Reducing Poisson error can offset classification error: a technique to meet clinical performance requirements

Charles B. Delahunt, Courosh Mehanian, Matthew P. Horning
CHARLES.DELAHUNT@GHLABS.ORG
Global Health Labs, Bellevue, WA

Abstract

Medical machine learning algorithms are typically evaluated based on their object-level accuracy vs. that of skilled clinicians, a challenging bar since trained clinicians are usually better classifiers than ML models. However, this metric does not fully capture the realities and requirements of the actual clinical task: it neglects the fact that humans, even with perfect object-level accuracy, are subject to non-trivial error from the Poisson statistics of rare events, because clinical protocols often specify a remarkably small sample size due to the exigencies of clinical work. For example, to quantitate malaria on a thin blood film a clinician examines only 2000 red blood cells (0.0004 μL), which can yield large Poisson variation in the actual number of parasites present, so that a perfect human's count can differ substantially from the true average load. In contrast, an ML system may be less accurate on an object detection level, but it may also have the option to examine much more blood (e.g. 0.1 μL, or 250×). Thus, while its parasite identification error is higher, the Poisson variability of its estimate is lower due to larger sample size. For both ML systems and humans, clinical performance depends on a combination of these two types of error.

To qualify for clinical deployment, an ML system's performance must match current standard of care, typically a demanding target. To achieve this, it may be possible to offset a system's imperfect accuracy by increasing its sample size to reduce Poisson error, and thus attain the same net clinical performance as a perfectly accurate human limited by protocols with smaller sample size.

In this paper, we analyse the mathematics of the relationship between Poisson error, classification error, and total error. This mathematical approach enables teams (software and hardware) optimizing ML systems to leverage a relative strength (larger sample sizes) to offset a relative weakness (classification accuracy). We illustrate the methods with two concrete examples: diagnosis and quantitation of malaria on blood films.

Keywords: Poisson error, malaria, quantitation

Data and Code Availability This work does not use any particular datasets. Python code is provided in the Appendix.

Institutional Review Board (IRB) This work does not required IRB approval.

1. Introduction

For ML solutions to be clinically useful, their performance must match or exceed current clinical requirements. During proof-of-concept, we typically evaluate ML models against a human defined ground truth, and judge its performance against human accuracy. However, this is not the only relevant axis of performance at the clinical task. Humans and ML have different strengths and weaknesses, and ML's strengths can be leveraged to offset its weaknesses to meet a clinical performance goal. Humans are highly adaptable and accurate at tasks like identifying malaria parasites, but they examine only a limited sample size due to time and fatigue constraints. These limited sample sizes are typically encoded in clinical protocols such as for how much blood to examine for malaria quantitation, or how many white blood cells (WBCs) to inventory for a differential blood count. In contrast, machines often have lower accuracy at tasks such as identifying malaria parasites, but they don't fatigue and can potentially examine a much larger sample. Examples of ML systems with test time sample size much larger than standard protocols are Das et al. (2022); Rees-Channer et al. (2023); Torres et al. (2018). In the

Torres study, the automated malaria diagnosis system examined markedly different volumes of blood at the two clinics, and had markedly better performance when it examined more blood.

In this paper we examine the relationship between Poisson error (i.e. the variation in counts of rare objects in a small sample), accuracy error (object classification error), and total error (deviation from the expected value in a very large sample). Limited sample sizes encoded in clinical protocols mean that humans, even with perfect classification accuracy, are subject to often substantial Poisson error relative to the background ground truth (e.g. average parasites per μL of blood). We study how an automated system can, in order to offset imperfect classifier accuracy, examine a larger sample and thus reduce its Poisson error, resulting in equal net error relative to a human following a clinical protocol. That is, machines can use a particular strength (high data throughput) to offset a weakness (classification inaccuracy) and thus match the performance on a clinical task of a skilled human who has high classification accuracy but lower data throughput.

We emphasize that increasing test time sample size does not affect a model's object-level accuracy. But at the patient level (i.e., the clinical task) object classification and Poisson statistics contribute distinct forms of error which roughly sum to give total error. So reducing Poisson error does enable a margin for higher object classification error while maintaining the same total error on the clinical task. In practice, ML teams optimize classification accuracy as much as possible. Then larger examined volume at test time can compensate for not-quite-sufficient classification accuracy in order to meet clinical performance requirements.

We assume the clinicians have perfect (100%) accuracy. Though not all field microscopy has perfect accuracy, well-trained and equipped microscopists are very, very skilled, and any ML system will typically need to match the "best case" human performance. In particular, absent rigorous studies regulatory agencies will likely assume a best case human performance. However, they are still subject to unavoidable Poisson error when a clinical protocol specifies that they examine a relatively small amount of substrate, e.g. 200 WBCs for a differential blood count, or 0.0625 μL of blood for malaria diagnosis. In this paper we describe how to calculate what increase in examined substrate volume is required to offset imperfect ML accuracy, such that the ML system does not exceed the total error allowed by current standards of care.

We address two widely-relevant use cases: (i) limits of detection, (ii) quantitation accuracy. We illustrate the calculation methods on specific clinical examples of these use cases, namely (i) diagnosis of malaria and (ii) quantitation of malaria parasites, on blood films. We provide equations and step-by-step methods to guide how to modify an operational parameter of an ML system (i.e. "sample size examined") to offset imperfect object classification accuracy.

Our mathematical derivations necessarily assume certain (well-principled) starting formulas for limit of detection (LoD) and quantitation. However, if the exact formulae described here might not apply in the context of a particular clinical task, the principle still holds: that total error on a clinical task involving examination of tissue can be broken into different error types, and weaker performance on the algorithm accuracy axis can be offset with stronger performance on the Poisson error axis. This paper provides a template for modeling the clinically relevant error function, which enables development teams (hardware and software) to best allocate their efforts to meet clinical performance requirements.

Our examples highlight the methods' relevance to parasitic diseases, and they have clear application in low resource regions where microscopy is a central tool. But the methods have utility whenever sample sizes specified by protocols imply non-trivial Poisson variability. "Parasite" can be replaced with "object of interest" (e.g. abnormal cell in cancer biopsy, or WBC in a differential WBC count). This paper is relevant to teams developing ML systems that move beyond the academic proof-of-concept and aspire to deployment in the clinic.

The next two sections describe mathematics for (i) diagnosis and LoD, and (ii) quantitation, in each case illustrated with concrete medical use cases in which some of the variables in the equations are known (e.g. fixed by clinical protocols).

2. Diagnosis and limit of detection

Diagnosis of malaria WHO (2010, 2016d) and neglected tropical diseases (e.g. schistosomiasis, lymphatic filariasis) WHO (2002, 2021) at the low parasitemias near LoD is an important case of rare object detection, in which an examined sample might contain only a few parasites and the exact number present in the sample is subject to Poisson variabil-

ity:

$$P(k\ events) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{1}$$

where k is an integer and λ is the expected number of events in a given interval.

In our case, "events" are parasites, "interval" is the volume of substrate (e.g. blood) examined, and the "expected number" is the parasitemia (e.g. parasites per μL) scaled by the examined volume. For large volumes and/or high parasitemias, this basically matches the binomial distribution, and diagnosis is not an issue (it does affect quantitation, as described in Section 3). At small volumes and low parasitemias, the Poisson distribution squeezes up against the y-axis, giving a larger standard error (i.e. std dev / mean = $\sigma(k)/\mu(k)$), and giving an asymmetrical distribution in which $k = 0$ has nontrivial probability. Python Van Rossum and Drake Jr (1995) code to plot Poisson distributions, used for computational sweeps, is given in Appendix A.2. This low parasitemia regime is relevant to diagnosis at LoD.

2.1. Example: Malaria

We assume that a human can perfectly classify objects in blood as a malaria parasite or artifact. Then, since there are no false positive artifacts to contend with, the LoD = N p/μL is the parasitemia at which the examined sample volume consistently (e.g. 95% of the time) contains at least one parasite .

Let n = number of parasites in the examined volume V; cV = clinically-relevant volume (1 μL for malaria), and N/cV be the parasitemia at LoD. Then we require that

$$P(n \geq 1 \mid V,\ N) \geq 0.95 \tag{2}$$

where the underlying distribution is Poisson.

The WHO guidelines for malaria microscopy WHO (2016b) specify that V contain 500 WBCs, ≈ 0.0625 (i.e. $1/16^{th}$) μL using the standard approximation of 8000 WBCs/μL. Plugging in values for N, we find an LoD of $\approx$50 p/μL (see Figure 2.1). We assume that our ML system has imperfect object-level accuracy, and follow the analysis in Delahunt et al. (2024). For a given patient, let object sensitivity = s and false positive rate f = FPs/cV. Then let $\boldsymbol{S}$ be the vector of object-level sensitivities s over each patient and $\boldsymbol{F}$ be the vector of FP rates f over each patient. We note that f varies from patient to patient and consider $\sigma(\boldsymbol{F})$, the standard deviation of FP rates over the population of patients. For simplicity we neglect the variation in s between patients and use the mean $\mu(\boldsymbol{S})$ (for more nuance on this, see Delahunt et al. (2024)).

We wish to calculate how large an examined volume V the ML system needs to examine, such that it will reliably (e.g. 95% of the time) both identify malaria-negative patients and also detect malaria-positive patients at the LoD of N p/μL. To achieve high patient-level specificity, we set a threshold T on the number of suspected parasites detected by the model, such that for most negative patients the suspected parasite count will come in below threshold. Assuming a Gaussian distribution, we can define T as:

$$T = \Big(\mu(\mathbf{F}) + 1.65\sigma(\mathbf{F})\Big)\ \frac{V}{1\ \mu L} \tag{3}$$

where the scaling term gives us the number of FPs in volume V. Then to achieve high sensitivity on samples with parasitemia at the LoD N p/μL we need the the number of suspected parasites (the sum of true positives TP and false positives FP) to equal or exceed this threshold. This must hold for most patients. So it must hold for a patient with a relatively "clean" sample, i.e. with very few FPs (the bottom end of the FP rate distribution):

$$\#FPs = \Big(\mu(\mathbf{F}) - 1.65\sigma(\mathbf{F})\Big)\ \frac{V}{1\ \mu L} \tag{4}$$

Since the suspected parasite count in V equals $\#TP\ +\ \#FP$, we need

$$\#TP > 3.3\sigma(\mathbf{F})\ \frac{V}{1\ \mu L} \tag{5}$$

Given sensitivity $\mu(\boldsymbol{S})$, we need at least $\#TP/\mu(\boldsymbol{S})$ true parasites present in the examined volume to find sufficient TPs to cross the threshold T (as required by Equation 5). Note that the number of TPs we need to have present in V depends on V, because as V increases so does the FP count spread and thus the threshold T. Let x = this required number of TPs.

To find the examined volume V needed to meet this spec, we can computationally sweep values of V and plug $\{V, x(V), N\}$ into the Poisson distribution to see which V gives $P(k \leq x) < 0.05$ (see Python code in Appendix A.2).

For example, if our model has $\mu(\mathbf{S}) = 0.85$ and $\sigma(\mathbf{F}) = 10/\mu L$, then attaining LoD = 50 p/μL is not feasible. However, an LoD = 70 p/μL can be attained by examining $V = 0.2\ \mu L$ (see Figure 2.1). This volume is larger than specified by WHO protocol, but is attainable for automated hardware.

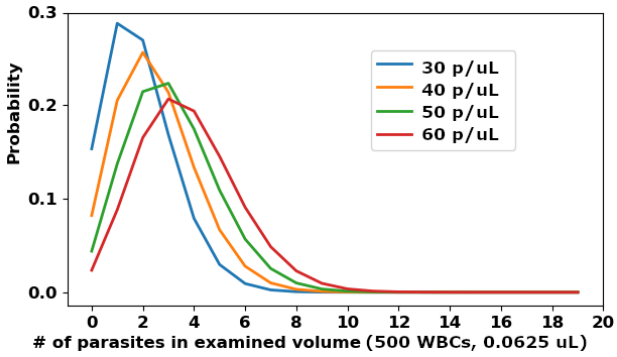

Figure 1: Poisson distributions showing number of parasites actually found in 500 WBCs' worth of blood, given various true parasitemias. The LoD is ≈50 p/μL.

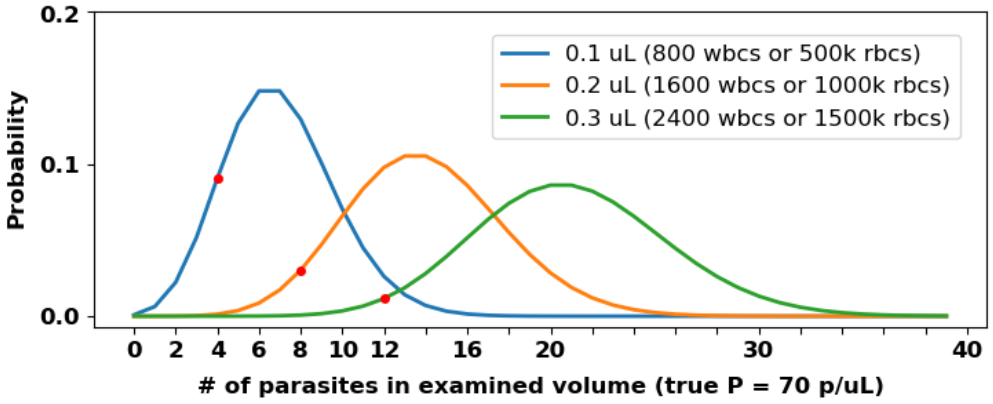

Figure 2: Estimating the volume required to match an LoD = 70, given $\mu(\boldsymbol{S}) = 0.85$ and $\sigma(\boldsymbol{F}) = 10/\mu L$. The required number of TPs (detected by the algorithm) for each volume are shown as red dots. For $V = 0.2$ μL, 95% of cases will exceed the required number.

3. Quantitation

Quantitation is sometimes a clinically relevant task. For example, to monitor drug-resistant malaria strains, sentinel sites dose *P. falciparum*-positive patients with a drug, then create and quantitate blood films every 4 to 6 hours, to see if the strain is developing resistance to the drug Ashley et al. (2014); White (2011); WHO (2016a). This use case involves a huge work burden for microscopists and has relaxed runtime requirements. It is thus an ideal target for automated systems.

Another example concerns mass drug administration of ivermectin to control onchocerciasis (river blindness, a serious Neglected Tropical Disease). In regions where the filaria *Loa loa* is co-endemic, patients with high *Loa loa* microfilaria burdens (25,000 - 30,000 mf/mL of blood) risk catastrophic side effects from ivermectin. Thus accurate quantitation at the 25,000 mf/mL cutoff is clinically vital for the "Test and Not Treat" protocol Gardon et al. (1997); D'Ambrosio et al. (2015). The D'Ambrosio paper discusses the likely impact of sample size on diagnostic accuracy.

Poisson error is introduced into quantitation, even at high parasitemias, when a relatively small sample is examined. Let $\mathcal{P}(p, V)$ = the Poisson distribution for parasitemia p p/cV and examined volume V. Then the standard error caused by Poisson variability at true parasitemia p is

$$\frac{\sigma(\mathcal{P}(p,V))}{\mu(\mathcal{P}(p,V))} = \frac{\sigma(\mathcal{P}(p,V))}{p\frac{V}{cV}} = \sqrt{\frac{1}{p}\frac{cV}{V}} \tag{6}$$

since for a Poisson distribution the variance equals the mean. The denominator is the expected number of parasites in V given parasitemia p. This unavoidable error is remarkably high for the small V_{PR} specified by some protocols.

3.1. Example: quantitation of malaria parasites

To quantitate blood films for malaria, WHO research protocols specify examining 500 WBCS (0.0625 μL) if $p < 16,000p/\mu L$, or 2000 red blood cells (RBCs), ($\approx 0.0004\ \mu L$) if $p > 16,000p/\mu L$ WHO (2016a). We wish to compare the errors in the algorithm quantitation to this Poisson error on V_{PR}. Following Mehanian et al. (2017), we define our formula for estimated parasitemia in a particular sample as

$$\hat{p} = \frac{(tp + fp) - \mu(\boldsymbol{F})\frac{V_E}{cV}}{\mu(\boldsymbol{S})}\ \frac{cV}{V_E} \tag{7}$$

where $tp + fp$ = the number of suspect parasites in examined volume V (both true parasites and misclassified distractors), V_E is the estimate (e.g. found by counting WBCs or RBCs) of V, and other terms are as defined previously. We let $\boldsymbol{V}_E$ = the vector of V_E over the population, in the useful case (below) that all examined volumes are the same V.

To motivate Equation 7, we note that in the first part of the equation (assuming Gaussian distributions of $\boldsymbol{F}$ and $\boldsymbol{S}$) the numerator is the Maximum Likelihood Estimate (MLE) for the number of true parasites detected by the algorithm in V_E, and the denominator is the MLE of the algorithm's sensitivity at parasite detection. So it is the most likely estimate of true parasites actually in V_E.

Equation 7 says: Count up the suspects; then subtract the expected number of FPs in V_E, to get an estimate of detected parasites tp; divide this by our expected sensitivity $\mu(\boldsymbol{S})$ to get an estimate of the actual number of parasites that were present; then normalize by estimated volume to get $\hat{p}/cV$.

This estimate contains three sources of error relative to true parasitemia p: classification (of both parasites and distractors), volume estimation, and Poisson variability. These all vary over the patient population. We seek a formula for the standard deviation of the combined error $sigma(\boldsymbol{p}_E)$ from these three sources.

Let the parasitemia = $p/\mu L$, and p_V = the actual number of parasites in V, so p_V has a Poisson distribution with mean $p\frac{V}{cV}$.
In what follows, Δ denotes the difference of the sample's variable from the mean population value, i.e. $S = \mu(\boldsymbol{S}) + \Delta S$, $F = \mu(\boldsymbol{F}) + \Delta S$, $V_E = V + \Delta V$, and $p_V = p\frac{V}{cV} + \Delta\mathcal{P}(p, V)$ where $\mathcal{P}(p, V)$ denotes Poisson variability. So

$$\begin{aligned} tp &= p_V\ (\mu(\boldsymbol{S}) + \Delta S) \\ &= (p\frac{V}{cV} + \Delta\mathcal{P}(p,V))(\mu(\boldsymbol{S}) + \Delta S) \end{aligned} \tag{8}$$

$$fp = (\mu(\boldsymbol{F}) + \Delta F)\frac{V}{cV} \tag{9}$$

To get a formula for the standard error of quantitation $\frac{\sigma(\boldsymbol{p}_E)}{p}$, we substitute these terms into Equation 7, apply some algebra, convert into standard deviations

over the population, then divide by p. The full derivation is in Appendix A.1:

$$\frac{\sigma(\boldsymbol{p}_E)}{p} = V_{SE} + \frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})}(1+V_{SE}) + \left(1+\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})}\right)\sqrt{\frac{1}{p}\frac{cV}{V}} + \frac{V_{SE}}{p}\frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})} + \frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})}\frac{(1+V_{SE})}{p} \tag{10}$$

where $V_{SE} = \frac{\sigma(\boldsymbol{V}_E)}{V}$ is a constant (see Appendix A.1 for details).

The crucial thing to note about Equation 10 is that the values on the RHS are known: $\mu(\boldsymbol{S}), \sigma(\boldsymbol{S}), \mu(\boldsymbol{F}), \sigma(\boldsymbol{F})$, and V_{SE} are readily-calculated performance statistics of the algorithm (for full details see Delahunt et al. (2024)). So Equation 10 is a simple function of $\{p, V\}$.

Equation 10 has the following structure:

(i) the first 2 terms involve only sensitivity and volume estimation, and are constant for all $\{p, V\}$;

(ii) only the third term involves Poisson variability, and it decreases with $\sqrt{pV}$; and

(iii) only the last 2 terms involve FP rates, and they decrease with p.

If Poisson and volume estimation error are ignored and only classification error is considered, i.e. when $\sigma(\mathcal{P}) = V_{SE} = 0$, then most terms disappear and it becomes the formula for the standard error of quantitation due to classifier inaccuracy as derived in Delahunt et al. (2019):

$$\frac{\sigma(\boldsymbol{p}_E)}{p} = \frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})} + \frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})}\frac{1}{p} \tag{11}$$

Equation 11 is useful in optimizing the classifier algorithm, because it isolates the algorithm's error contributions. But to model quantitation error fully, for example when co-ordinating with a hardware team to determine the required blood volume to examine, $\sigma(\mathcal{P})$ and V_{SE} should be included for two reasons: They are likely to be among the largest contributors to total quantitation error; and $\sigma(\mathcal{P})$ is a piece that is under control of the hardware (since it depends on examined volume).

The contributions of the various terms in Equation 10 are plotted in Figure 3.1. For most parasitemias p, the biggest non-Poisson contributions are from $\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})}$ and V_{SE}, while $\frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})}$, (i.e. the biggest of the terms involving $\sigma(\boldsymbol{F})$) dominates at low p. This plot is a useful tool to highlight where algorithm improvements are most needed.

It also allows us to answer our original question: How much extra volume must we examine to reduce Poisson error enough to offset algorithm errors, such that the automated system's total error matches the Poisson error of a perfect clinician examining a protocol volume V_{PR}?

We can answer this by first plotting the curve $\frac{1}{p}\sigma(\mathcal{P}(p, V_{PR}))$ over values of p, then plotting curves $\frac{1}{p}\sigma(\boldsymbol{p}_E)(p, V)$ from Equation 10 for a set of volumes V. This is illustrated in Figure 4.1 for a (pretty strong) hypothetical algorithm with $\mu(\boldsymbol{S}) = 0.95, \sigma(\boldsymbol{S}) = 0.03, \mu(\boldsymbol{F}) = 50, \sigma(\boldsymbol{F}) = 10$, and $V_{SE} = 0.02$. We find that when applied to V_{PR} the algorithm has far higher standard error than a perfect human following protocol, but examining 0.4 μL suffices to closely match the human's error at most parasitemias, especially if, as is likely, the human volume estimate also has error (see 3.3 for details). A larger examined blood volume can be attained in two ways: (i) by examining more RBCs on thin film (as in Noul (2023)); or (ii) by staying on thick films (as in Mehanian et al. (2017); Das et al. (2022); Rees-Channer et al. (2023)) since machines do not need to switch to thin film at 16,000 p/μL, giving the large advantage shown in Figure 4.1.

3.2. Example: differential white blood cell counts

A common diagnostic is a differential WBC count Doig and Thompson (2017); Gulati et al. (2013); Brihi and Pathak (2024), either 5-part (where each WBC type is inventoried) or 3-part (where only lymphocytes, granulocytes, and monocytes are inventoried by size). Relative numbers and proportions of WBC types can give insight into, for example, whether an infection is bacterial, viral, fungal, or parasitic. When microscopy is used (e.g. in low resource settings) this is a time-consuming task, and thus clinical protocols specify examining only 200 (sometimes 100) WBCs. Since some WBC types represent only 1% of the total count, this protocol implies high Poisson error. Given knowledge of an ML system's WBC classification accuracy (e.g. Fan et al. (2022)), we can calculate how many WBCs it must examine to attain the same total error as a perfect microscopist examining 200 WBCs. To do this, we apply Equation 10 to each WBC type over the expected range of percentages (the analogue of parasitemia), and identify

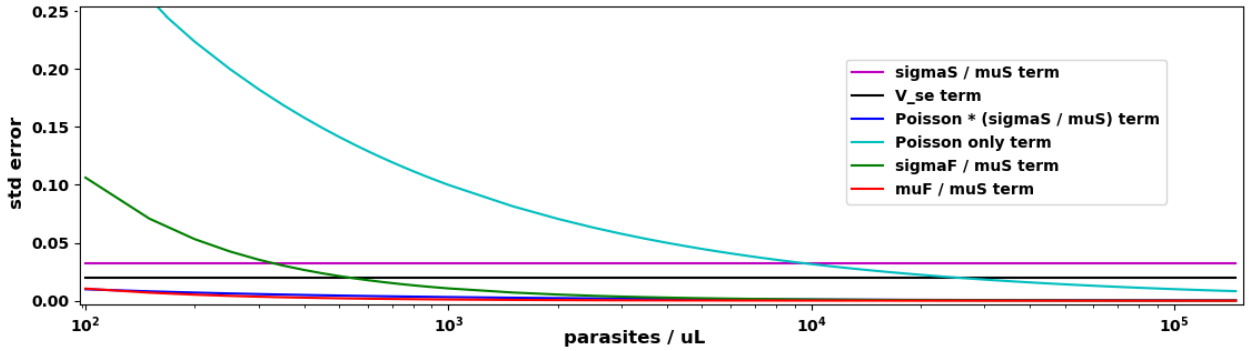

Figure 3: The size of various terms in the standard error of quantitation, Equation 10, vs. parasitemia. At low p errors due to Poisson and $\sigma(\boldsymbol{F})$ dominate, while at high p errors due to $\sigma(\boldsymbol{S})$ and V_{SE} dominate.

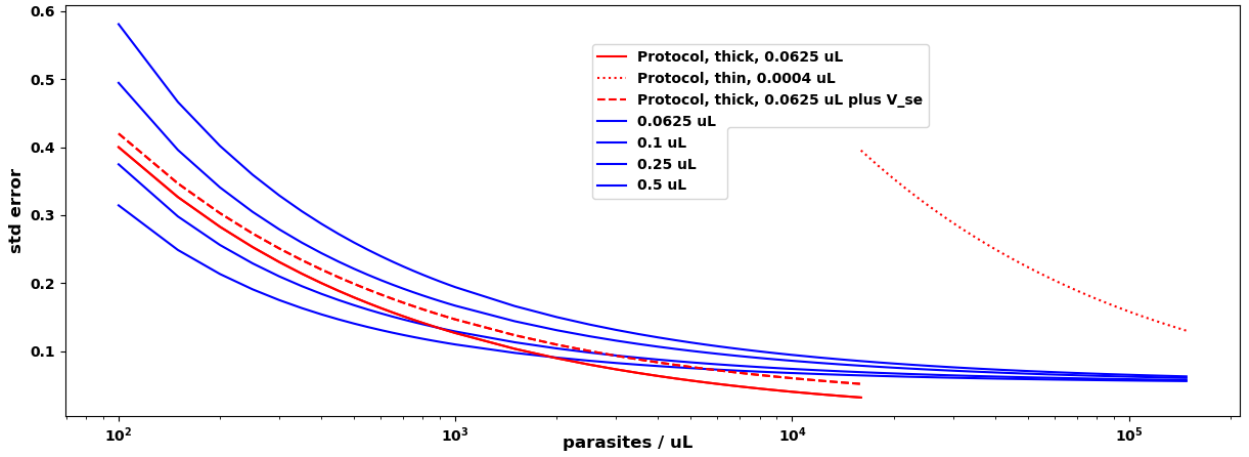

Figure 4: Combined standard error of quantitation vs. true parasitemia, for: (In RED) Humans with perfect accuracy, one without and one with volume estimation error, examining protocol volumes; (In BLUE) An imperfect algorithm with performance per the text, at various examined volumes V. Larger V compensates for its classification and volume errors.

the weakest link, i.e., the cell type that requires the largest sample size to make its total error match that of the microscopist following standard protocol. All terms involving V_{SE} in Equation 10 zero out, since the number of total WBCs inventoried is automatically exact.

3.3. Errors in estimating examined volume

As noted in Delahunt et al. (2024), another source of quantitation error is error in estimating the volume examined. According to WHO protocols, on thick films blood volume is estimated by counting WBCs and using the approximation 8000 WBCs/μL WHO (2016b) (6000/μL in Peru Ministerio de Salud (2003)). It is estimated on thin films by examining microscope fields-of-view, ballparking RBC count per field-of-view, then using the approximation 5e6 RBCs/μL WHO (2016c). In research situations a grid system is sometimes used, where 5 μL of blood is evenly spread across a grid of fixed size, and volume is estimated by area WHO (2016a).

Two details of protocol acknowledge an imprecision in human quantitations for malaria. These details may reflect an expectation of errors in parasite counting and volume estimation in addition to known Poisson error: First, when possible two or more manual quantitations are averaged WWARN (2023). This carries high operational cost however, and is not typical for diagnostic settings. Second, the WHO proficiency standards define someone whose quantitations are within 25% of ground truth at least half the time on a defined set of 15 blood films as having "Level 1" proficiency WHO (2016d).

In the WBC/RBC counting cases, machines have a substantial advantage. Although expert humans are extremely skilled, one can reasonably expect nontrivial error when manually counting hundreds of cells while moving through microscope fields of view and concurrently tallying parasite counts, or when depending on bulk estimates of RBCs per field without careful counting. We are not aware of any studies of human counting error in this context, so we left it out of our analysis above. However, perfect human volume estimation is likely an unrealistic assumption.

4. Discussion

To deploy into clinical settings, an ML system's performance must match current standards of care, typically a challenging requirement. In this paper we have described how standard protocols for human microscopy assessment specify, by necessity, relatively small sample sizes and thus have high levels of unavoidable Poisson variability.

While trained humans are extraordinarily strong and adaptable object classifiers, they are still fully subject to this Poisson error. Automated ML systems may struggle to attain the same object classification accuracy as trained humans, but they excel on other axes of performance. In particular, the clinical deployment of a diagnostic algorithm typically requires, as part of the total system, an automated data acquisition device, e.g. a slide scanner, whose role in the system must also optimized. If the clinical use case and the deployed system allow for increasing the examined sample size, this opens the possibility to reduce the Poisson error associated with the automated system, offsetting an algorithm's higher classification error.

We have also described mathematical methods to analyze the relationship between Poisson error, algorithm error, and total error. These methods can inform the principled use of increased sample size to offset algorithm error and thus achieve performance equivalent to a perfectly accurate human on a clinical task. This approach enables development teams to best allocate effort as to where to reduce error.

If the medical use case is not sufficiently similar to the "count objects in the sample" problem, our paper offers perspective/approach rather than methodology (since the equations will not directly apply). Medical use cases are too varied to admit of a general solution. Sometimes the given equations will apply (e.g. the Diagnosis equations for histopathology to identify and tally aberrant cells in a tissue sample). If they do not apply, then the ML team may need to derive their own equations or heuristic code. For example, the quantitation equations apply directly to differential blood cell counts at a basic level (i.e., quantitating each cell type), but a medical diagnosis is based on ratios of these counts, a further layer of analysis would be ideal. We also note that absent bespoke equations, a legitimate option is to simply try to maximize the volume examined by the hardware, knowing that this will likely help the system achieve clinical performance goals.

A second use of the quantitation equations is to guide algorithm development by highlighting which forms of object-level error are most productive to reduce. For example, the equations might show that

due to low parasitemia ranges being targeted, an algorithm's FP error is the limiting factor.

We note that the ability to examine a larger sample is subject to technical, cost, and use case constraints. For example, scanning more sample takes more time. In the drug resistance monitoring use case, time-to-result is relaxed and allows more scanning time; while point-of-care settings require fast diagnosis, limiting scanning time and thus examined volume. Financial cost may also constrain increases to sample size. For example, Noul (2023) has a one-use cartridge to stain a thin film, so increasing examined volume beyond one cartridge's worth doubles the cost of consumables; while a system scanning standard thick films can examine extra volume at no extra cost. In another example, the one-use capillary in D'Ambrosio et al. (2015) has a fixed blood volume. This suffices for the *Loa loa* "test and not treat" use case but perhaps not for diagnosing the much lower parasitemias in lymphatic filariasis, and increasing volume would require two capillaries per patient, doubling consumable costs.

The ability to drive down Poisson error offers teams that are optimizing automated ML systems for deployment a separate, valuable axis for improvement, an axis which humans cannot realistically leverage. Improvement on this axis, involving both hardware and software, can offset lower ML performance on tasks at which humans excel, such as object classification detached from the constraints of clinical protocols, enabling ML systems to meet the rigorous clinical performance standards of care required for deployment.

Acknowledgments

Funded by Global Health Labs, Inc.
(www.ghlabs.org).

References

E.A. Ashley, M. Dhorda, R.M. Fairhurst, C. Amaratunga, and et al. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *New England J of Medicine*, 2014.

J.E. Brihi and S. Pathak. Normal and abnormal complete blood count with differential. *StatPearls [Internet]*, 2024. URL https://www.ncbi.nlm.nih.gov/books/NBK604207/.

D. Das, R. Vongpromed, M. Dhorda, and et al. Field evaluation of the diagnostic performance of EasyScan GO: a digital malaria microscopy device based on machine-learning. *Malaria J*, 2022.

C. B. Delahunt, M. S. Jaiswal, M. P. Horning, S. Janko, C. M. Thompson, S. Kulhare, and et al. Fully-automated patient-level malaria assessment on field-prepared thin blood film microscopy images, with supplemental information. *IEEE GHTC Proceedings, arXiv version*, 2019. https://arxiv.org/abs/1908.01901.

C.B. Delahunt, N. Gachuhi, and M.P Horning. Metrics to guide development of machine learning algorithms for malaria diagnosis. *in review*, 2024. https://arxiv.org/abs/2209.06947.

K. Doig and L.A. Thompson. A methodical approach to interpreting the white blood cell parameters of the complete blood count. *Clin Lab Sci*, 2017. URL https://hwmaint.clsjournal.ascls.org.

M. V. D'Ambrosio, D.A. Fletcher, and et al. Point-of-care quantification of blood-borne filarial parasites with a mobile phone microscope. *Science Trans Medicine*, 2015.

B.E. Fan, D.T.Y. Chen, S-Y. Hsu, Yang H-H., S.T.B.A Latiff, K.G.E. Lim, Ong Y.X., M.S. Wong, S. Winkler, and K. Ponnudurai. Real-World Evaluation of a Semi-Supervised Artificial-Intelligence Model Trained on 185,412 Cells for Identification of White Blood Cells. *MICCAI-CLINICCAI*, 2022.

J. Gardon, M. Boussinesq, and et al. Serious reactions after mass treatment of onchocerciasis with ivermectin in an area endemic for loa loa infection. *Lancet*, 1997.

G. Gulati, J. Song, A.D. Florea, and J. Gong. Purpose and criteria for blood smear scan, blood smear examination, and blood smear review. *Ann Lab Med.*, 2013. doi: 10.3343/alm.2013.33.1.1.

C. Mehanian, M. Horning, and et al. Computer-automated malaria diagnosis and quantitation using convolutional neural networks. *ICCV*, 2017.

Ministerio de Salud. *Manual de Procedimientos de Laboratoria Para el Diagnostico de Malaria*, 2003. Lima, Peru.

Noul. miLab platform. 2023. S. Korea. https://noul.kr/en/milab-platform/.

R.R. Rees-Channer, C.M. Bachman, P.L. Chiodini, and et al. Evaluation of an automated microscope using machine learning for the detection of malaria in travelers returned to the UK. *Frontiers Malaria*, 2023.

K. Torres, C.M. Bachman, V.D. Gamboa, and et al. Automated microscopy for routine malaria diagnosis: a field comparison on Giemsa-stained blood films in Peru. *Malaria J*, 2018.

Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

N. White. The parasite clearance curve. *Malaria J*, 2011.

WHO. *Prevention and control of schistosomiasis and soil-transmitted helminthiasis*, 2002. World Health Organization, Geneva, Switzerland.

WHO. *Basic malaria microscopy. Part I. Learner's guide. 2nd ed, (esp units 7, 8, and 9)*, 2010. World Health Organization, Geneva, Switzerland.

WHO. *Microscopy for the detection, identification and quantification of malaria parasites on stained thick and thin blood films in research settings, ver 1*, 2016a. World Health Organization, Geneva, Switzerland.

WHO. *Microscopy examination of thick and thin blood films for identification of malaria parasites (esp SOPs 8 and 9)*, 2016b. World Health Organization, Geneva, Switzerland.

WHO. *Malaria Microscopy Standard Operating Procedure MM-SOP-09: Malaria Parasite Counting*, 2016c. World Health Organization, Geneva, Switzerland.

WHO. Malaria microscopy quality assurance manual v2. 2016d. World Health Organization, Geneva, Switzerland.

WHO. *Diagnostic target product profiles for monitoring, evaluation and surveillance of schistosomiasis control programmes*, 2021. World Health Organization, Geneva, Switzerland.

WWARN. Obare method calculator. 2023. https://www.wwarn.org/obare-method-calculator.

Appendix A.

A.1. Derivation of quantitation error

Following Mehanian et al. (2017), suppose our formula for estimated parasitemia, for a sample with true parasitemia p/cV, is:

$$\hat{p} = \frac{(tp + fp) - \mu(\boldsymbol{F})\frac{V_E}{cV}}{\mu(\boldsymbol{S})} \frac{cV}{V_E} \tag{12}$$

where $tp + fp =$ the number of suspect parasites in examined volume V (both true parasites and misclassified distractors),
V_E is the estimate of V (e.g. found by counting WBCs or RBCs),
and other terms are as defined previously.
This formula says: we count the suspects then subtract the number of FPs which we expect in V_E, to leave our estimate of detected parasites; we divide this by our estimate of sensitivity $\mu(\boldsymbol{S})$ to get an estimate of the actual number of parasites that were present; we then normalize by estimated volume to get $\hat{p}/cV$.

This estimate contains three sources of error relative to true parasitemia p: classification (of both parasites and distractors), volume estimation, and Poisson variability. Each of these varies by sample (a particular patient's blood film). We will substitute these errors into Equation 7:

Let $P_V =$ the true number of parasites in V, so P_V has a Poisson distribution with mean $p\frac{V}{cV}$.
In what follows, Δ denotes the deviation of the sample's variable from the mean population value, i.e. $S = \mu(\boldsymbol{S}) + \Delta S$, $F = \mu(\boldsymbol{F}) + \Delta S$, $V_E = V + \Delta V$, and $P_V = p\frac{V}{cV} + \Delta\mathcal{P}(p, V)$ where $\mathcal{P}(p, V)$ denotes Poisson variability. So

$$\begin{aligned} tp &= P_V\ (\mu(\boldsymbol{S}) + \Delta S) \\ &= (p\frac{V}{cV} + \Delta\mathcal{P}(p, V))(\mu(\boldsymbol{S}) + \Delta S) \end{aligned} \tag{13}$$

$$fp = (\mu(\boldsymbol{F}) + \Delta F)\frac{V}{cV} \tag{14}$$

Then substituting Equations 13,14 into Equation 12 gives the following quantitation error for the sample:

$$\begin{aligned} \Delta p = \hat{p} - p = -p\ +\ \hat{p} = -p\ + \\ \left[\left(p\frac{V}{cV} + \Delta\mathcal{P}(p, V)\right)(\mu(\boldsymbol{S}) + \Delta S) + \right. \\ \left. (\mu(\boldsymbol{F}) + \Delta F)\frac{V}{cV} - \mu(\boldsymbol{F})\frac{V_E}{cV}\right]\frac{1}{\mu(\boldsymbol{S})}\frac{cV}{V_E} \end{aligned} \tag{15}$$

Distribute everything:

$$\begin{aligned}\Delta p = -p+ \\ &\left(p\frac{V}{cV}\frac{\mu(\boldsymbol{S})}{\mu(\boldsymbol{S})}\frac{cV}{V_E}\right) + \left(\Delta\mathcal{P}(p,V)\frac{\mu(\boldsymbol{S})}{\mu(\boldsymbol{S})}\frac{cV}{V_E}\right) + \\ &\left(p\frac{V}{cV}\frac{\Delta S}{\mu(\boldsymbol{S})}\frac{cV}{V_E}\right) + \left(\Delta\mathcal{P}(p,V)\frac{\Delta S}{\mu(\boldsymbol{S})}\frac{cV}{V_E}\right) + \\ &\left(\frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})}\frac{V}{cV}\frac{cV}{V_E}\right) + \left(\frac{\Delta F}{\mu(\boldsymbol{S})}\frac{V}{cV}\frac{cV}{V_E}\right) - \frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})}\frac{V_E}{cV}\frac{cV}{V_E}\end{aligned} \tag{16}$$

Cancel terms, and substitute $V = V_E - \Delta V$:

$$\begin{aligned}\Delta p = -p + &\left(p\frac{V_E-\Delta V}{V_E}\right) + \left(\Delta\mathcal{P}(p,V)\frac{cV}{V_E}\right) + \\ &\left(p\frac{V_E-\Delta V}{V_E}\frac{\Delta S}{\mu(\boldsymbol{S})}\right) + \left(\Delta\mathcal{P}(p,V)\frac{\Delta S}{\mu(\boldsymbol{S})}\frac{cV}{V_E}\right) + \\ &\left(\frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})}\frac{V_E-\Delta V}{V_E}\right) + \left(\frac{\Delta F}{\mu(\boldsymbol{S})}\frac{V_E-\Delta V}{V_E}\right) - \frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})}\end{aligned} \tag{17}$$

Continue simplifying:

$$\begin{aligned}\Delta p = -p + p - p\frac{\Delta V}{V_E} + \Delta\mathcal{P}(p,V)\frac{cV}{V_E} + \\ p\frac{\Delta S}{\mu(\boldsymbol{S})} - p\frac{\Delta V}{V_E}\frac{\Delta S}{\mu(\boldsymbol{S})} + \Delta\mathcal{P}(p,V)\frac{\Delta S}{\mu(\boldsymbol{S})}\frac{cV}{V_E} + \\ \frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})} - \frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})}\frac{\Delta V}{V_E} + \frac{\Delta F}{\mu(\boldsymbol{S})} - \frac{\Delta F}{\mu(\boldsymbol{S})}\frac{\Delta V}{V_E} - \frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})}\end{aligned} \tag{18}$$

To assess standard deviation of the quantitation error, $\sigma(\Delta\hat{p})$, over the population, we convert the Δ terms to std devs $\sigma(\ )$.

Note that in the interaction terms, $\Delta V, \Delta\mathcal{P}, \Delta S$, and ΔF are independent (relative to the other interaction term) random variables with zero mean ($\Delta\mathcal{P}$ has very close to zero mean at parasitemias relevant for quantitation, though this does not hold for the very low parasitemias at LoD). So

$$\sigma(\Delta V\Delta S) = \sigma(\Delta V)\sigma(\Delta S), \text{etc.} \tag{19}$$

Following the definitions in Section 2.1, $\sigma(\Delta S) = \sigma(\boldsymbol{S})$ and $\sigma(\Delta S) = \sigma(\boldsymbol{S})$. We denote the vector of ΔVs over the population as $\boldsymbol{V}_E$ and the vector of parasitemia estimate errors as $\boldsymbol{p}_E$.

To facilitate computation later, we make an approximation when converting to std devs at population level:

$$\frac{\sigma(\Delta\,\boldsymbol{V})}{V_E} \approx \frac{\sigma(\Delta\,\boldsymbol{V})}{V} \tag{20}$$

if the volume estimator (e.g. WBC or RBC counter) is decent (e.g. if standard error = 0.1 they differ by a factor < 1.05). Then:

$$\begin{aligned}\sigma(\boldsymbol{p}_E) = \\ &p\frac{\sigma(\boldsymbol{V}_E)}{V} + \sigma(\mathcal{P}(p,V))\frac{cV}{V} + p\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})} + \\ &p\frac{\sigma(\boldsymbol{V}_E)}{V}\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})} + \sigma(\mathcal{P}(p,V))\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})}\frac{cV}{V} + \\ &\frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})}\frac{\sigma(\boldsymbol{V}_E)}{V} + \frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})} + \frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})}\frac{\sigma(\boldsymbol{V}_E)}{V}\end{aligned} \tag{21}$$

To further simplify, we note that $\frac{\sigma(\boldsymbol{V}_E)}{V}$ is most likely constant at usable V: this is the standard error of the volume estimator, and the error in the count of WBCs or RBCs will likely scale with the total true number, e.g. if $\sigma(\boldsymbol{V}_E) = 20$ for $V = 500$, then $\sigma(\boldsymbol{V}_E) = 40$ for $V = 1000$. Letting $\frac{\sigma(\boldsymbol{V}_E)}{V} = V_{SE}$ we have the cleaner formula

$$\begin{aligned}\sigma(\boldsymbol{p}_E) = \\ &pV_{SE} + \sigma(\mathcal{P}(p,V))\frac{cV}{V} + p\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})} + \\ &pV_{SE}\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})} + \sigma(\mathcal{P}(p,V))\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})}\frac{cV}{V} + \\ &\frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})}V_{SE} + \frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})} + \frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})}V_{SE}\end{aligned} \tag{22}$$

Then the standard error of quantitation is

$$\begin{aligned}\frac{\sigma(\boldsymbol{p}_E)}{p} = \\ &V_{SE} + \frac{\sigma(\mathcal{P}(p,V))}{p}\frac{cV}{V} + \frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})} + \\ &V_{SE}\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})} + \frac{\sigma(\mathcal{P}(p,V))}{p}\frac{cV}{V}\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})} + \\ &\frac{V_{SE}}{p}\frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})} + \frac{1}{p}\frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})} + \frac{V_{SE}}{p}\frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})}\end{aligned} \tag{23}$$

Grouping and reordering terms gives

$$\begin{aligned}\frac{\sigma(\boldsymbol{p}_E)}{p} = \ &V_{SE} + \\ &\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})}(1+V_{SE}) + \left(1+\frac{\sigma(\boldsymbol{S})}{\mu(\boldsymbol{S})}\right)\sqrt{\frac{1}{p}\frac{cV}{V}} + \\ &\frac{V_{SE}}{p}\frac{\mu(\boldsymbol{F})}{\mu(\boldsymbol{S})} + \frac{\sigma(\boldsymbol{F})}{\mu(\boldsymbol{S})}\frac{(1+V_{SE})}{p}\end{aligned} \tag{24}$$

where we used the fact that for a Poisson distribution the variance equals the mean, so $\sigma(\mathcal{P}(p,V)) = \sqrt{p\frac{V}{cV}}$.

This is Equation 10 in the main text.

A.2. Python code to plot Poisson distributions

```
import numpy as np
from matplotlib import pyplot as plt
from scipy.stats import poisson

pPerUL = 100
vols = np.array((0.01, 0.02, 0.05, 0.1))
numWbcs = vols * 8000
numRbcs = vols * 5000
mu = pPerUL * vols
numDraws = 10000
k = np.arange(0, 20)

probK = np.zeros((len(k), len(vols)))
cumProbK = np.zeros((len(k), len(vols)))
stdProbK = np.zeros(len(vols))

for i in range(len(vols)):
   for j in range(len(k)):
       probK[j, i] = poisson.pmf(k = k[j], mu=mu[i])
       cumProbK[j,i] = poisson.cdf(k = k[j], mu=mu[i])
       stdProbK[i] = np.std(poisson.rvs(mu = mu[i],size=numDraws ))

stdError = stdProbK / mu

# plot distributions:
print('vols = ' + str(vols))
print('std error = ' + str(np.round(stdError,2)))

tickKwargs = 'fontweight':'bold','fontsize':12
legendKwargs = 'fontsize':12

plt.figure()
plt.xlabel('# of parasites in examined volume (true P = 100 p/uL)',
fontweight='bold', fontsize=12)
plt.ylabel('Probability', fontweight='bold', fontsize=12)
for i in range(len(vols)):
   plt.plot(k, probK[:, i], linewidth=2, label = str(vols[i]) + ' uL (' +
str(int(numWbcs[i])) + ' wbcs or ' +      str(int(numRbcs[i])) + 'k rbcs)')
plt.legend(**legendKwargs)
plt.xticks(range(0,21,2), **tickKwargs)
plt.yticks(np.arange(0, 0.5, 0.1), **tickKwargs)
```

A.3. Python code to plot standard error of quantitation

```
import os
import numpy as np
from matplotlib import pyplot as plt, rc

# parasitemias:
p = list(range(100, 1001, 50)) + list(range(1000, 10000, 500)) +
```

```
list(range(10000, 150000, 2000))
p = np.array(p)
# Volumes examined:
Vrbc = 0.0004 # 2000 RBCs if > 16k p/uL. TDR
Vwbc = 0.0625 # 500 WBCs if p < 16k/uL. TDR.
VwbcWho = 0.025 # 200 WBCs if p > 400, < 16k
Vauto = 0.125 # 1000 WBCs
Vexam = np.array([0.0625, 0.1, 0.25, 0.5]) # volume examined by algorithm

# Algorithm performance statistics:
muS = 0.95
sigmaS = 0.03
muF = 50
sigmaF = 10
Vse = 0.02 # ie std dev = 2% of total count
VseHuman = 0.02

# Various fixed Vs:
PoisVrbc = np.sqrt(1 / (p * Vrbc))
PoisVwbc = np.sqrt(1 / (p * Vwbc))
PoisVauto = np.sqrt(1 / (p * Vauto))

# Populate for range of Vs:
PoisVexam = np.zeros((len(Vexam), len(p)))
PoisAndSigSTerm = np.zeros((len(Vexam), len(p)))
muFTerm = np.zeros((len(Vexam), len(p)))
sigmaFTerm = np.zeros((len(Vexam), len(p)))

for i in range(len(Vexam)):
   PoisVexam[i, :]  = np.sqrt(1 / (p * Vexam[i]))
   PoisAndSigSTerm[i, :]  = (sigmaS / muS) * np.sqrt(1 / (p * Vexam[i]))
   muFTerm[i, :]  = (Vse / p) * (muF / muS)
   sigmaFTerm[i, :]  = sigmaF / muS *(1 + Vse) / p
# Constant:
sigmaSConstantTerm = (sigmaS / muS) * (1 + Vse)

totalStdError = np.zeros((len(Vexam), len(p)))
for i in range(len(Vexam)):
   totalStdError[i, :]  = Vse + sigmaSConstantTerm +
PoisVexam[i,:]  + PoisAndSigSTerm[i, :]  + muFTerm[i, :]  + sigmaFTerm[i, :]

----------------------------------------
#%% Plot Poisson error (only) for select V's including protocols:
plt.figure()
rc('font',weight='bold')
# Vwbc:
inds = np.where(p < 16001)[0]
plt.semilogx(p[inds], PoisVwbc[inds], color='r', label='Protocol, thick, 500
WBCs')
# Vrbc:
inds = np.where(p >= 16000)[0]
```

```
plt.semilogx(p[inds], PoisVrbc[inds], linestyle='--', color='r',
label='Protocol, thin, 2000 RBCs')
# Autoscope:
plt.semilogx(p, PoisVauto, 'b', label='Machine, 1000 WBCs')

plt.legend()
plt.xlabel('parasites / uL', fontweight='bold', fontsize=12)
plt.ylabel('std error', fontweight='bold', fontsize=12)

----------------------------------------
#%% Plot some of the important terms for one Vexam:
ind = np.where(Vexam == 0.1)[0][0]
plt.figure()
rc('font', weight='bold')

plt.semilogx(p, sigmaSConstantTerm *np.ones(len(p)),'m',
label='sigmaS / muS term')
plt.semilogx(p, Vse *np.ones(len(p)),'k', label='Vse term')
plt.semilogx(p,PoisAndSigSTerm[ind,:],'b', label='Poisson * (sigmaS / muS)
term')
plt.semilogx(p,PoisVexam[ind,:],'c', label='Poisson only term')
plt.semilogx(p,sigmaFTerm[ind,:],'g', label='sigmaF / muS term')
plt.semilogx(p,muFTerm[ind,:],'r', label='muF / muS term')

plt.legend()
plt.xlabel('parasites / uL', fontweight='bold', fontsize=12)
plt.ylabel('std error', fontweight='bold', fontsize=12)
plt.title('Components of std error equation for:  '  + 'V = ' +
str(Vexam[ind]) + ', Vse = ' + str(Vse) +
'muS = ' + str(muS) + ', sigmaS = ' + str(sigmaS) + ', muF = ' + str(muF) + ',
sigmaF = ' + str(sigmaF), fontweight='bold')

----------------------------------------
#%% Plot Poisson error for protocol and total error for selection of Vs:
plt.figure()
rc('font',weight='bold')

# Vwbc:
inds = np.where(p < 16001)[0]
plt.semilogx(p[inds], PoisVwbc[inds], color='r',
label='Protocol, thick, ' + str(Vwbc) + ' uL')
inds = np.where(p < 16001)[0]
# Vrbc:
inds = np.where(p >= 16000)[0]
plt.semilogx(p[inds], PoisVrbc[inds], linestyle=':', color='r',
label='Protocol, thin, ' + str(Vrbc) + ' uL')
# Vwbc + some volume estimation error:
inds = np.where(p < 16001)[0]
plt.semilogx(p[inds], PoisVwbc[inds] + VseHuman, color='r', linestyle='--',
label='Protocol, thick, ' + str(Vwbc) + ' uL plus se')
# Machine
```

```
for i in range(len(PoisVexam)):
   plt.semilogx(p, totalStdError[i], 'b', label=str(Vexam[i]) + ' uL')
# Rerun Vwbc to make them foreground:
inds = np.where(p < 16001)[0]
plt.semilogx(p[inds], PoisVwbc[inds], color='r')
plt.semilogx(p[inds], PoisVwbc[inds] + VseHuman, color='r', linestyle='--')
plt.legend()
plt.xlabel('parasites / uL', fontweight='bold', fontsize=12)
plt.ylabel('std error', fontweight='bold', fontsize=12)
```