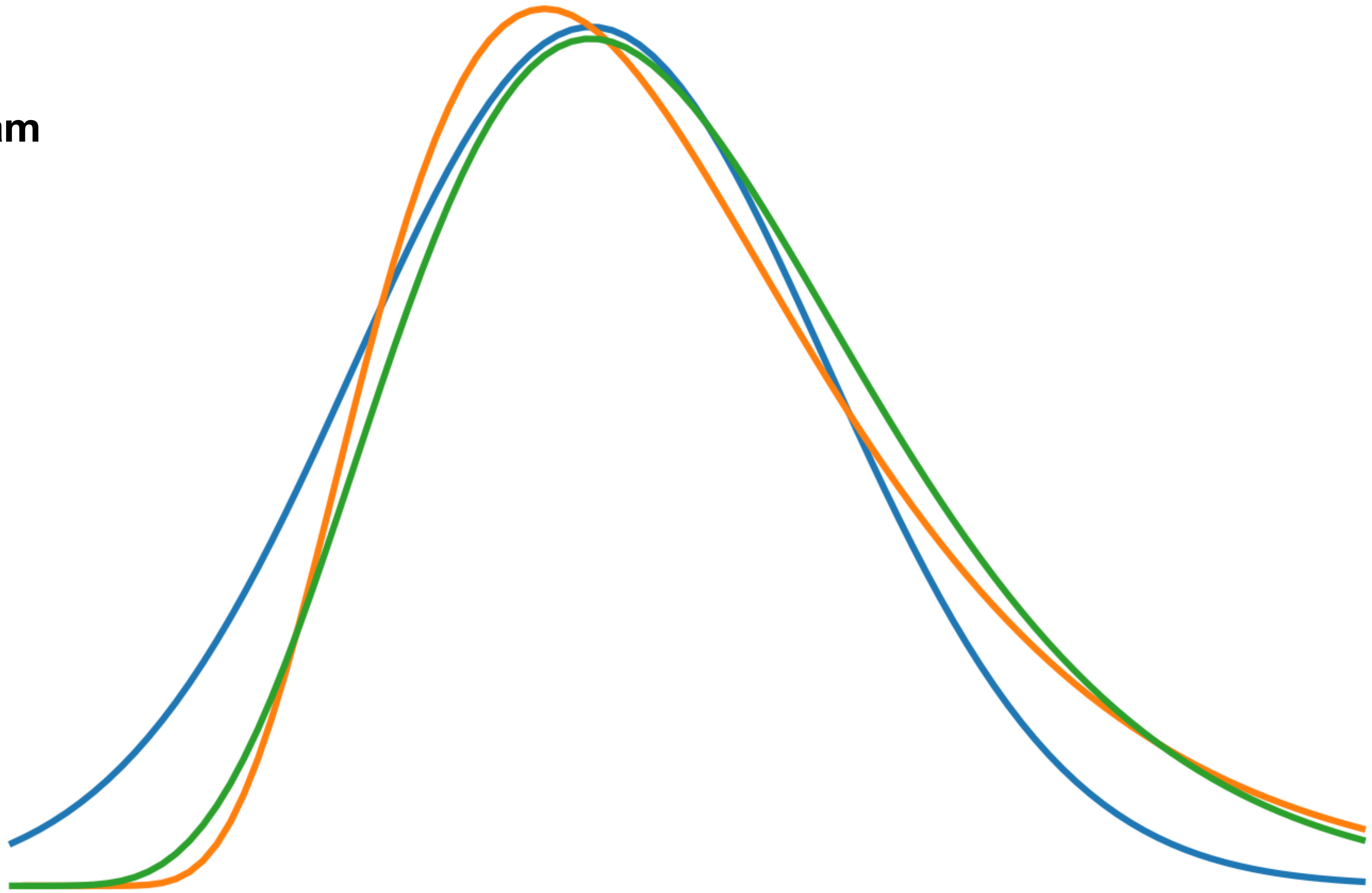
When can Poisson random variables be approximated as Gaussian?

Matthew Hayman[1], Willem Marais[2], Hristo Chipilski[3], Robert Stillwell[1], Scott Spuler[1]

[1]National Center for Atmospheric Research, Earth Observing Lab
[2]University of Wisconsin Space Science and Engineering Center
[3]National Center for Atmospheric Research Advanced Studies Program

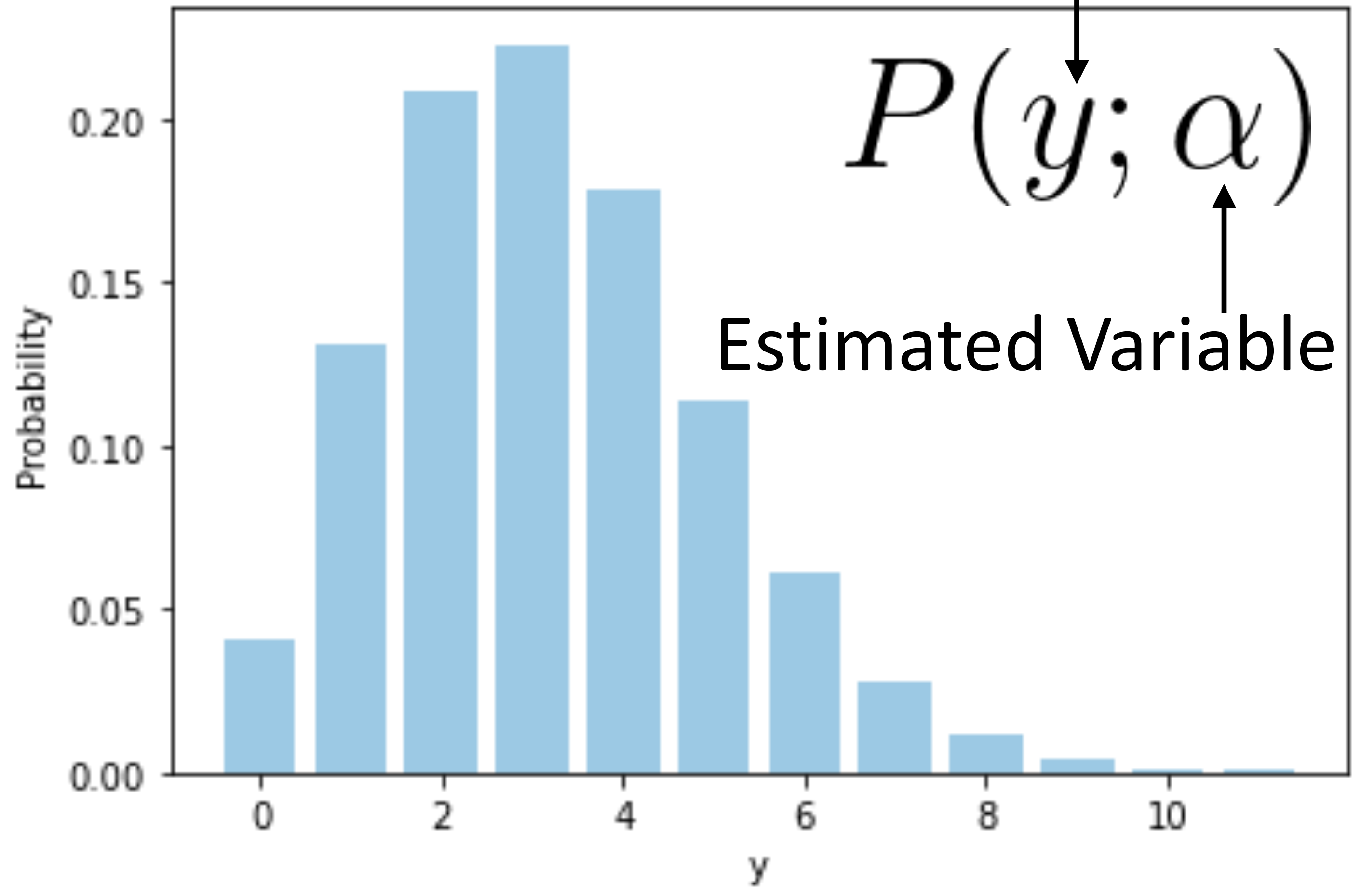


30th International Laser Radar Conference
2022 June 29
P2.27

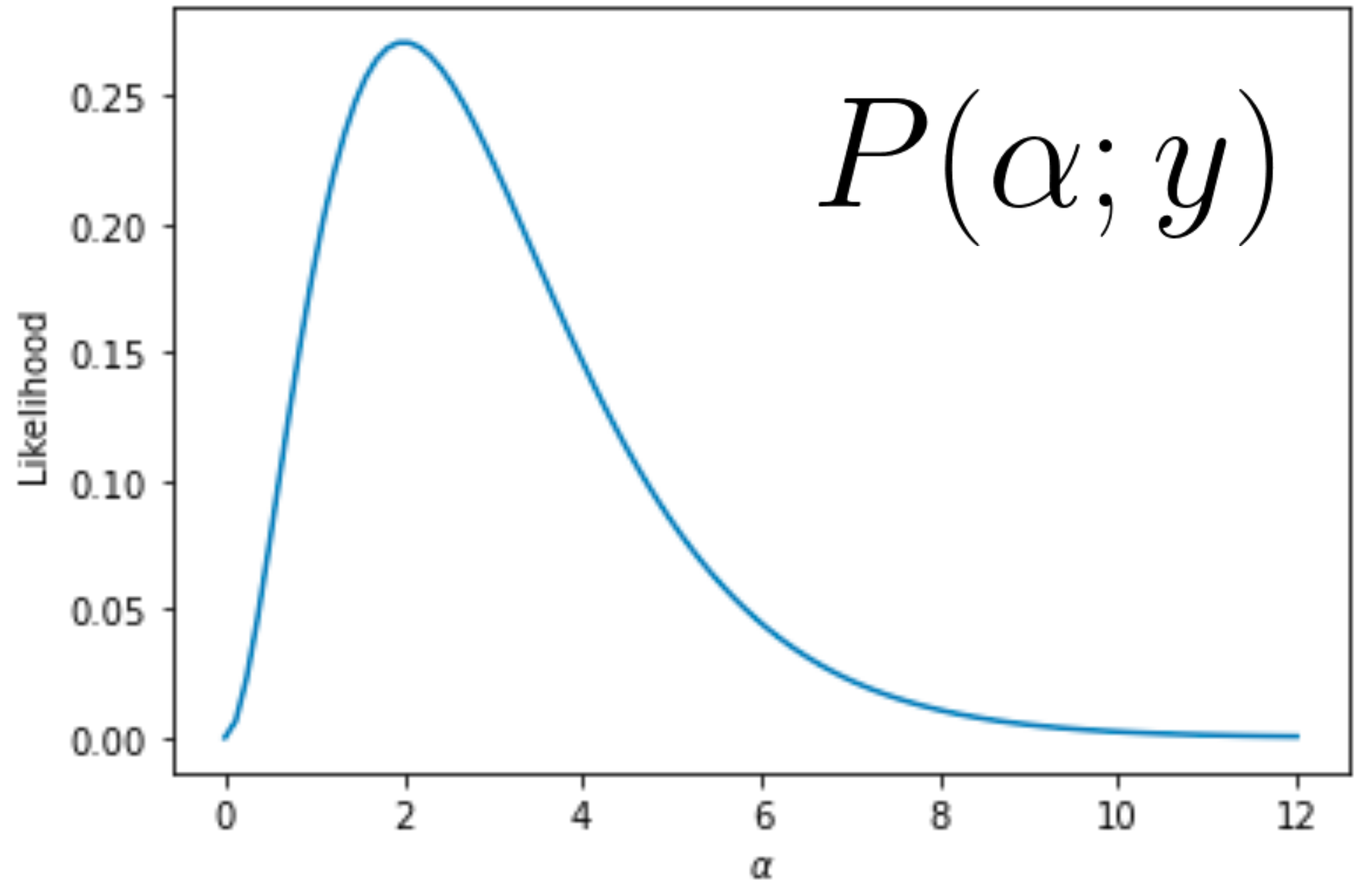
Maximum Likelihood Estimation

Probability Distribution

Observation

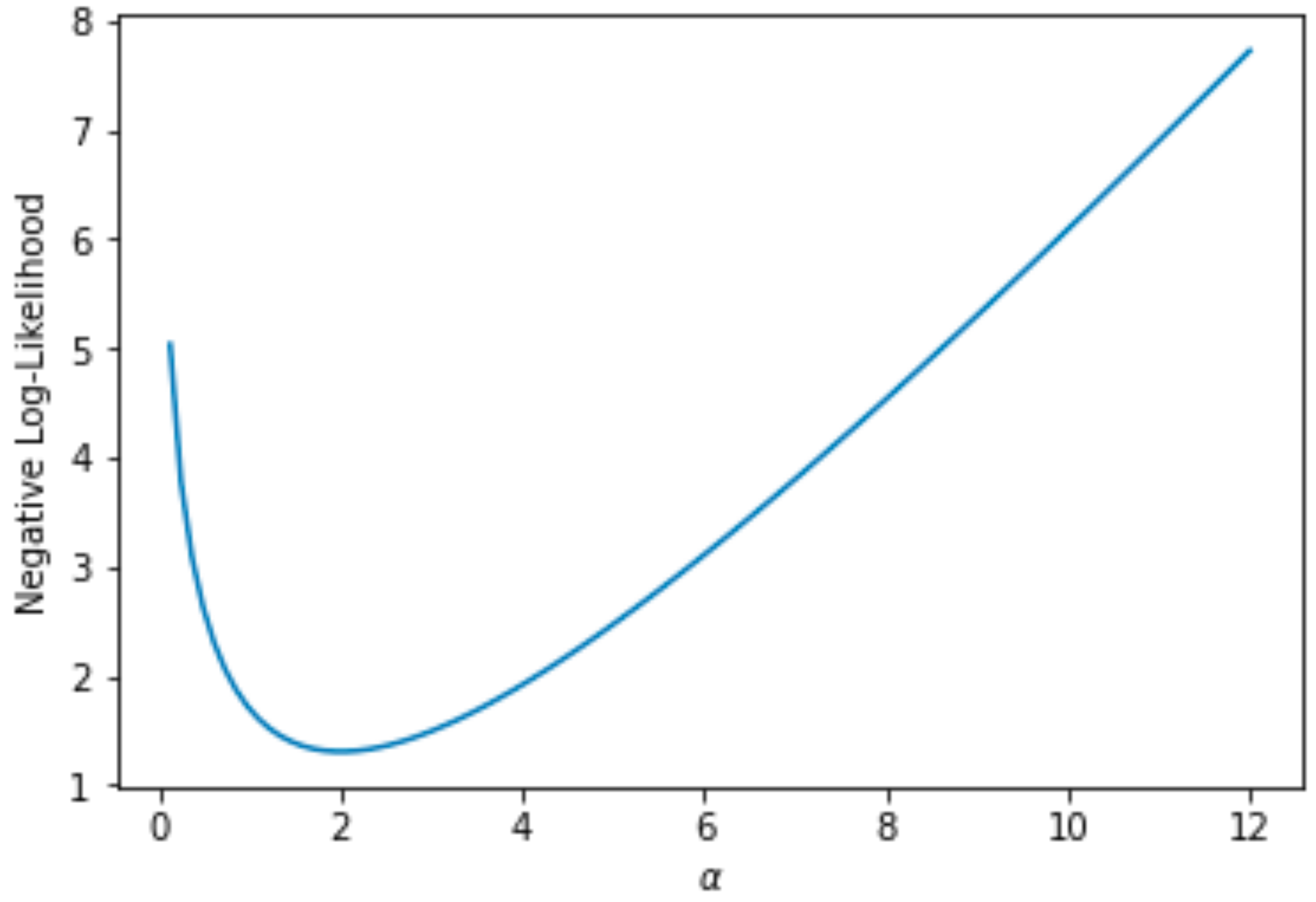


Likelihood



$$\mathcal{L}(\alpha; y) = -\ln P(\alpha; y)$$

Negative Log-Likelihood



The Three Noise Models

Poisson Negative Log-Likelihood

$$\mathcal{L}(\alpha; y) = \alpha - y \ln \alpha$$

Gaussian Approximation Negative Log-Likelihood

$$\boxed{\begin{aligned} \mu &= \alpha \\ \sigma^2 &= \alpha \end{aligned}}$$

This approximation is not usually used in lidar applications

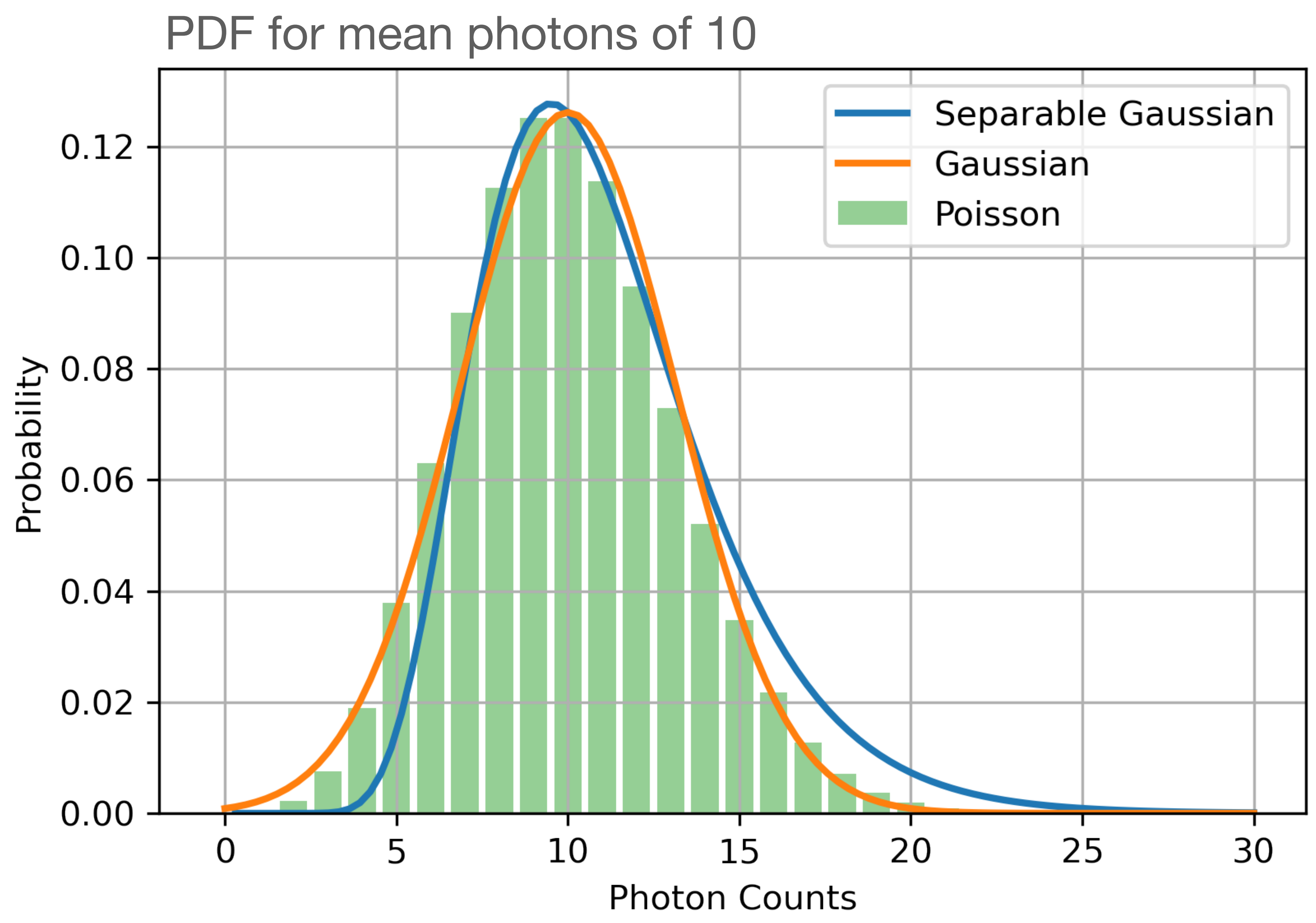
$$\mathcal{L}(\alpha; y) = \frac{1}{2} \ln(2\pi\alpha) + \frac{(y-\alpha)^2}{2\alpha}$$

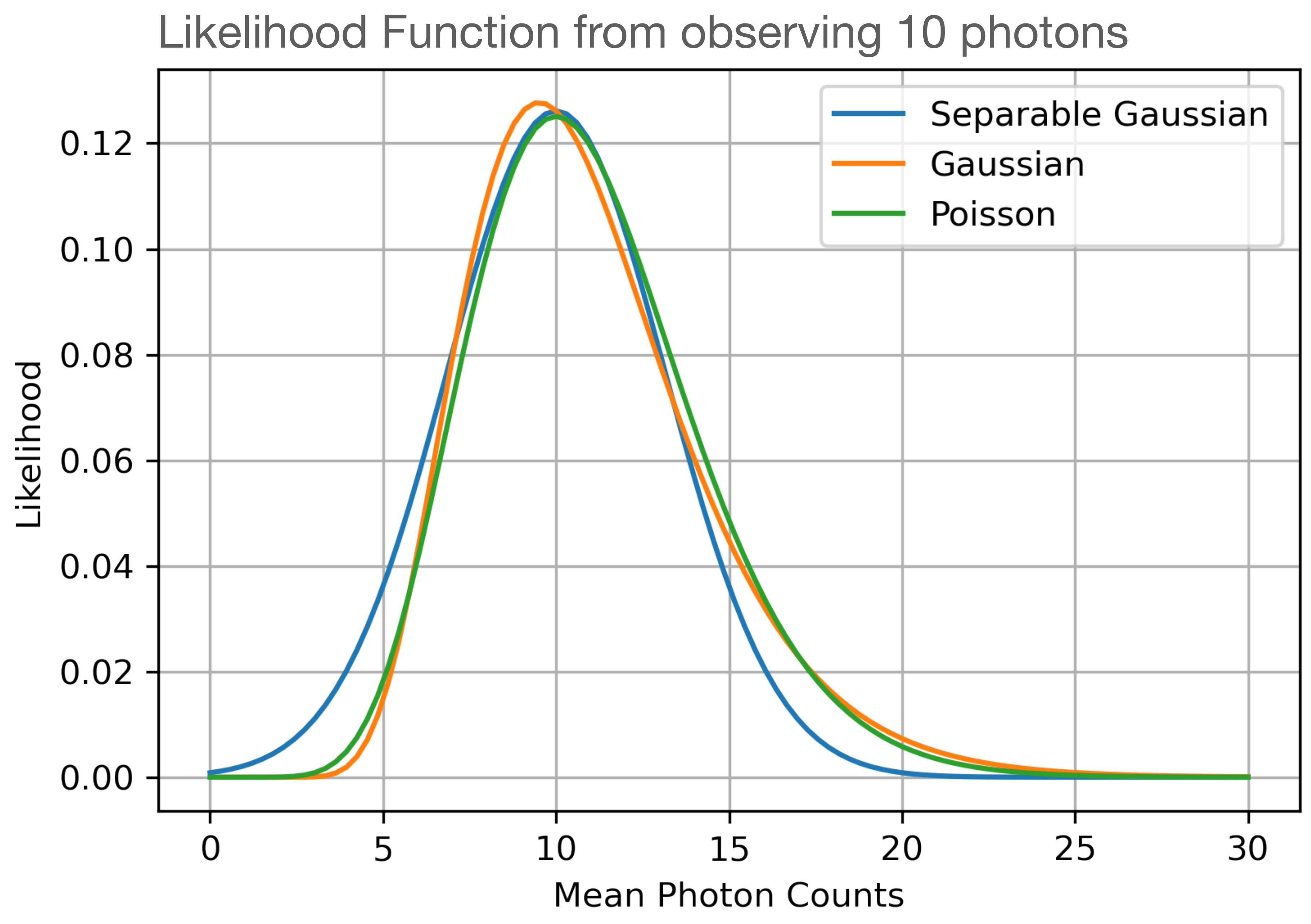
Separable Gaussian Approximation Negative Log-Likelihood

$$\boxed{\begin{aligned} \mu &= \alpha \\ \sigma^2 &= y \end{aligned}}$$

This approximation is less rigorous but more often used in lidar applications

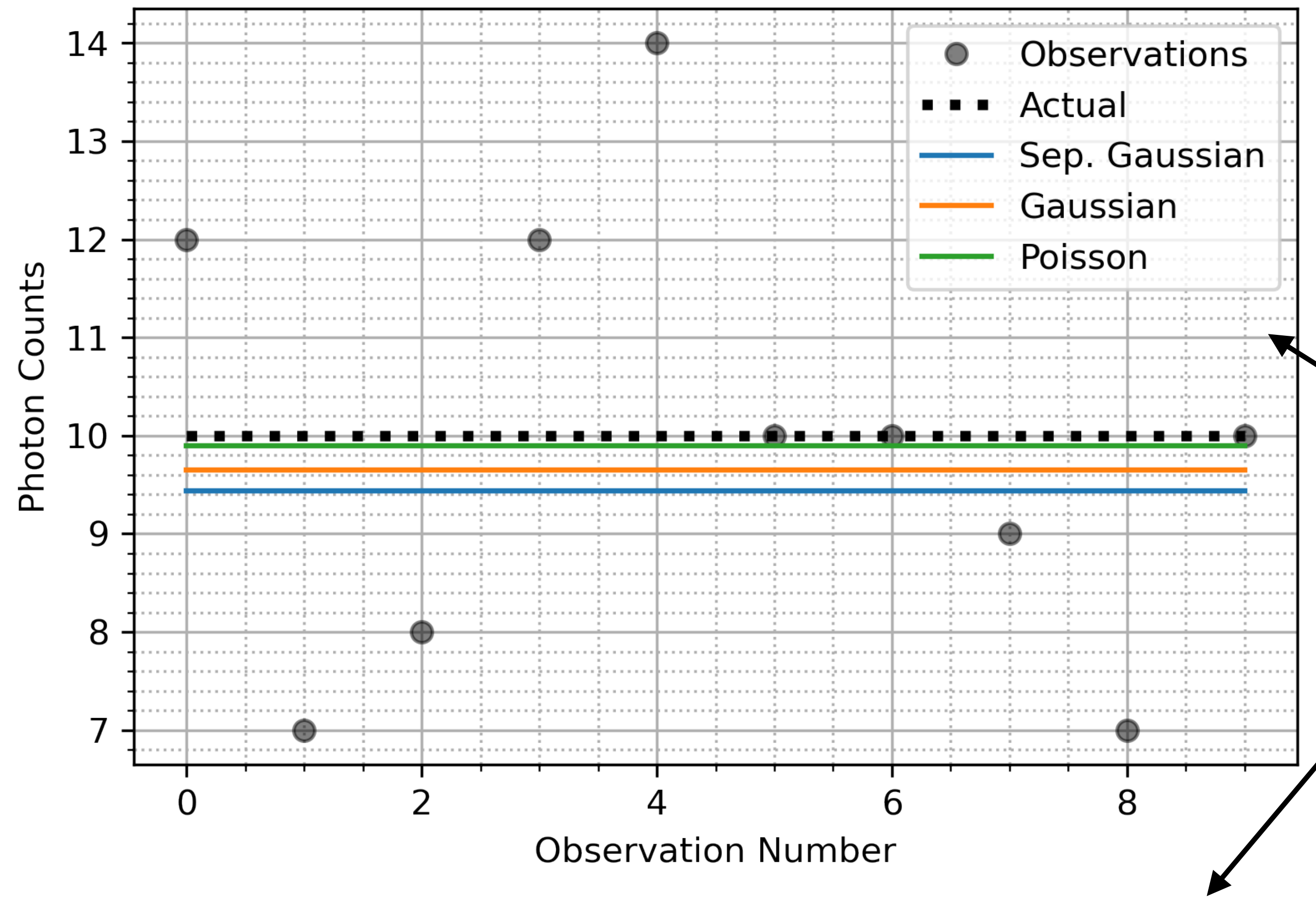
$$\mathcal{L}(\alpha; y) = \frac{(y-\alpha)^2}{2y}$$



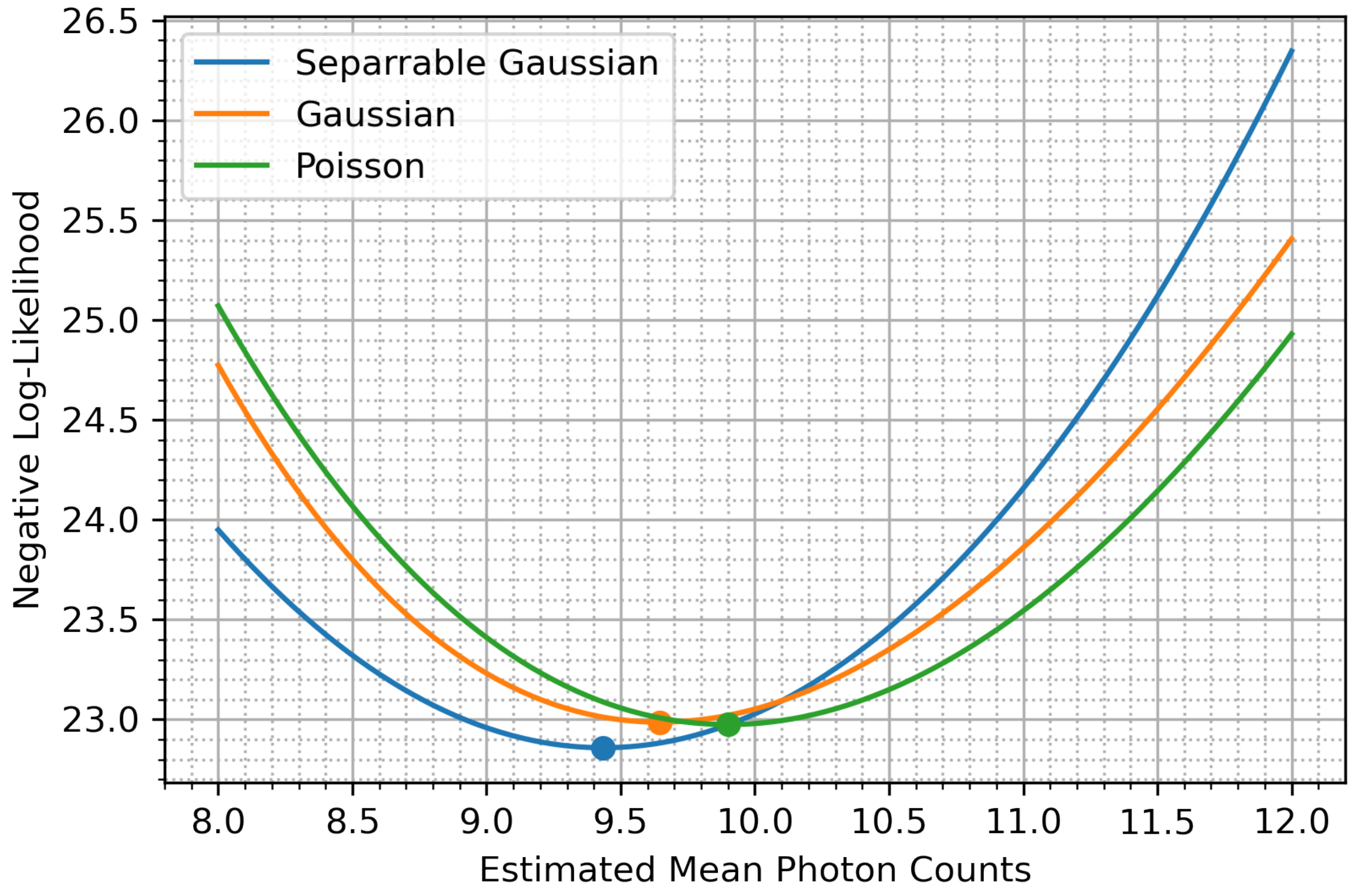


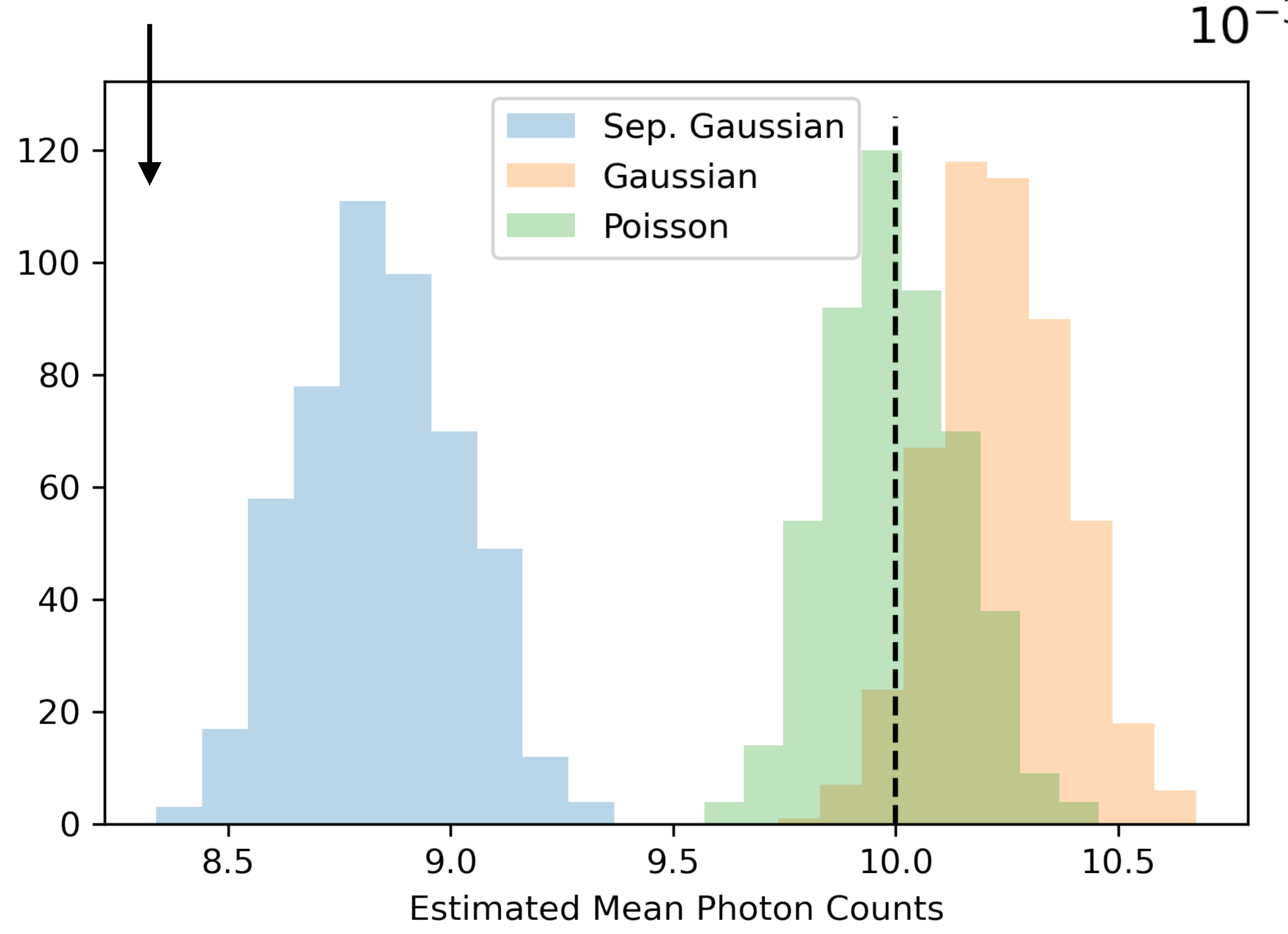
Experiment 1

Estimate Constant Signal

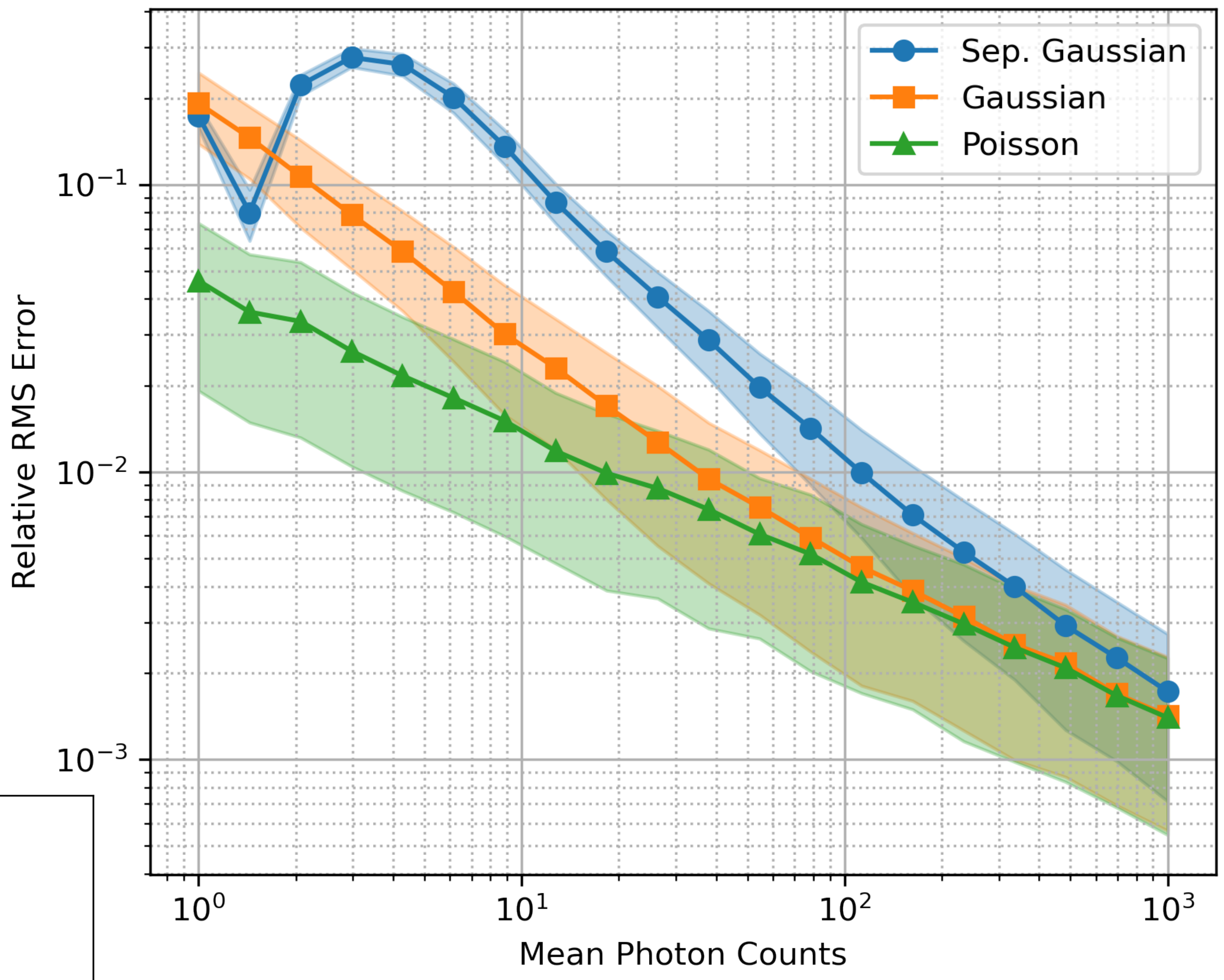


1. Random Poisson observations are generated
2. Fit is performed by minimizing the NLL of the three noise models
3. Repeat 500 times





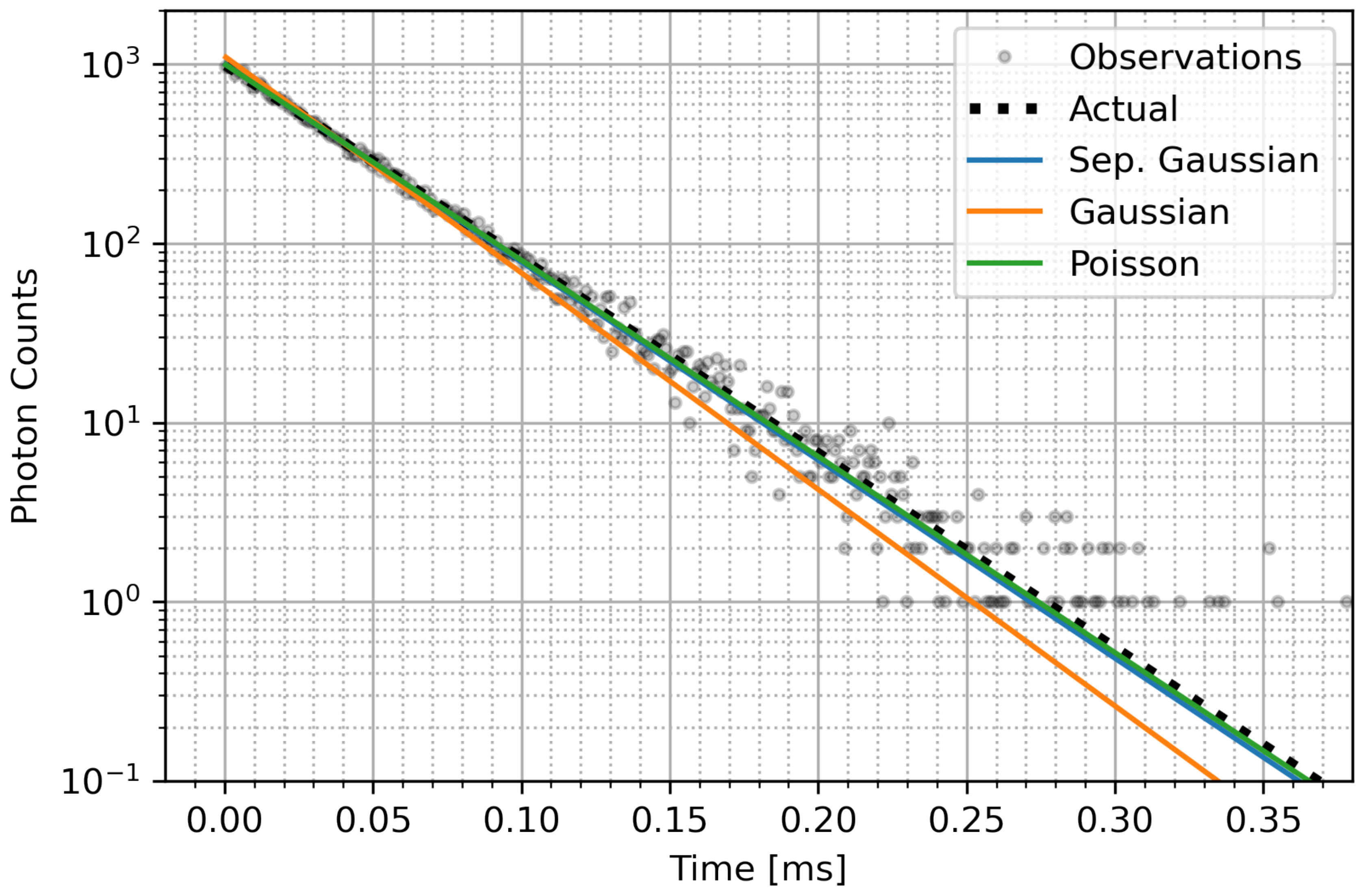
Mean and variance of the Relative RMS Error as a function of mean photons



- The Gaussian approximation resembles Poisson when the mean photons is greater than 300
- Separable Gaussian approximation does not resemble Poisson for mean photons considered here

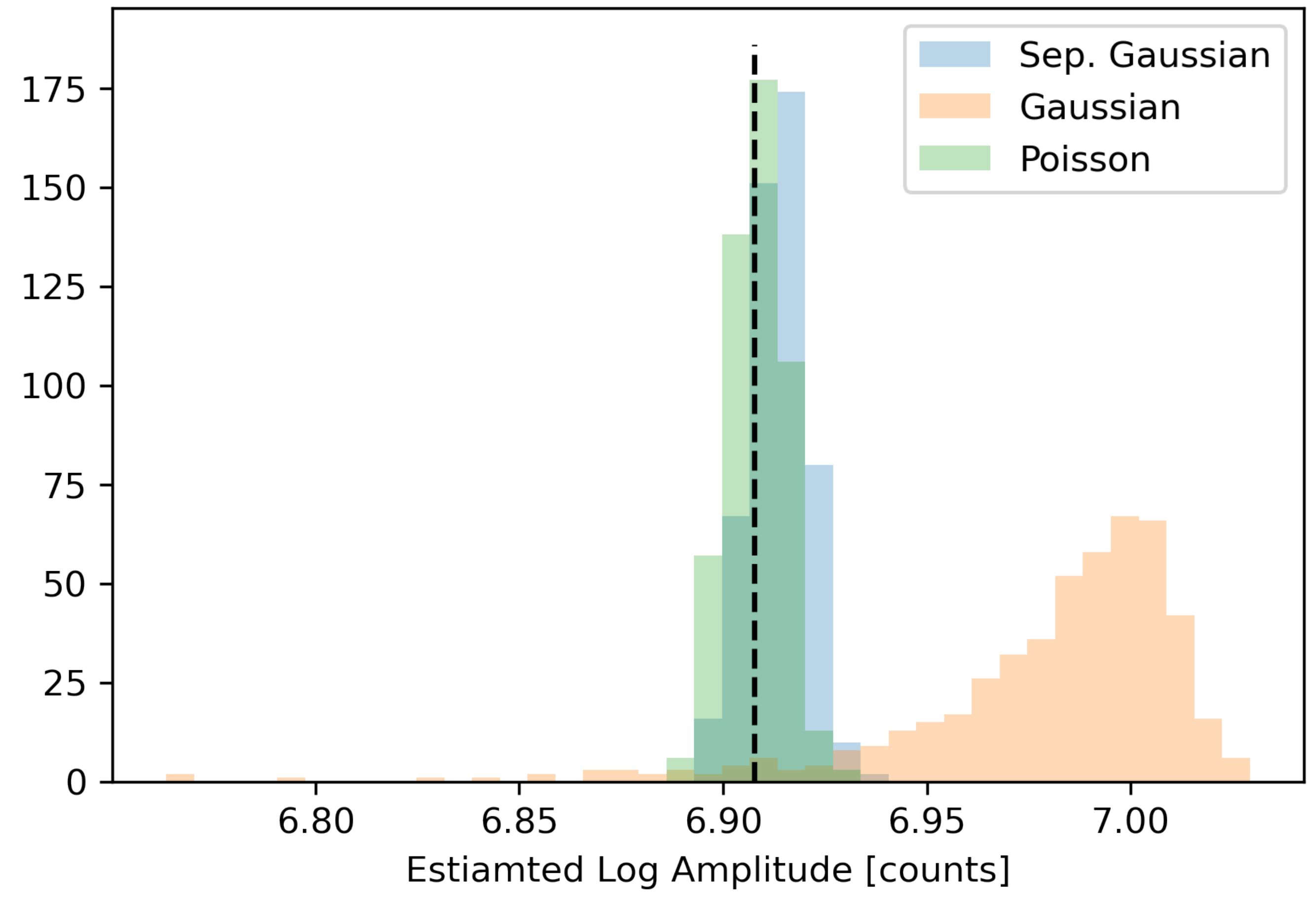
Experiment 2

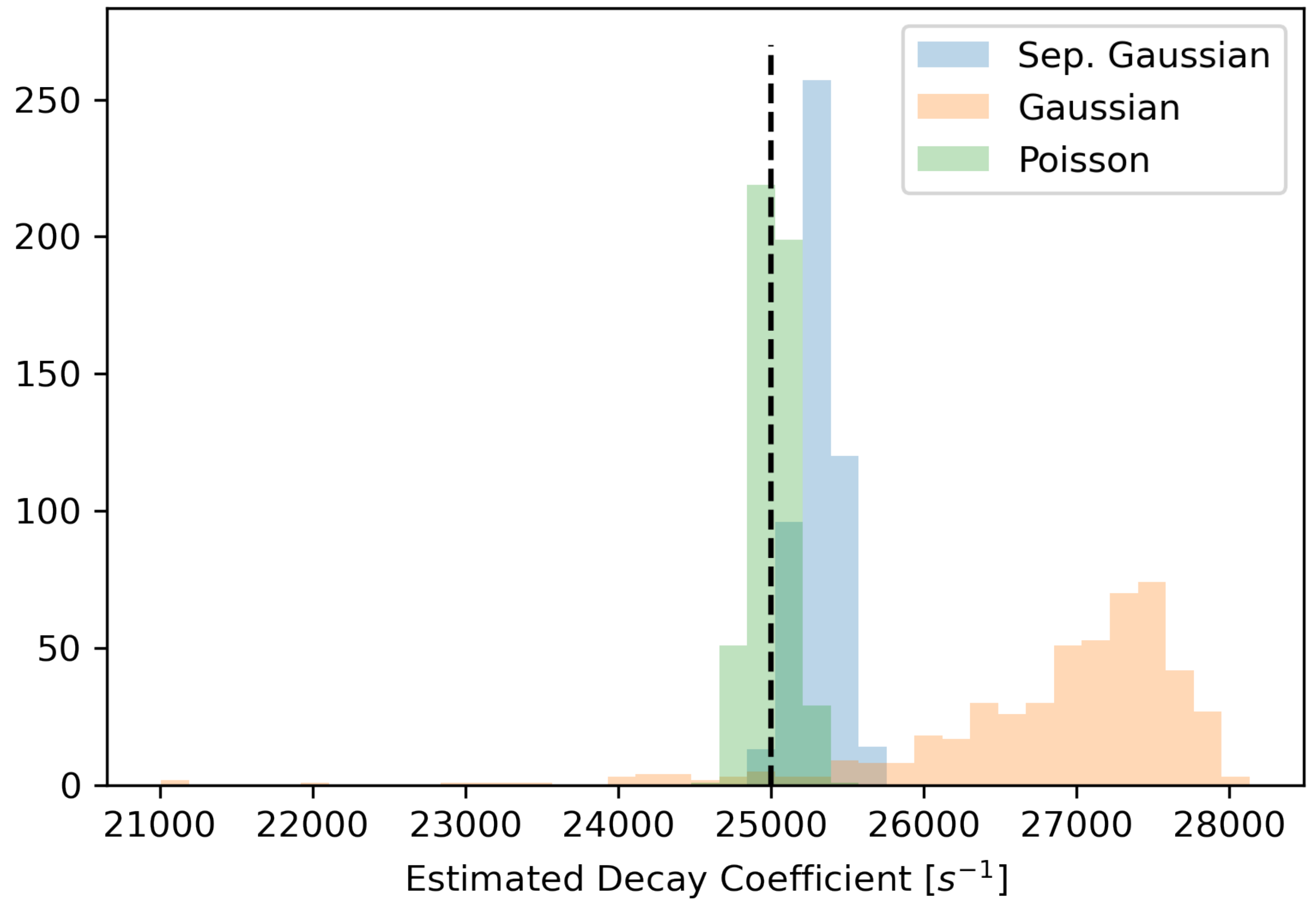
Exponential Fit



$$\alpha(t) = b\exp(-at)$$

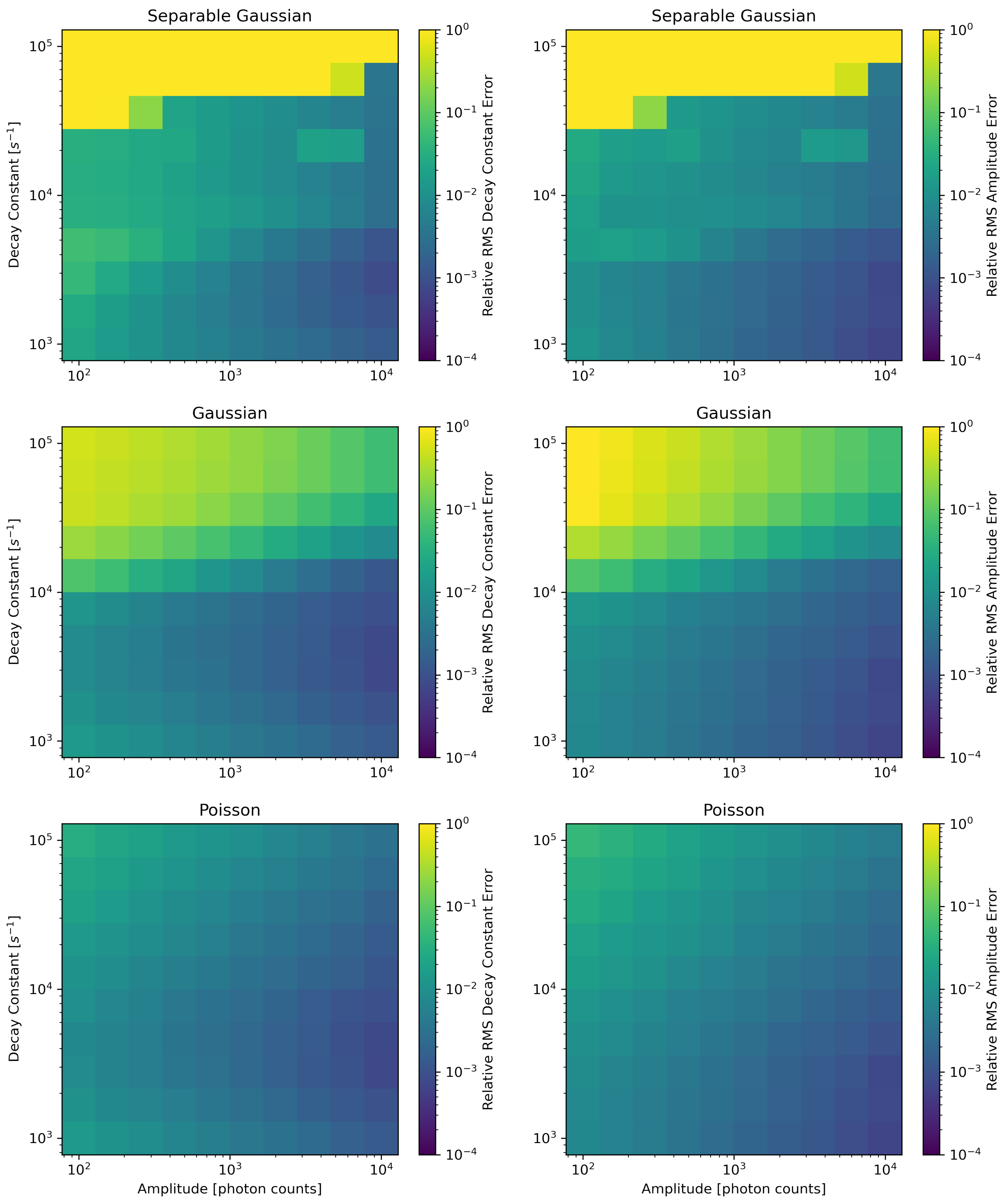
- Regions with low photon counts tend to bias the estimate of the overall function
- Relative performance of Separable Gaussian and Gaussian depends on the region considered





Decay Constant Relative RMS Error

Amplitude Relative RMS Error



Conclusions

When can Poisson random variables be approximated as Gaussian?

- Validity of each approximation depends on mean photon counts
  - Gaussian > 300 counts
  - Separable Gaussian > 1,000 - 10,000 counts
  - The approximations are probably poor candidates for high resolution processing
- The Separable Gaussian approximation typically requires higher counts to be valid
- Error from these approximations are likely to be application dependent
  - Further work is needed to quantify these effects in lidar retrievals
- Efforts are underway to avoid noise approximation in data assimilation

Funding:

NSF AGS-1930907

National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977