_{i+1}-q_i)\left(1+4\frac{q_{i+1}-q_i}{q_{i+1}}+\frac{(q_{i+1}-q_i)^2}{q_{i+1}^2}\right)}{q_{i+1}\frac{q_i^4}{q_{i+1}^4}}= \tag{102}$$

$$\delta_i\left(\frac{p_iq_i^2}{4(q_{i+1}-q_i)}\right)\frac{q_{i+1}^3+4q_{i+1}^2(q_{i+1}-q_i)+q_{i+1}(q_{i+1}-q_i)^2}{q_i^4}= \tag{103}$$

$$\delta_i\left(\frac{p_iq_i^2}{4(q_{i+1}-q_i)}\right)\frac{q_{i+1}^3+4q_{i+1}^3-4q_iq_{i+1}^2+q_{i+1}^3-2q_iq_{i+1}^2+q_i^2q_{i+1}}{q_i^4}= \tag{104}$$

$$\delta_i\left(\frac{p_iq_i^2}{4(q_{i+1}-q_i)}\right)\frac{6q_{i+1}^3-6q_iq_{i+1}^2+q_i^2q_{i+1}}{q_i^4}= \tag{105}$$

$$\delta_i\frac{6p_iq_i^2q_{i+1}^2(q_{i+1}-q_i)}{4q_i^4(q_{i+1}-q_i)}+\delta_i\frac{p_iq_i^4q_{i+1}}{4q_i^4(q_{i+1}-q_i)}= \tag{106}$$

$$\delta_i\frac{3p_iq_{i+1}^2}{2q_i^2}+\delta_i\frac{p_iq_{i+1}}{4(q_{i+1}-q_i)}. \tag{107}$$

Now order $s=-2$

$$\delta_i\left(\frac{p_i+p_{i+1}}{4}-\frac{(p_{i+1}-p_i)q_{i+1}}{2(q_{i+1}-q_i)}+\frac{(p_{i+1}-3p_i)q_{i+1}^2}{4(q_{i+1}-q_i)^2}\right)\mathrm{Li}_{-2}(\cdot)= \tag{108}$$

$$\delta_i\left(\frac{(p_i+p_{i+1})(q_{i+1}-q_i)^2-2(p_{i+1}-p_i)q_{i+1}(q_{i+1}-q_i)+(p_{i+1}-3p_i)q_{i+1}^2}{4(q_{i+1}-q_i)^2}\right)\frac{\frac{q_{i+1}-q_i}{q_{i+1}}\left(1+\frac{q_{i+1}-q_i}{q_{i+1}}\right)}{\left(1-\frac{q_{i+1}-q_i}{q_{i+1}}\right)^3}= \tag{109}$$

$$\delta_i\left(\frac{\begin{array}{c}p_iq_{i+1}^2-2p_iq_iq_{i+1}+p_iq_i^2+p_{i+1}q_{i+1}^2-2p_{i+1}q_iq_{i+1}+p_{i+1}q_i^2-\\ 2p_{i+1}q_{i+1}^2+2p_iq_{i+1}^2+2p_{i+1}q_iq_{i+1}-2p_iq_iq_{i+1}+p_{i+1}q_{i+1}^2-3p_iq_{i+1}^2\end{array}}{4(q_{i+1}-q_i)^2}\right)\frac{(q_{i+1}-q_i)\left(1+\frac{q_{i+1}-q_i}{q_{i+1}}\right)}{q_{i+1}\frac{q_i^3}{q_{i+1}^3}}= \tag{110}$$

$$\delta_i\left(\frac{p_iq_i^2+p_{i+1}q_i^2-4p_iq_iq_{i+1}}{4(q_{i+1}-q_i)}\right)\frac{q_{i+1}^2+q_{i+1}(q_{i+1}-q_i)}{q_i^3}= \tag{111}$$

$$\delta_i\frac{(p_iq_i^2q_{i+1}+p_{i+1}q_i^2q_{i+1}-4p_iq_iq_{i+1}^2)(q_{i+1}-q_i)}{4q_i^3(q_{i+1}-q_i)}+\delta_i\frac{p_iq_i^2q_{i+1}^2+p_{i+1}q_i^2q_{i+1}^2-4p_iq_iq_{i+1}^3}{4q_i^3(q_{i+1}-q_i)}= \tag{112}$$

$$\delta_i\frac{p_iq_iq_{i+1}+p_{i+1}q_iq_{i+1}-4p_iq_{i+1}^2}{4q_i^2}+\delta_i\frac{p_iq_iq_{i+1}^2+p_{i+1}q_iq_{i+1}^2-4p_iq_{i+1}^3}{4q_i^2(q_{i+1}-q_i)}= \tag{113}$$

$$\delta_i\frac{-4p_iq_{i+1}^2}{4q_i^2}+\delta_i\frac{p_iq_iq_{i+1}(q_{i+1}-q_i)+p_{i+1}q_iq_{i+1}(q_{i+1}-q_i)+p_iq_iq_{i+1}^2+p_{i+1}q_iq_{i+1}^2-4p_iq_{i+1}^3}{4q_i^2(q_{i+1}-q_i)}= \tag{114}$$

$$\delta_i\frac{-p_iq_{i+1}^2}{q_i^2}+\delta_i\frac{2p_iq_iq_{i+1}^2-p_iq_i^2q_{i+1}+2p_{i+1}q_iq_{i+1}^2-p_{i+1}q_i^2q_{i+1}-4p_iq_{i+1}^3}{4q_i^2(q_{i+1}-q_i)}= \tag{115}$$

$$\delta_i\frac{-p_iq_{i+1}^2}{q_i^2}+\delta_i\frac{-p_iq_{i+1}^3}{q_i^2(q_{i+1}-q_i)}+\delta_i\frac{p_iq_{i+1}^2}{2q_i(q_{i+1}-q_i)}+\delta_i\frac{p_{i+1}q_{i+1}^2}{2q_i(q_{i+1}-q_i)}+\delta_i\frac{-p_iq_{i+1}}{4(q_{i+1}-q_i)}+\delta_i\frac{-p_{i+1}q_{i+1}}{4(q_{i+1}-q_i)}. \tag{116}$$

Onwards to order $s = -1$

$$\delta_i \left(\frac{-3p_i}{4} - \frac{(-2p_i - p_{i+1})q_{i+1}}{2(q_{i+1} - q_i)} + \frac{(p_i - 2p_{i+1})q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_{-1}(\cdot) = \tag{117}$$

$$\delta_i \left(\frac{-3p_i(q_{i+1} - q_i)^2 + 2(2p_i + p_{i+1})q_{i+1}(q_{i+1} - q_i) + (p_i - 2p_{i+1})q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \frac{\frac{q_{i+1} - q_i}{q_{i+1}}}{\left(1 - \frac{q_{i+1} - q_i}{q_{i+1}}\right)^2} = \tag{118}$$

$$\delta_i \left(\frac{\begin{array}{r} -3p_i q_{i+1}^2 + 6p_i q_i q_{i+1} - 3p_i q_i^2 + 4p_i q_{i+1}^2 + 2p_{i+1} q_{i+1}^2 - \\ 4p_i q_i q_{i+1} - 2p_{i+1} q_i q_{i+1} + p_i q_{i+1}^2 - 2p_{i+1} q_{i+1}^2 \end{array}}{4(q_{i+1} - q_i)^2} \right) \frac{(q_{i+1} - q_i)}{q_{i+1} \frac{q_i^2}{q_{i+1}^2}} = \tag{119}$$

$$\delta_i \left(\frac{2p_i q_{i+1}^2 + 2p_i q_i q_{i+1} - 2p_{i+1} q_i q_{i+1} - 3p_i q_i^2}{4(q_{i+1} - q_i)} \right) \frac{q_{i+1}}{q_i^2} = \tag{120}$$

$$\delta_i \frac{2p_i q_{i+1}^3 + 2p_i q_i q_{i+1}^2 - 2p_{i+1} q_i q_{i+1}^2 - 3p_i q_i^2 q_{i+1}}{4q_i^2(q_{i+1} - q_i)} = \tag{121}$$

$$\delta_i \frac{p_i q_{i+1}^3}{2q_i^2(q_{i+1} - q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1} - q_i)} + \delta_i \frac{-p_{i+1} q_{i+1}^2}{2q_i(q_{i+1} - q_i)} + \delta_i \frac{-3p_i q_{i+1}}{4(q_{i+1} - q_i)} \tag{122}$$

And, prancing on from stage left, order $s = 0$,

$$\delta_i \left(\frac{-p_i - 3p_{i+1}}{4} - \frac{(p_i - p_{i+1})q_{i+1}}{(q_{i+1} - q_i)} + \frac{(3p_i - p_{i+1})q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \mathrm{Li}_0(\cdot) = \tag{123}$$

$$\delta_i \left(\frac{-(p_i + 3p_{i+1})(q_{i+1} - q_i)^2 - 4(p_i - p_{i+1})q_{i+1}(q_{i+1} - q_i) + (3p_i - p_{i+1})q_{i+1}^2}{4(q_{i+1} - q_i)^2} \right) \frac{\frac{q_{i+1} - q_i}{q_{i+1}}}{1 - \frac{q_{i+1} - q_i}{q_{i+1}}} = \tag{124}$$

$$\delta_i \left(\frac{\begin{array}{r} -p_i q_{i+1}^2 + 2p_i q_i q_{i+1} - p_i q_i^2 - 3p_{i+1} q_{i+1}^2 + 6p_{i+1} q_i q_{i+1} - 3p_{i+1} q_i^2 - \\ 4p_i q_{i+1}^2 + 4p_i q_i q_{i+1} + 4p_{i+1} q_{i+1}^2 - 4p_{i+1} q_i q_{i+1} + 3p_i q_{i+1}^2 - p_{i+1} q_{i+1}^2 \end{array}}{4(q_{i+1} - q_i)^2} \right) \frac{(q_{i+1} - q_i)}{q_{i+1} \frac{q_i}{q_{i+1}}} = \tag{125}$$

$$\delta_i \frac{-2p_i q_{i+1}^2 - p_i q_i^2 - 3p_{i+1} q_i^2 + 6p_i q_i q_{i+1} + 2p_{i+1} q_i q_{i+1}}{4q_i(q_{i+1} - q_i)} = \tag{126}$$

$$\delta_i \frac{-p_i q_{i+1}^2}{2q_i(q_{i+1} - q_i)} + \delta_i \frac{-p_i q_i}{4(q_{i+1} - q_i)} + \delta_i \frac{-3p_{i+1} q_i}{4(q_{i+1} - q_i)} + \delta_i \frac{3p_i q_{i+1}}{2(q_{i+1} - q_i)} + \delta_i \frac{p_{i+1} q_{i+1}}{2(q_{i+1} - q_i)} \tag{127}$$

Finally, lets get funky with order $s = 1$

$$\delta_i \left(\frac{p_i + p_{i+1}}{2} - \frac{p_{i+1} q_{i+1}}{(q_{i+1} - q_i)} + \frac{(p_{i+1} - p_i)q_{i+1}^2}{2(q_{i+1} - q_i)^2} \right) \mathrm{Li}_1(\cdot) = \tag{128}$$

$$-\delta_i \left(\frac{(p_i + p_{i+1})(q_{i+1} - q_i)^2 - 2p_{i+1} q_{i+1}(q_{i+1} - q_i) + (p_{i+1} - p_i)q_{i+1}^2}{2(q_{i+1} - q_i)^2} \right) \log \left(1 - \frac{q_{i+1} - q_i}{q_{i+1}} \right) = \tag{129}$$

$$-\delta_i \left(\frac{\begin{array}{r} p_i q_{i+1}^2 - 2p_i q_i q_{i+1} + p_i q_i^2 + p_{i+1} q_{i+1}^2 - 2p_{i+1} q_i q_{i+1} + p_{i+1} q_i^2 - \\ 2p_{i+1} q_{i+1}^2 + 2p_{i+1} q_i q_{i+1} + p_{i+1} q_{i+1}^2 - p_i q_{i+1}^2 \end{array}}{2(q_{i+1} - q_i)^2} \right) \log \left(\frac{q_i}{q_{i+1}} \right) = \tag{130}$$

$$\delta_i \left(\frac{2p_i q_i q_{i+1} - p_i q_i^2 - p_{i+1} q_i^2}{2(q_{i+1} - q_i)^2} \right) (\log(q_i) - \log(q_{i+1})) \tag{131}$$

Lets stick all of the above terms together into one horrifying and bloated Harkonnen of an equation,

$$
\begin{aligned}
&-\delta_i \frac{p_i + p_{i+1}}{2} \log(q_{i+1}) + \delta_i \frac{3p_i q_{i+1}^2}{2q_i^2} + \delta_i \frac{p_i q_{i+1}}{4(q_{i+1} - q_i)} + \\
&\delta_i \frac{-p_i q_{i+1}^2}{q_i^2} + \delta_i \frac{-p_i q_{i+1}^3}{q_i^2(q_{i+1} - q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1} - q_i)} + \delta_i \frac{p_{i+1} q_{i+1}^2}{2q_i(q_{i+1} - q_i)} + \delta_i \frac{-p_i q_{i+1}}{4(q_{i+1} - q_i)} + \delta_i \frac{-p_{i+1} q_{i+1}}{4(q_{i+1} - q_i)} + \\
&\delta_i \frac{p_i q_{i+1}^3}{2q_i^2(q_{i+1} - q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1} - q_i)} + \delta_i \frac{-p_{i+1} q_{i+1}^2}{2q_i(q_{i+1} - q_i)} + \delta_i \frac{-3p_i q_{i+1}}{4(q_{i+1} - q_i)} + \\
&\delta_i \frac{-p_i q_{i+1}^2}{2q_i(q_{i+1} - q_i)} + \delta_i \frac{-p_i q_i}{4(q_{i+1} - q_i)} + \delta_i \frac{-3p_{i+1} q_i}{4(q_{i+1} - q_i)} + \delta_i \frac{3p_i q_{i+1}}{2(q_{i+1} - q_i)} + \delta_i \frac{p_{i+1} q_{i+1}}{2(q_{i+1} - q_i)} + \\
&\delta_i \left(\frac{2p_i q_i q_{i+1} - p_i q_i^2 - p_{i+1} q_i^2}{2(q_{i+1} - q_i)^2} \right) (\log(q_i) - \log(q_{i+1})) =
\end{aligned} \tag{132}
$$

Start by considering only the log terms, as they are a special kind of fiddly,

$$
-\delta_i \frac{p_i + p_{i+1}}{2} \log(q_{i+1}) + \delta_i \left(\frac{2p_i q_i q_{i+1} - p_i q_i^2 - p_{i+1} q_i^2}{2(q_{i+1} - q_i)^2} \right) (\log(q_i) - \log(q_{i+1})) = \tag{133}
$$

$$
\delta_i \left(\frac{\begin{aligned} &2p_i q_i q_{i+1} \log(q_i) - p_i q_i^2 \log(q_i) - p_{i+1} q_i^2 \log(q_i) - 2p_i q_i q_{i+1} \log(q_{i+1}) + \\ &p_i q_i^2 \log(q_{i+1}) + p_{i+1} q_i^2 \log(q_{i+1}) - (p_i + p_{i+1})(q_{i+1} - q_i)^2 \log(q_{i+1}) \end{aligned}}{2(q_{i+1} - q_i)^2} \right) = \tag{134}
$$

$$
\begin{aligned}
&\frac{\delta_i}{2(q_{i+1} - q_i)^2} \big(2p_i q_i q_{i+1} \log(q_i) - p_i q_i^2 \log(q_i) - p_{i+1} q_i^2 \log(q_i) - 2p_i q_i q_{i+1} \log(q_{i+1}) + p_i q_i^2 \log(q_{i+1}) + \\
&p_{i+1} q_i^2 \log(q_{i+1}) - p_i q_{i+1}^2 \log(q_{i+1}) + 2p_i q_i q_{i+1} \log(q_{i+1}) - p_i q_i^2 \log(q_{i+1}) - p_{i+1} q_{i+1}^2 \log(q_{i+1}) + \\
&2p_{i+1} q_i q_{i+1} \log(q_{i+1}) - p_{i+1} q_i^2 \log(q_{i+1}) \big) =
\end{aligned} \tag{135}
$$

$$
\begin{aligned}
&\frac{\delta_i}{2(q_{i+1} - q_i)^2} \big(2p_i q_i q_{i+1} \log(q_i) - p_i q_i^2 \log(q_i) - p_{i+1} q_i^2 \log(q_i) + 2p_{i+1} q_i q_{i+1} \log(q_{i+1}) - \\
&p_i q_{i+1}^2 \log(q_{i+1}) - p_{i+1} q_{i+1}^2 \log(q_{i+1}) \big) =
\end{aligned} \tag{136}
$$

Note that

$$
\begin{aligned}
&-(p_i \log(q_i) + p_{i+1} \log(q_{i+1}))(q_{i+1} - q_i)^2 = \\
&-p_i q_i^2 \log(q_i) + 2p_i q_i q_{i+1} \log(q_i) - \underline{p_i q_{i+1}^2 \log(q_i)} - \underline{p_{i+1} q_i^2 \log(q_{i+1})} + 2p_{i+1} q_i q_{i+1} \log(q_{i+1}) - p_{i+1} q_{i+1}^2 \log(q_{i+1})
\end{aligned} \tag{137}
$$

is almost, but not quite, a match for the numerator — two of the logs are swapped (underlined). This can be corrected by adding

$$
p_i q_{i+1}^2 \log(q_i) + p_{i+1} q_i^2 \log(q_{i+1}) - p_i q_{i+1}^2 \log(q_{i+1}) - p_{i+1} q_i^2 \log(q_i) = \tag{138}
$$

$$
(p_i q_{i+1}^2 - p_{i+1} q_i^2) \log(q_i) + (p_{i+1} q_i^2 - p_i q_{i+1}^2) \log(q_{i+1}) = \tag{139}
$$

$$
(p_{i+1} q_i^2 - p_i q_{i+1}^2) (\log(q_{i+1}) - \log(q_i)) \tag{140}
$$

and rewrite the log terms as

$$
-\delta_i \frac{p_i \log(q_i) + p_{i+1} \log(q_{i+1})}{2} + \delta_i \frac{p_{i+1} q_i^2 - p_i q_{i+1}^2}{2(q_{i+1} - q_i)^2} (\log(q_{i+1}) - \log(q_i)) \tag{141}
$$

which is really rather nice all things considered (there is the question of if it has a geometric meaning, which could lead to a much more elegant derivation), with the right kind of numerical stability.

Now work through all of the non-log terms:

$$\begin{aligned}
&\delta_i \frac{3p_i q_{i+1}^2}{2q_i^2} + \delta_i \frac{p_i q_{i+1}}{4(q_{i+1}-q_i)} + \\
&\delta_i \frac{-p_i q_{i+1}^2}{q_i^2} + \delta_i \frac{-p_i q_{i+1}^3}{q_i^2(q_{i+1}-q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{p_{i+1} q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{-p_i q_{i+1}}{4(q_{i+1}-q_i)} + \delta_i \frac{-p_{i+1} q_{i+1}}{4(q_{i+1}-q_i)} + \\
&\delta_i \frac{p_i q_{i+1}^3}{2q_i^2(q_{i+1}-q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{-p_{i+1} q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{-3p_i q_{i+1}}{4(q_{i+1}-q_i)} + \\
&\delta_i \frac{-p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{-p_i q_i}{4(q_{i+1}-q_i)} + \delta_i \frac{-3p_{i+1} q_i}{4(q_{i+1}-q_i)} + \delta_i \frac{3p_i q_{i+1}}{2(q_{i+1}-q_i)} + \delta_i \frac{p_{i+1} q_{i+1}}{2(q_{i+1}-q_i)} = \qquad (142)
\end{aligned}$$

$$\begin{aligned}
&\delta_i \frac{3p_i q_{i+1}^2}{2q_i^2} + \delta_i \frac{-2p_i q_{i+1}^2}{2q_i^2} + \delta_i \frac{-2p_i q_{i+1}^3}{2q_i^2(q_{i+1}-q_i)} + \delta_i \frac{p_i q_{i+1}^3}{2q_i^2(q_{i+1}-q_i)} + \\
&\delta_i \frac{p_{i+1} q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{-p_{i+1} q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{-p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \\
&\delta_i \frac{-p_i q_{i+1}}{4(q_{i+1}-q_i)} + \delta_i \frac{-p_{i+1} q_{i+1}}{4(q_{i+1}-q_i)} + \delta_i \frac{p_i q_{i+1}}{4(q_{i+1}-q_i)} + \delta_i \frac{-p_i q_i}{4(q_{i+1}-q_i)} + \\
&\delta_i \frac{-3p_i q_{i+1}}{4(q_{i+1}-q_i)} + \delta_i \frac{-3p_{i+1} q_i}{4(q_{i+1}-q_i)} + \delta_i \frac{6p_i q_{i+1}}{4(q_{i+1}-q_i)} + \delta_i \frac{2p_{i+1} q_{i+1}}{4(q_{i+1}-q_i)} = \qquad (143)
\end{aligned}$$

$$\delta_i \frac{p_i q_{i+1}^2}{2q_i^2} + \delta_i \frac{-p_i q_{i+1}^3}{2q_i^2(q_{i+1}-q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{(3p_i + p_{i+1})q_{i+1} - (p_i + 3p_{i+1})q_i}{4(q_{i+1}-q_i)} = \qquad (144)$$

$$\delta_i \frac{p_i q_{i+1}^3 - p_i q_i q_{i+1}^2 - p_i q_{i+1}^3}{2q_i^2(q_{i+1}-q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{(3p_i + p_{i+1})q_{i+1} - (p_i + 3p_{i+1})q_i}{4(q_{i+1}-q_i)} = \qquad (145)$$

$$\delta_i \frac{-p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{p_i q_{i+1}^2}{2q_i(q_{i+1}-q_i)} + \delta_i \frac{(3p_i + p_{i+1})q_{i+1} - (p_i + 3p_{i+1})q_i}{4(q_{i+1}-q_i)} = \qquad (146)$$

$$\delta_i \frac{(3p_i + p_{i+1})q_{i+1} - (p_i + 3p_{i+1})q_i}{4(q_{i+1}-q_i)} \qquad (147)$$

Finally, putting it all together you get

$$\begin{aligned}
&-\delta_i \int_0^1 ((1-t)p_i + tp_{i+1}) \log((1-t)q_i + tq_{i+1}) dt = \\
&-\delta_i \frac{p_i \log(q_i) + p_{i+1}\log(q_{i+1})}{2} + \delta_i \frac{p_{i+1} q_i^2 - p_i q_{i+1}^2}{2(q_{i+1}-q_i)^2} (\log(q_{i+1}) - \log(q_i)) + \delta_i \frac{(3p_i + p_{i+1})q_{i+1} - (p_i + 3p_{i+1})q_i}{4(q_{i+1}-q_i)} \qquad (148)
\end{aligned}$$

where only the first term is needed if $q_{i+1} = q_i$. This works, as verified by comparison to numerical integration. Need to switch evaluation strategy when too close to singularity.

A.1 Behaviour around singularity

Lets demonstrate that it's mathematically, if not computationally, stable. To ignore the second and third term when $q_{i+1} = q_i$ need to show that

$$\lim_{q_{i+1} \to q_i} \left[\delta_i \frac{p_{i+1} q_i^2 - p_i q_{i+1}^2}{2(q_{i+1}-q_i)^2} (\log(q_{i+1}) - \log(q_i)) + \delta_i \frac{(3p_i + p_{i+1})q_{i+1} - (p_i + 3p_{i+1})q_i}{4(q_{i+1}-q_i)} \right] = 0. \qquad (149)$$

This is not the case if the two terms are considered independently — it is only the case if they are merged, despite the lovely aesthetics of them as separate terms. To show this limit start by converting the $\log(\cdot)$ of the second term into an infinite sequence, again (Yay!),

$$2\delta_i \frac{p_{i+1}q_i^2 - p_i q_{i+1}^2}{2(q_{i+1}-q_i)^2} \sum_{n\in\{1,3,5,\ldots\}} \frac{1}{n}\left(\frac{\frac{q_{i+1}}{q_i}-1}{\frac{q_{i+1}}{q_i}+1}\right)^n = \tag{150}$$

$$\delta_i \frac{p_{i+1}q_i^2 - p_i q_{i+1}^2}{(q_{i+1}-q_i)^2} \sum_{n\in\{1,3,5,\ldots\}} \frac{1}{n}\frac{(q_{i+1}-q_i)^n}{(q_{i+1}+q_i)^n} = \tag{151}$$

$$\delta_i (p_{i+1}q_i^2 - p_i q_{i+1}^2) \sum_{n\in\{1,3,5,\ldots\}} \frac{1}{n}\frac{(q_{i+1}-q_i)^{n-2}}{(q_{i+1}+q_i)^n} \tag{152}$$

Note that the limit is trivially true for $n = 3$ onwards, so you only need to consider the $n = 1$ case, which can be dropped into the rest of the limit statement

$$\delta_i (p_{i+1}q_i^2 - p_i q_{i+1}^2)\frac{(q_{i+1}-q_i)^{-1}}{(q_{i+1}+q_i)} + \delta_i \frac{(3p_i+p_{i+1})q_{i+1} - (p_i+3p_{i+1})q_i}{4(q_{i+1}-q_i)} = \tag{153}$$

$$\delta_i \frac{4p_{i+1}q_i^2 - 4p_i q_{i+1}^2 + (3p_i+p_{i+1})q_{i+1}(q_{i+1}+q_i) - (p_i+3p_{i+1})q_i(q_{i+1}+q_i)}{4(q_{i+1}-q_i)(q_{i+1}+q_i)} = \tag{154}$$

$$\delta_i \frac{4p_{i+1}q_i^2 - 4p_i q_{i+1}^2 + (3p_i+p_{i+1})q_{i+1}^2 + (2p_i-2p_{i+1})q_i q_{i+1} - (p_i+3p_{i+1})q_i^2}{4(q_{i+1}-q_i)(q_{i+1}+q_i)} = \tag{155}$$

$$\delta_i \frac{(p_{i+1}-p_i)q_{i+1}^2 - 2(p_{i+1}-p_i)q_i q_{i+1} + (p_{i+1}-p_i)q_i^2}{4(q_{i+1}-q_i)(q_{i+1}+q_i)} = \tag{156}$$

$$\delta_i \frac{(p_{i+1}-p_i)(q_{i+1}-q_i)^2}{4(q_{i+1}-q_i)(q_{i+1}+q_i)} = \tag{157}$$

$$\delta_i \frac{(p_{i+1}-p_i)(q_{i+1}-q_i)}{4(q_{i+1}+q_i)}, \tag{158}$$

for which the required limit is simply true. The above is suggestive of an alternate way to write the equation out, so lets consider the unused terms from the second equation,

$$\delta_i (p_{i+1}q_i^2 - p_i q_{i+1}^2) \sum_{n\in\{3,5,7,\ldots\}} \frac{1}{n}\frac{(q_{i+1}-q_i)^{n-2}}{(q_{i+1}+q_i)^n} = \tag{159}$$

$$\delta_i (p_{i+1}q_i^2 - p_i q_{i+1}^2) \sum_{n\in\{1,3,5,\ldots\}} \frac{1}{n+2}\frac{(q_{i+1}-q_i)^{n}}{(q_{i+1}+q_i)^{n+2}} = \tag{160}$$

$$\delta_i \frac{p_{i+1}q_i^2 - p_i q_{i+1}^2}{(q_{i+1}+q_i)^2} \sum_{n\in\{1,3,5,\ldots\}} \frac{1}{n+2}\left(\frac{q_{i+1}-q_i}{q_{i+1}+q_i}\right)^n \tag{161}$$

which appears to be the end of the line. But writing it all together gets

$$-\delta_i \int_0^1 ((1-t)p_i + tp_{i+1})\log((1-t)q_i + tq_{i+1})dt =$$
$$-\delta_i \frac{p_i\log(q_i) + p_{i+1}\log(q_{i+1})}{2} + \delta_i \frac{(p_{i+1}-p_i)(q_{i+1}-q_i)}{4(q_{i+1}+q_i)} + \delta_i \frac{p_{i+1}q_i^2 - p_i q_{i+1}^2}{(q_{i+1}+q_i)^2} \sum_{n\in\{1,3,5,\ldots\}} \frac{1}{n+2}\left(\frac{q_{i+1}-q_i}{q_{i+1}+q_i}\right)^n \tag{162}$$

which is a good fallback when in the $q_i \approxeq q_{i+1}$ condition as the terms of the infinite series move towards zero quickly in typical usage (see main text for quantification).

A.2 Hulk smash fraction of minor irritation

This is super unnecessary, but to demonstrate maximum paranoia it's good to be absolutely certain and reprove results from *The On-Line Encyclopedia of Integer Sequences*. Firstly[7],

$$\frac{1}{n(n+1)(n+2)} = \frac{n(n+3)}{4(n+1)(n+2)} - \frac{(n-1)(n+2)}{4n(n+1)} \tag{163}$$

$$4 = n^2(n+3) - (n-1)(n+2)^2 \tag{164}$$

$$4 = n^3 + 3n^2 - (n-1)(n^2+4n+4) \tag{165}$$

$$4 = n^3 + 3n^2 - n^3 - 4n^2 - 4n + n^2 + 4n + 4 \tag{166}$$

$$4 = 4 \tag{167}$$

Then[8]

$$\frac{1}{n(n+1)} = \frac{n}{n+1} - \frac{n-1}{n} \tag{168}$$

$$1 = n^2 - (n-1)(n+1) \tag{169}$$

$$1 = n^2 - n^2 + 1 \tag{170}$$

$$1 = 1 \tag{171}$$

Combine the above to get

$$\frac{1}{n(n+1)(n+2)} = \frac{n(n+1)(n+3)}{4(n+2)} - \frac{n^2(n+3)}{4(n+1)} - \frac{(n-1)n(n+2)}{4(n+1)} + \frac{(n-1)^2(n+2)}{4n}, \tag{172}$$

$$4 = (n-1)^2(n+1)(n+2)^2 - n^3(n+2)(n+3) - (n-1)n^2(n+2)^2 + n^2(n+1)^2(n+3), \tag{173}$$

$$4 = (n^2-2n+1)(n+1)(n^2+4n+4) - n^3(n^2+5n+6) - (n-1)n^2(n^2+4n+4) + n^2(n^2+2n+1)(n+3), \tag{174}$$

$$4 = (n^2-2n+1)(n^3+5n^2+8n+4) - (n^5+5n^4+6n^3) - (n-1)(n^4+4n^3+4n^2) + n^2(n^3+5n^2+7n+3), \tag{175}$$

$$4 = (n^5+3n^4-n^3-7n^2+4) - (n^5+5n^4+6n^3) - (n^5+3n^4-4n^2) + (n^5+5n^4+7n^3+3n^2), \tag{176}$$

$$4 = n^5+3n^4-n^3-7n^2+4-n^5-5n^4-6n^3-n^5-3n^4+4n^2+n^5+5n^4+7n^3+3n^2, \tag{177}$$

$$4 = (1-1-1+1)n^5 + (3-5-3+5)n^4 + (-1-6+7)n^3 + (-7+4+3)n^2 + 4, \tag{178}$$

$$4 = 4, \tag{179}$$

which can also be written as

$$\frac{1}{n(n+1)(n+2)} = \frac{(n-1)^2(n+2)}{4n} - \frac{n^2(n+3)+(n-1)n(n+2)}{4(n+1)} + \frac{n(n+1)(n+3)}{4(n+2)}. \tag{180}$$

[7] https://oeis.org/A007531
[8] https://oeis.org/A002378