Poisson Statistics

Francisca Vasconcelos*
MIT Department of Physics
(Dated: October 9, 2019)

The Poisson distribution models the number of events that will occur in a given time interval, given the event rate, λ. The binomial distribution similarly models the number of successes, given the number of trials, n, and probability of success, p. In this work, we create a physical Poisson process with four different event rates. Using the data, we experimentally demonstrate that the Poisson process is the limit of the binomial distribution, for $\lambda = np$ as $n \to \infty$ and $p \to 0$.

I. INTRODUCTION & THEORY

The Poisson distribution was introduced by Simeon Denis Poisson in 1837, in an effort to model the number of wrongful convictions in a given country [1]. This distribution is of particular importance in physics because it can be used to simulate any physical counting process in which event occurrences are, or can be approximated as, independent. This allows physicists to model fairly complex phenomena, such as the number of photons that will hit a photo-detector over the course of a minute, the number of radioactive particles that will decay in an hour, and even the likelihood of an asteroid hitting earth in the next ten years. Furthermore, the Poisson distribution has several nice features, which make it ideal for statistical error analysis.

I.1. Poisson Distribution

Given an event rate, λ time^{-1}, the Poisson discrete probability distribution assigns a likelihood to k events occurring in one unit of time.

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

Time can be defined as seconds, hours, years, etc. Intuitively we would expected the majority of the probability density to fall around the event rate, λ. However, the Poisson distribution additionally accounts for random fluctuations that may cause counts to arrive before or after their ideal arrival time (the further from the ideal arrival time, the less likely). A particularly useful feature of the Poisson distribution is that its mean and variance are both equal to λ.

I.2. Binomial Distribution

The binomial discrete probability distribution is similar to the Poisson distribution in that it assigns a likelihood to k events occurring. However, k is now considered to be the number of successes in n trials, given that the modeled event has a probability of success p.

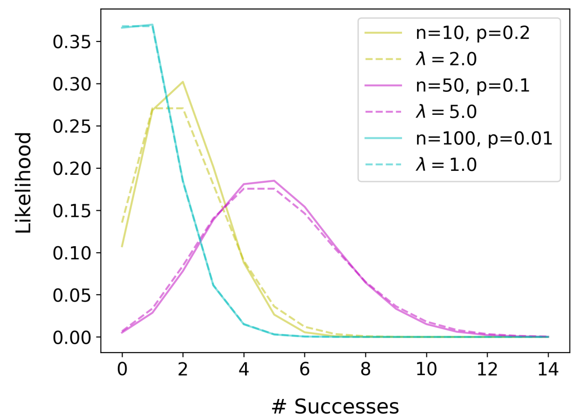

FIG. 1. We demonstrate the Poisson Limit Theorem, with the Poisson distribution (dotted lines) becoming a better approximation of the binomial distribution (solid line) as we increase n and decrease p, for $\lambda = np$.

$$P(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \tag{2}$$

A common example of a binomial process is coin tossing, in which we would assign k=# heads, n=# coin tosses, and p=50% (likelihood of landing heads for a fair coin). For plots of the shapes of this distribution and the Poisson distribution, refer to *Appendix A*.

I.3. Poisson Limit Theorem

The Poisson Limit Theorem, or Law of Rare Events, states that the Poisson distribution can be used to approximate the binomial distribution, in the case that $\lambda = np$ for large n ($n \to \infty$) and small p ($p \to 0$). We demonstrate this convergence graphically in Fig. 1 [2].

II. EXPERIMENTAL SETUP

Abstractly speaking, we aimed to create four physical Poisson processes, with target event rates of: 1, 5, 10,

* francisc@mit.edu

and 100 counts/second. As illustrated in Fig. 2, we performed two different types of measurements on each of these processes:

100-Shots: Given a Poisson source, we performed 100 counting measurements, each of duration 1 second. At the end of each measurement interval the number of counts was recorded and reset.

1-Shot: Given a Poisson source, we performed a single counting measurement of 100 second duration and recorded the total number of counts.

In terms of the actual experimental setup, we used the measurement chain depicted in Fig. 2. A ^{137}Cs gamma ray source was set a variable distance away from an NaI scintillator. The signal from this scintillator was amplified by a photomultplier tube and converted to an electrical signal via a voltage divider. This signal was then passed through a pre-amplifier and amplifier, before being discriminated into discrete counts, which were recorded after a given time interval.

In order to create the four desired Poisson count rates, we made use of 3 variable parameters in our measurement chain: the distance between source and detector, the fine/coarse gain of the amplifier, and the discriminator level of the counter. As the effects of the different parameters on the count rates was not independent, we made multiple simultaneous parameter changes in order to achieve the desired rate. The relative changes of these parameters for the 5, 10, and 100 count/second rates relative to the 1 count/second rate is shown in Table I.

TABLE I. Variable parameters in the measurement chain that were used to achieve different count rates. The actual values are listed for the ~ 1 count rate, whereas the relative values are shown for the other three rates.

Rate $[\frac{counts}{sec}]$	Distance $[cm]$	Gain	Discriminator $[V]$
~ 1	25	46.8	8
~ 5	100%	441%	100%
~ 10	100%	440%	62.5%
~ 100	100%	443%	65.6%

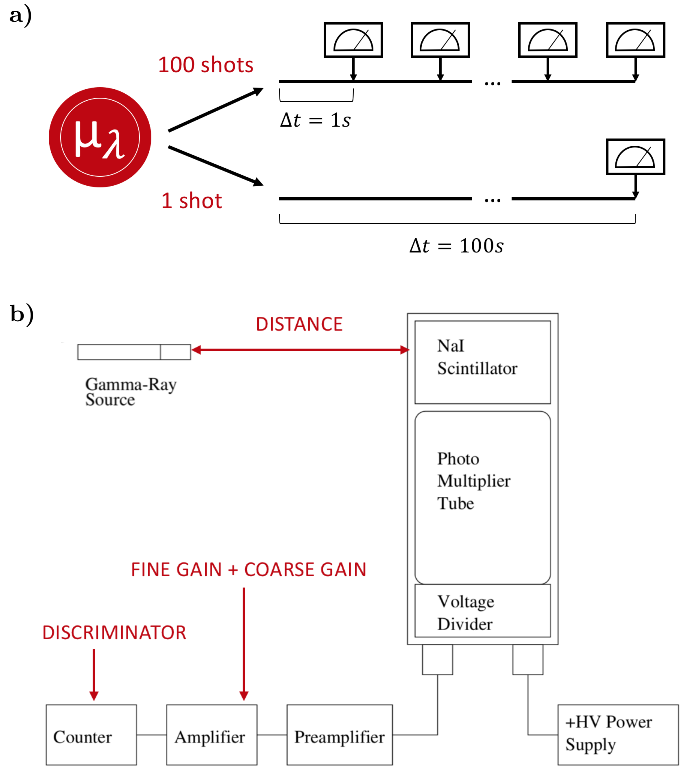

FIG. 2. **(a)** For each created physical Poisson process, we performed two types of measurements: 100-shots (100x 1s measurement)and 1-shot (1x 100s measurement). **(b)** The experimental measurement chain used to count emissions from a ^{137}Cs gamma ray source. By changing the parameters indicated in red, different Poisson event rates were achieved.[a]

[a] Image Source: JLAB Experiment Manual

III. DATA ANALYSIS

III.1. Experimentally Achieved Poisson Rates

As described in the *Experimental Setup* section, both a 1-shot and 100-shots measurement were performed on each of the created Poisson sources. Although we aimed to create the Poisson rates of 1, 5, 10, and 100, analysis of the measured data indicates that we actually achieved Poisson rates of 0.53, 6.15, 10.84, and 107.63. This analysis was performed in two different ways, however we focus primarily on the approach utilizing cumulative averaging and residual analysis. For details on the second approach, utilizing least-squares based curve fitting, refer to *Appendix B*.

III.1.1. Cumulative Averaging

The cumulative average represents the change in our estimate of the Poisson rate over time. For a sequence number, j, of the count, the cumulative average is given by

$$r_c(j) = \frac{\sum_{i=1}^{i=j} x_i}{\sum_{i=1}^{i=j} t_i}, \tag{3}$$

where x_i is the number of counts in time t_i. The cumulative average for each of the created Poisson processes is represented by a blue line in Fig 3. The dark blue dashed lines demonstrate the convergence of each of these cumulative averages to an asymptote as the number of 1 second measurements increases to 100 (in the ~ 1 rate case this convergence was not clear). The y-intercept of the asymptote was deemed to be the Poisson rate of the 100-shot measurement. However, as can be seen with the red

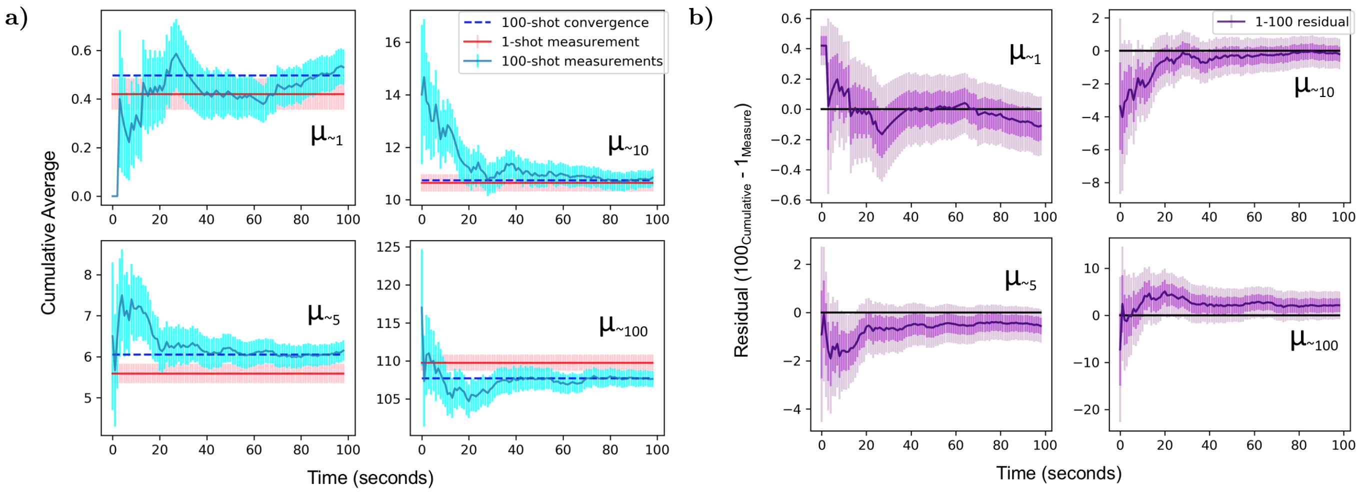

FIG. 3. **(a)** The cumulative average of 100-shot measurements plotted with the 1-shot measured value average across time, for each of the Poisson rates. **(b)** Residuals of the difference between the 100-shot CDF and 1-shot average plotted over time. Note that the dark purple uncertainty bars correspond to 1σ uncertainty, while the light purple correspond to 2σ uncertainty. Although the 1-shot and 100-shot measured values seem to be converging to significantly different quantities, residual analysis shows that the deviations are not statistically significant.

lines, the 1-shot measurements indicated a different Poisson rate, given by $\frac{x}{100}$, where x is the number of counts measured in the 100 second interval.

To calculate the uncertainties of these quantities, we utilize the fact that the uncertainty of a bin in a histogram can be modeled as $\sqrt{N}$, where N is the number of counts in the bin. In the 1-shot measurement case, we derive an uncertainty of

$$\sigma^{(1)} = \frac{\sqrt{x}}{100}, \tag{4}$$

and in the 100-shots measurement case, we derive an uncertainty of

$$\sigma_j^{(100)} = \frac{\sqrt{\sum_{i=1}^{i=j} x_i}}{\sum_{i=1}^{i=j} t_i}. \tag{5}$$

For a derivation of these uncertainties, refer to *Appendix C*.

III.1.2. Residual Analysis

As illustrated by the cumulative average plots in Fig 3, the Poisson rates in the 1-shot and 100-shot measurements converge to different values for 3 out of the 4 experiments. From an experimental perspective, this is concerning, since the 1-shot measurement was performed immediately after the 100-shot measurement, with no equipment changes or reason to expect a different outcome. Statistically, the fact that there is minimal overlap in the uncertainties of these quantities is also concerning. So, we performed analysis to determine whether these deviations could have just been statistical fluctuations or indicate a systematic error.

This analysis was performed using the residual of the 1-shot and 100-shot measurements, defined as

$$R_j = r_c(j) - \frac{x}{100}. \tag{6}$$

Utilizing propagation of errors, the uncertainty of the residual is

$$\sigma_j^{(R)} = \sqrt{\left(\sigma_j^{(100)}\right)^2 + \left(\sigma^{(1)}\right)^2}. \tag{7}$$

In Fig 3, $1\sigma_j^{(R)}$ is plotted in darker purple, while $2\sigma_j^{(R)}$ is plotted in lighter purple. Ideally, the residual would be zero. However, given that for all four experiments the zero line falls within $2\sigma_j^{(R)}$, and nearly $1\sigma_j^{(R)}$ generally, it appears that these were statistical fluctuations and not an issue with the experiment itself.

III.2. Experimental Poisson Limit Theorem Demonstration

The main theoretical goal of this experiment was to demonstrate the Poisson Limit Theorem discussed in the *Introduction & Theory* section. To do so, we arbitrarily choose to focus on the 100 counts/second Poisson data. Utilizing a least-squares based curve-fitting software, we fit a Poisson distribution to our data, indicated by the dashed magenta curve in Fig 6 (the underlying data histogram and uncertainties are plotted in grey). By leaving the count rate as a free parameter, the curve fitting specified a Poisson rate of 107.63, with $\chi^2 = 97.72$, $df = 5$, and $\chi_r^2 = 19.54$. Although the fit is not great, we are more interested in how well this Poisson distribution approximates a binomial distribution.

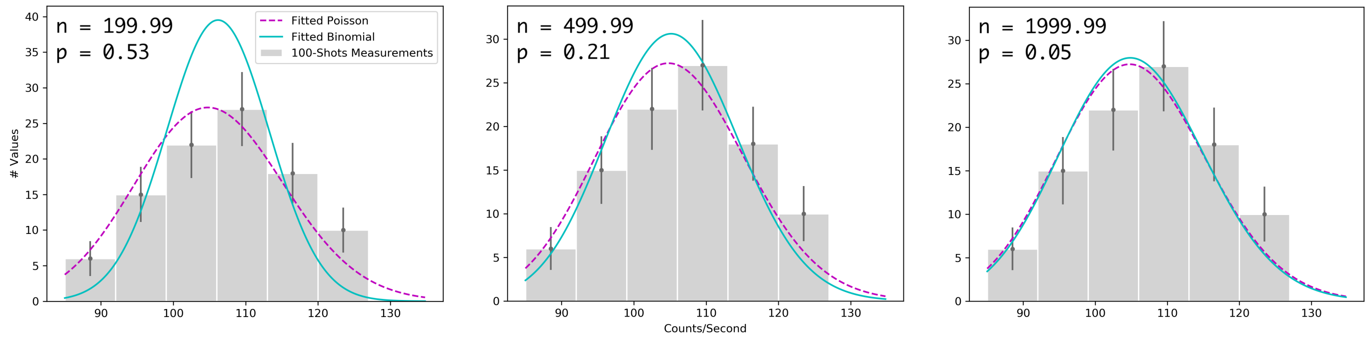

FIG. 4. We demonstrate the Poisson Limit Theorem, with the Poisson distribution (dotted lines) becoming a better approximation of the binomial distribution (solid line) as we increase n and decrease p, for $\lambda = np$.

Using the same data, we performed a new least-squares fit to a binomial distribution with free parameters n and p. As dictated by the Poisson Limit Theorem, the binomial distribution can be approximated by the Poisson distribution for $\lambda = np$. In order to assess the effect of the sizes of n and p on the overlap of these functions, we ran three separate curve-fits, each time changing the initial starting guess for n and p. Setting $\lambda = 107.65$ (the Poisson rate found by our previous curve fit), we then selected 3 arbitrary initialization values for n: 200, 500, 2000. Thus, p was respectively initialized to the $\frac{\lambda}{n}$ values of: 0.53, 0.21, 0.05. The results of these curve fits (plotted in cyan solid lines) were binomials parametrized almost exactly by the provided initialization. Notice how as the value of n increases and, respectively, p decreases, the Poisson distribution becomes a better approximation of the binomial distribution. To quantitatively verify this, we calculated the Bhattacharyya distance between the two probability distributions. This is defined as

$$D_B(p, b) = -ln\left(\sum_{x \in X} \sqrt{p(x)b(x)}\right), \tag{8}$$

where p is the Poisson distribution and b is the binomial distribution. The Bhattacharyya distance is maximized with a value of $D_B(p, b) = 1$ when the two distributions are identical. It is minimized with a value of $D_B(p, b) = 0$ when the two distributions are orthogonal. We report our calculated Bhattacharyya distances for each of the binomial distributions in Table II.

TABLE II. The Bhattacharyya distances, D_B, calculated between a binomial distribution (parametrized by n and p) and a Poisson distribution (parametrized by $\lambda = np = 107.65$).

n	p	D_B
199.99	0.53	0.918
499.99	0.21	0.979
1999.99	0.05	0.986

IV. RESULTS & CONCLUSIONS

Theoretically, we described the Poisson and binomial distributions, as well as their relationship through the Poisson Limit Theory. Using ^{137}Cs, we attempted to experimentally create Poisson processes of rates: 1, 5, 10, and 100 counts/second. Two different measurement procedures were used to determine the experimental count rates. Using a cumulative averaging (with residual analysis) approach, we claim the following experimentally achieved rates: 0.530 ± 0.097, 6.15 ± 0.34, 10.74 ± 0.46, and 107.71 ± 1.47. Finally, we performed varying binomial curve fits to our data. By increasing n and decreasing p of our binomial distribution, we saw the binomial-Poisson Bhattacharyya distance approach 1, verifying the Poisson Limit Theory both qualitatively and quantitatively.

ACKNOWLEDGMENTS

FV gratefully acknowledges Ghadah Alshalan's equal partnership, as well as the guidance and advice of the JLAB course staff and faculty.

[1] S. D. Poisson, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile precédées des règles générales du calcul des probabilités par sd poisson* (Bachelier, 1837).

[2] "Poisson theorem." Encyclopedia of Mathematics.

Appendix A: Shapes of Poisson and Binomial Distributions

As illustrated in Fig. 5, as λ increases, the Poisson distribution becomes broader and shorter, moving away from the value 0. For the binomial distribution, the shape of the distribution changes significantly depending on the values of n and p.

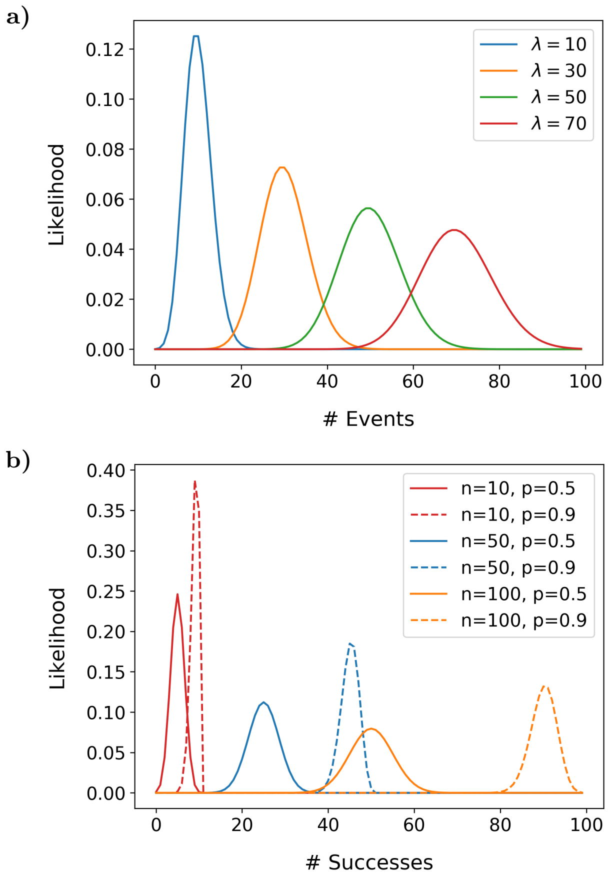

FIG. 5. **(a)** Poisson distributions for various event rates, λ. **(b)** Binomial distributions for various trial numbers, n, and success probabilities, p.

Appendix B: Measurement Uncertainty Derivation

To calculate the uncertainties of the 1-shot and 100-shots measurement quantities, we utilize the fact that the uncertainty of a bin in a histogram can be modeled as $\sqrt{N}$, where N is the number of counts in the bin. In the case of the 1-shot measurement, it is trivial to calculate the uncertainty, as we can treat the measurement as a single bin of a histogram, with x counts. Thus, the uncertainty is $\sqrt{x}$ and we scale by the 100 second averaging constant to achieve an uncertainty of

$$\sigma^{(1)} = \frac{\sqrt{x}}{100}. \tag{B1}$$

The 100-shot measurement case is a bit more tricky, so we will solve for it using induction. In the case of our first measurement period, the cumulative average is

$$r_c(1) = \frac{x_1}{t_1}. \tag{B2}$$

Utilizing the logic of Eqn B1, our uncertainty in this measurement is

$$\sigma_1 = \frac{\sqrt{x_1}}{t_1}. \tag{B3}$$

For the cumulative average consisting of our second measurement period, we have the expression

$$r_c(2) = \frac{1}{t_1 + t_2}(x_1 + x_2). \tag{B4}$$

The uncertainties of each of the individual counts in the sum are

$$\sigma_a = \frac{\sqrt{x_1}}{t_1 + t_2} \quad \text{and} \quad \sigma_b = \frac{\sqrt{x_2}}{t_1 + t_2}. \tag{B5}$$

Utilizing propagation of errors, the overall uncertainty is

$$\sigma_2 = \sqrt{\sigma_a^2 + \sigma_b^2} = \frac{1}{t_1 + t_2}\sqrt{x_1 + x_2}. \tag{B6}$$

Thus, if we continue this logic for all j measurement intervals, the uncertainty is

$$\sigma_j = \frac{\sqrt{\sum_{i=1}^{i=j} x_i}}{\sum_{i=1}^{i=j} t_i}. \tag{B7}$$

Appendix C: Curve-Fitting to Find Poisson Rates

Utilizing the sci-py curve_fit() function, modified to perform a χ^2 optimization, we achieved the Poissonian curve fits shown in Fig 6. From these fits we received an estimates of the Poisson rate and uncertainty/χ^2 information.

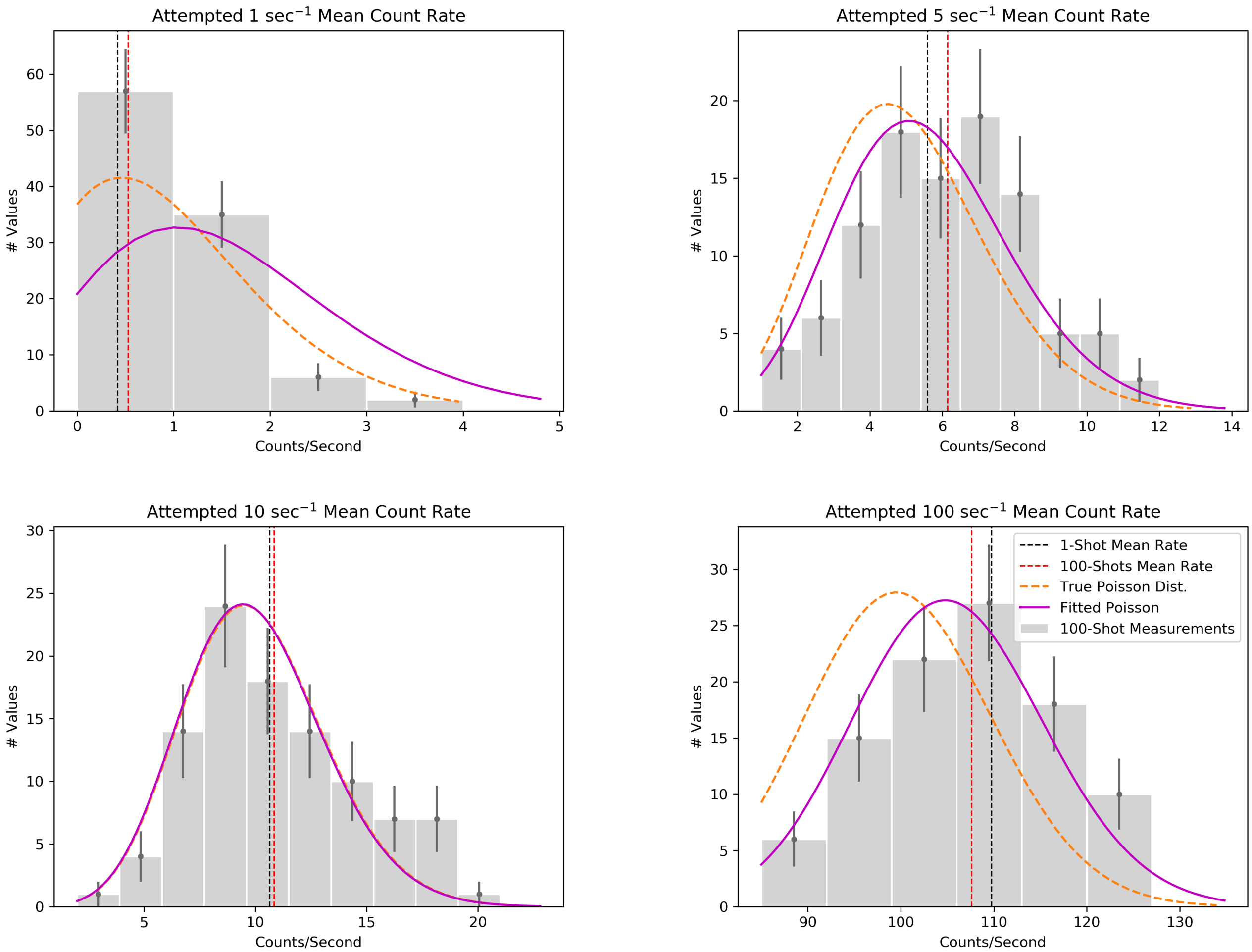

FIG. 6. Analysis of the experimentally achieved Poisson distribution relative to the True (or desired) Poisson rate distribution.