

Dokan Development Log

Alexander Huss

November 18, 2024

Contents

1	Production runtime estimation	1
2	Optimal distribution of resources	2

1 Production runtime estimation

We begin with a single low-statistics ($n_0 = \text{ncall_start}$) job to get a first rough estimate of the runtime (t_0). This is followed by a second job that uses this initial estimate to target a production run with `min_runtime` (if this is not supplied by the user, we may default it to 75% of `max_runtime` or similar): $n_1 = (n_0/t_0) \cdot t_{\min}$. The actual runtime t_1 should lie in the ballpark of t_{\min} but will in practice typically lie below due to the larger relative overhead in the very low-statistics jobs.

Once we have at least those two production jobs complete, we will want to find a dynamic way of estimating the runtime. To this end, we'll determine the time per event $\hat{t}_i \equiv t_i/n_i$ for each completed run and find the mean and error on that quantity. Given that the short runs will in general give time estimates that are less accurate (overheads, etc.), we will *weight* \hat{t}_i in such a way that the longer-running jobs contribute more. We could use either n_i or t_i but it seems more convenient to use n_i . The weighted average and the sample variance is then given by

$$\langle \hat{t} \rangle = \frac{\sum_i \hat{t}_i n_i}{\sum_j n_j} = \frac{\sum_i t_i}{\sum_j n_j}, \quad \langle \hat{t}^2 \rangle = \frac{\sum_i (\hat{t}_i)^2 n_i}{\sum_j n_j} = \frac{\sum_i (t_i)^2 / n_i}{\sum_j n_j}, \quad (1)$$

$$\sigma_{\hat{t}}^2 = \frac{\sum_i (\hat{t}_i - \langle \hat{t} \rangle)^2 n_i}{\sum_j n_j} = \langle \hat{t}^2 \rangle - \langle \hat{t} \rangle^2 = \frac{\sum_i ((t_i)^2 / n_i - t_i^2)}{\sum_j n_j}. \quad (2)$$

The two terms in the parenthesis under the sum of the numerator, will be accumulated while we iterate over each production job (`sumt2` and `sumt`, resp.). In an intermediate implementation we also had a buffer for the `sumt2` accumulator, $\text{sumt2} = \sum_i [(t_i \cdot 0.01)^2 + (t_i)^2/n_i]$, which can be used to inflate the error a bit? We dropped it now for our implementation. Instead, we will use the sample variance ($\sigma_{\hat{t}}$) instead of the variance on the mean ($\sigma_{\langle \hat{t} \rangle}^2 = \sigma_{\hat{t}}^2 / \sum_j n_j$), to be more conservative.

We will aim to hit `max_runtime` with each job by setting $n_{i+1} = t_{\max} / (\langle \hat{t} \rangle + 5\sigma_{\hat{t}})$, leaving a 5σ buffer. Note that in reality the formulae should account for the fact that each job can in principle run different number of iterations, $n_i = n_i^{\text{ncall}} \times n_i^{\text{niter}}$, which is straightforward.

2 Optimal distribution of resources

We assume that we have some initial trial runs at our disposal to estimate uncertainties for the different channels. The total uncertainty is obtained from statistically independent channels

$$\delta\sigma_{\text{tot}} = \sqrt{\sum_{i \in \text{chan}} (\delta\sigma_i)^2} \quad (3)$$

For deriving the optimal distribution of compute resources into the different channels, we *assume* that we have reached the asymptotic Monte Carlo scaling for the errors ($\delta\sigma_i \sim 1/\sqrt{N_i}$), where the total number of “events” invested so far into channel i is denoted by N_i . From the previous (trial-) runs we can also determine the average time per event, $\tau_i \equiv T_i/N_i$.

Let’s denote the new set of events that is going to be invested into channel i as N'_i ; we can then determine the estimated uncertainty on that channel as

$$\delta\sigma'_i = \frac{\sqrt{N_i} \delta\sigma_i}{\sqrt{N_i + N'_i}} \quad (4)$$

with a *runtime* investment of $T'_i = \tau_i N'_i$. For the expected total uncertainty we therefor obtain

$$(\delta\sigma'_{\text{tot}})^2 = \sum_{i \in \text{chan}} (\delta\sigma'_i)^2 = \sum_{i \in \text{chan}} (\delta\sigma_i)^2 \frac{N_i}{N_i + N'_i} \quad (5)$$

This will be the quantity we wish to *minimize* with the *boundary condition* that the total time T'_{tot} we can invest is fixed by how we choose to distribute

our resources into the different channels

$$T'_{\text{tot}} = \sum_{i \in \text{chan}} T'_i = \sum_{i \in \text{chan}} N'_i \tau_i \quad (6)$$

We therefore get the Lagrange function as a function of the event-count distribution

$$\mathcal{L}(\{N'_i\}) = \sum_{i \in \text{chan}} (\delta\sigma_i)^2 \frac{N_i}{N_i + N'_i} - \lambda \left[T_{\text{tot}} - \sum_{i \in \text{chan}} \tau_i N'_i \right] \quad (7)$$

and the Lagrange multiplier λ .

The optimal distribution is obtained from the Euler-Lagrange functions

$$\frac{\partial \mathcal{L}}{\partial N'_i} = 0 \quad \Rightarrow \quad (\delta\sigma_i)^2 \frac{N_i}{(N_i + N'_i)^2} = \lambda \tau_i \quad (8)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \quad \Rightarrow \quad T_{\text{tot}} = \sum_{i \in \text{chan}} \tau_i N'_i \quad (9)$$

where we obtain an intermediate condition $\lambda > 0$ and the optimal distribution is found to be

$$N'_i = \delta\sigma_i \sqrt{\frac{N_i}{\tau_i} \frac{T_{\text{tot}} + T'_{\text{tot}}}{\sum_{i \in \text{chan}} \delta\sigma_i \sqrt{T_i}}} - N_i \quad (10)$$