A APPENDIX: PROOFS

Proof of Theorem 1. That $p'_r$ can be computed using the claimed number of evaluations of $p$ follows immediately from Definition 5. Further, the approximation claim holds trivially for $r \leq g_r$ since $p'$ and $p$ are identical in this regime. In what follows, we argue that for any $r > g_r$, $p'_r(r)$, which is $p(g_j)$ for $j = \lceil \log_{1 + \epsilon} r \rceil$, is a $(1 + \epsilon)$-approximation of $p(r)$. By definition:

$$p(r) = \frac{p(g_j)}{g_j + \sum_{i=g_j+1}^r v(t_i)}$$

Since each $v(t_i) \geq 0$, we have:

$$p(r) \geq \frac{p(g_j) g_j}{r} \geq p(g_j) \frac{1 + \epsilon}{1 + \epsilon}$$

where the second inequality can be verified by plugging in the definition of $g_j$ for $j$ as defined above. Thus, $p(g_j)$ is no more than $1 + \epsilon$ times $p(r)$.

For the other direction, we rely on the weak monotonicity assumption A.1. In particular, this implies $p(r) \leq p(g_j)$ whenever $r \geq g_j + m$, finishing the proof for such $r$. When $r < g_j + m$, however, $p(r)$ may be larger than $p(g_j)$. Nevertheless, we use $p(g_j + m) \leq p(g_j)$ and the definition of $p$ to argue that $p(r)$ can be no more than $1 + \epsilon$ times $p(g_j)$ when $r < g_j + m$. To this end, consider the sequence of true labels $v(t_i) \in \{0, 1\}$ for $i = g_j + 1, g_j + 2, \ldots, g_j + m$. Since $p(g_j + m) \leq p(g_j)$, it follows from the definition of $q$ that at most a $p(g_j)$ fraction of these $m$ labels is $1$. Under these conditions, it can be verified that the largest possible value of $p(r)$ occurs at $r = r^* = g_j + \lfloor p(g_j) m \rfloor$ when all true labels $v(t_i)$ for $i = g_j + 1, \ldots, r^*$ are $1$. In this case:

$$\frac{p(r^*)}{p(g_j)} \leq \frac{p(g_j) g_j + \lfloor p(g_j) m \rfloor}{g_j + \lfloor p(g_j) m \rfloor} \leq \frac{g_j + m}{g_j + p(g_j) m} \leq 1 + \frac{m}{g_j} \leq 1 + \frac{1 + \epsilon}{g_j} \leq 1 + \epsilon$$

Plugging these bounds back into the summation over $j$, we obtain:

$$g_k p(g_k) \geq g_t p(g_t) + \sum_{j=t}^{k-1} (g_{j+1} - g_j) p\Delta(g_{j+1}) \geq Y_-(\epsilon, \ell, k, \Delta)$$

This finishes the proof.

Proof of Lemma 1. At each point $g_k$, we have $s$ samples in the set $X_k$. Applying Eqn. (2), it follows that $q(g_k) = S_k/s$, which is the average of the true labels of the $s$ samples in $X_k$, provides a $\beta$-approximation of $p(g_k)$. ■

Proof of Theorem 2. By construction, $q(r) = p(r)$ in Algorithm 1 when $r \in \{1, \ldots, \ell\}$. From Lemma 1, we further have $q(r)$ is a $\beta$-approximation of $p(r)$ for $r \in \{g_{\ell+1}, \ldots, g_L\}$. It now follows from Corollary 1 that step $p'_r$ is a $(1 + \epsilon)$-approximation of $p$.

For the number of annotated samples used in the whole process, we have $\ell$ evaluations of $v$ at points $1, \ldots, g_{\ell}$ for computing $q(1), \ldots, q(g_{\ell})$, which subsumes the $s$ samples needed for $X_L$. For each of $X_{\ell+1}, \ldots, X_L$ in subsequent steps, on average $\frac{1}{1+\epsilon} s$ new samples are drawn. This yields evaluations of $v$ at a total of $g_{\ell}$ initial points and at $(L-\ell) \frac{1}{1+\epsilon} s = \frac{\epsilon}{(L-\ell)} \frac{2(1+\epsilon)^{\ell}}{\ln(1+\epsilon)^{\ell}} \ln \frac{\beta}{\ell}$ points chosen via stratified sampling. ■

Proof of Lemma 3. By the definition of $p(g_k)$, we have:

$$g_k p(g_k) = \sum_{i=1}^{g_k} v(t_i)$$

As observed earlier, $g_{j+1} - g_j \geq (1 + \epsilon)^{j+1} - (1 + \epsilon)^j - 1 = (1 + \epsilon)^j - 1$. When $j \geq \ell$, this implies $g_{j+1} - g_j \geq (1 + \epsilon)^\ell - 1 \geq m$. From the strong monotonicity assumption A.2, the inner summation may be bounded as:

$$p\Delta(g_{j+1}) \leq \sum_{i=g_{j+1}}^{g_{j+1}+1} v(t_i) \leq p\Delta(g_j)$$

This finishes the proof.

Proof of Lemma 4. Using the same argument as in the proof of Lemma 3, for any $j \geq \ell$, we have $g_{j+1} - g_j \geq (1 + \epsilon)^\ell - 1 \geq m$. From the strong monotonicity assumption A.2, the inner summation may be bounded as:

$$p\Delta(g_{j+1}) \leq \sum_{i=g_{j+1}}^{g_{j+1}+1} v(t_i) \leq p\Delta(g_j)$$

This finishes the proof.
by the strong monotonicity assumption A.2, we have \( p_\Delta(g_{j+1}) \leq p_\Delta(g_j) \). However, in general, \( p_\Delta(g_{j+1}) \) may be arbitrarily smaller than \( p_\Delta(g_j) \), making it difficult to lower bound the ratio of \( L \) and \( U \) using the ratio of these local precision terms.

Similar to Ermon et al. (2013), we instead collect all “coefficients” of \( p_\Delta(g_j) \) for each \( j \in \{\ell, \ell + 1, \ldots, k\} \) in the expressions for \( L \) and \( U \), resp., and bound each pair of corresponding coefficients. The computations here are more intricate than in prior work because of the non-integrality of \( \epsilon \); the previous derivation was for the special case where \( \epsilon \) is effectively 1.

Using the assumption \( p(g_{\ell}) \geq p_\Delta(g_{\ell}) \), it can be verified that for the claim it suffices to show two properties:

\[
\gamma g_{\ell} \geq g_{\ell+1} \tag{7}
\]
\[
\gamma (g_{j+1} - g_j) \geq g_{j+2} - g_{j+1} \tag{8}
\]

for all \( j \in \{\ell + 1, \ldots, k - 1\} \). For the first property:

\[
\frac{g_{\ell+1}}{g_{\ell}} = \frac{(1 + \epsilon)^{\ell+1}}{(1 + \epsilon)^{\ell}} \leq \frac{(1 + \epsilon)^{\ell+1} + 1}{(1 + \epsilon)^{\ell} - 1} = 1 + \epsilon + \frac{2 + \epsilon}{(1 + \epsilon)^{\ell} - 1} \leq 1 + \epsilon + \frac{2 + \epsilon}{\epsilon(1 + \epsilon)^{\ell} - 1} \leq 1 + \epsilon + \frac{2 + \epsilon}{m} = \gamma
\]

where the second inequality follows from the precondition \( \epsilon \in (0, 1] \). For the second property:

\[
\frac{g_{j+2} - g_{j+1}}{g_{j+1} - g_j} = \frac{(1 + \epsilon)^{j+2} - [(1 + \epsilon)^{j+1}]}{(1 + \epsilon)^{j+1} - (1 + \epsilon)^{j+1} + 1} \leq \frac{(1 + \epsilon)^{j+2} - (1 + \epsilon)^{j+1} + 1}{(1 + \epsilon)^{j+1} - (1 + \epsilon)^{j+1} - 1} \leq 1 + \epsilon + \frac{2 + \epsilon}{(1 + \epsilon)^{\ell} - 1} \leq 1 + \epsilon + \frac{2 + \epsilon}{m} = \gamma
\]

The second inequality follows from the observation that the ratio under consideration here is a decreasing function of \( j \). This finishes the proof.

**Proof of Theorem 3.** By construction, \( q^-(r) = q^+(r) = p(r) \) in Algorithm 2 when \( r \in \{1, \ldots, \ell\} \). From Lemma 2, we further have \( q^+(r) \) and \( q^-(r) \) are \( \gamma \)-approximations of \( p(r) \) from below and above, resp., for \( r \in \{g_{\ell+1}, \ldots, g_L\} \). It now follows from Corollary 1 that \( \text{step}^-_{\epsilon,\ell} \) and \( \text{step}^+_{\epsilon,\ell} \) are \( (1 + \epsilon) \)-approximations of \( p \) from below and above, resp.

That PAULA uses a total of \( g_{\ell} + \Delta(L - \ell) \) evaluations of \( v \) follows from the observation that it computes \( p \) exactly at points \( 1, \ldots, g_{\ell} \), which in total require \( g_{\ell} \) evaluations, and \( p_\Delta \) at \( L - \ell \) additional points, each of which requires \( \Delta \) evaluations. \( \blacksquare \)