Bounding parameter uncertainty in single molecule localization

C.W. Seitz

June 27, 2022

1 An ideal single element detector

Suppose that we have an imaginary single element detector k, with a magical shape, that has the potential to detect every photon emitted by a fluorophore. The fluorophore emits photons according to a Poisson process with rate η, which can be shown to have mean $\mu = \eta \Delta t$ where Δt is the exposure time. The only source of photon-loss, is the quantum efficiency of k, which is a probability $0 \leq \gamma \leq 1$ that a photon is detected. We then have two random variables at play (1) how many photons will be emitted in Δt and (2) whether not each photon emitted will be detected:

$$M_k \sim \text{Poisson}(\eta) \quad N_k \sim \text{Bern}(\gamma, 1 - \gamma)$$

As mentioned, the probability M_k photons will be detected in the exposure time Δt is is a discrete Poisson random variable, which has the form

$$\text{Poisson}(n; \eta) = \frac{\exp\left(-\eta \Delta t\right)\left(\eta \Delta t\right)^n}{n!}$$

Based on our intuition, we might expect that throwing out each photon with probability γ will effectively reduce the rate of the Poisson process by a factor of γ. However, we would like to prove this in a mathematically rigorous way. We start by noting that we can define a new variable $S_k = M_k N_k$ such that our measurment is sampled from the density $P(S_k)$. Let's write the joint distribution over all three variables

$$P(S_k, M_k, N_k) = P(S_k | M_k = m_k, N_k = n_k) P(M_k | N_k = n_k) P(N_k = n_k)$$

Note that $m_k, s_k \in \Omega$ where $\Omega = \{z \in \mathbb{Z} : z \geq 0\}$ We can marginalize this distribution over M_k and N_k

$$P(S_k) = \sum_{m_k} \sum_{n_k} P(S_k | M_k = m_k, N_k = n_k) P(M_k | N_k n_k) P(N_k = n_k)$$

Expanding this expression out over the support of N_k gives

$$P(S_k) = \sum_{m_k} P(S_k|M_k = m_k, N_k = 0)P(M_k|N_k = 0)P(N_k = 0)$$
$$+ P(S_k|M_k = m_k, N_k = 1)P(M_k|N_k = 1)P(N_k = 1)$$

At this point it is quite clear that we have

$$P(S_k) = \sum_{m_k} \gamma\delta(s_k - m_k)P(M_k = m_k) + (1 - \gamma)\delta(s_k)P(M_k = m_k)$$
$$= \gamma P(M_k) + \delta(s_k)(1 - \gamma)$$

Interestingly, this result is fairly intuitive - the distribution $P(S_k)$ is given by multiplying the probability mass $P(M_k)$ by a factor of γ and inflating the value at zero by the difference between the new and old distributions. It is worth nothing that adding $1 - \gamma$ to the probability mass at zero is effectively reducing the Poisson rate η to $\gamma\eta$, since the probability of observing an event is uniform in the interval $(t, t + \Delta t)$. Therefore, we conclude that

$$P(S_k) = \text{Poisson}(S_k; \lambda_k)$$

where we have used the *effective rate* $\lambda_k = \gamma\eta_k$.

1.1 Corrupting the signal

In the real-world, noise is unavoidable. Detectors often suffer from dark noise (thermal noise) and there may also be a background signal. Considering the former first, we define another r.v. $W_{\theta,k}$ which represents Gaussian dark noise of the detector element. Unless otherwise specified we will always assume that $W_k \sim \mathcal{N}(m_k, \sigma_{w,k}^2)$. Adding in dark noise leads to yet another distribution, this time over the corrupted signal H_k. Notice that while S_k has units of photons [p], that H_k has units of photoelectrons $[e^-]$. The conversion factor between the two is the gain of the detector element g_k $[e^-\text{p}^{-1}]$. Moving on, we employ a similar trick as before, evaluating the joint density $P(S_k, H_k)$

$$P(S_k, H_k) = P(H_k|S_k)P(S_k)$$
$$= \frac{1}{Z} \exp\left(-\frac{(H_k - g_k n_k)^2}{\sigma_k^2}\right) \frac{\exp(-\lambda_k) \lambda_k^{n_k}}{n_k!}$$

Marginalizing over S_k gives the desired distribution over H_n
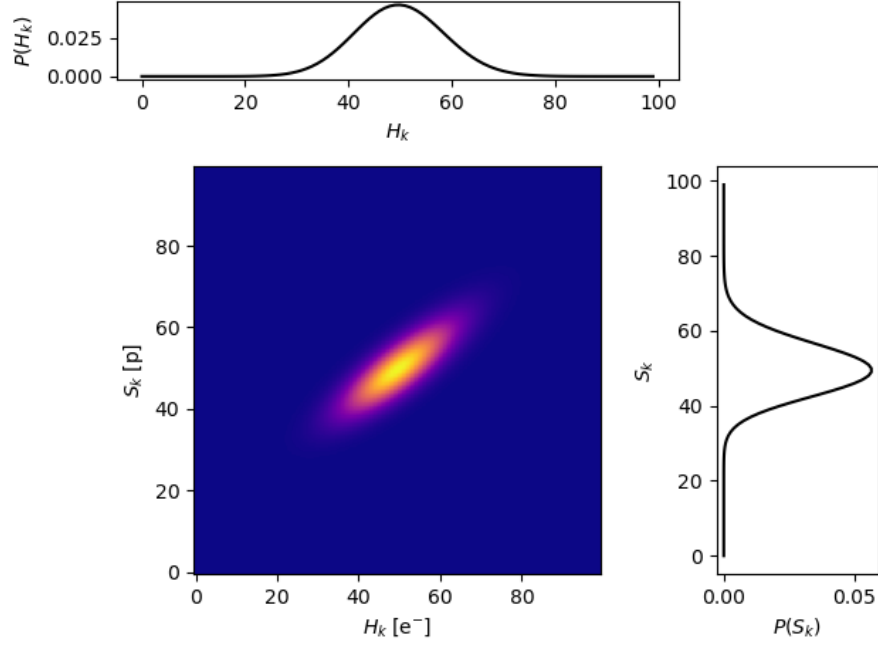
(a) Theoretical prediction of $P(S_k, H_k)$ and marginal distributions for $g_k = 1$, $\lambda_k = 10$ cpms, and $\Delta t = 5$ms and $\sigma_k = 5\ e^-$
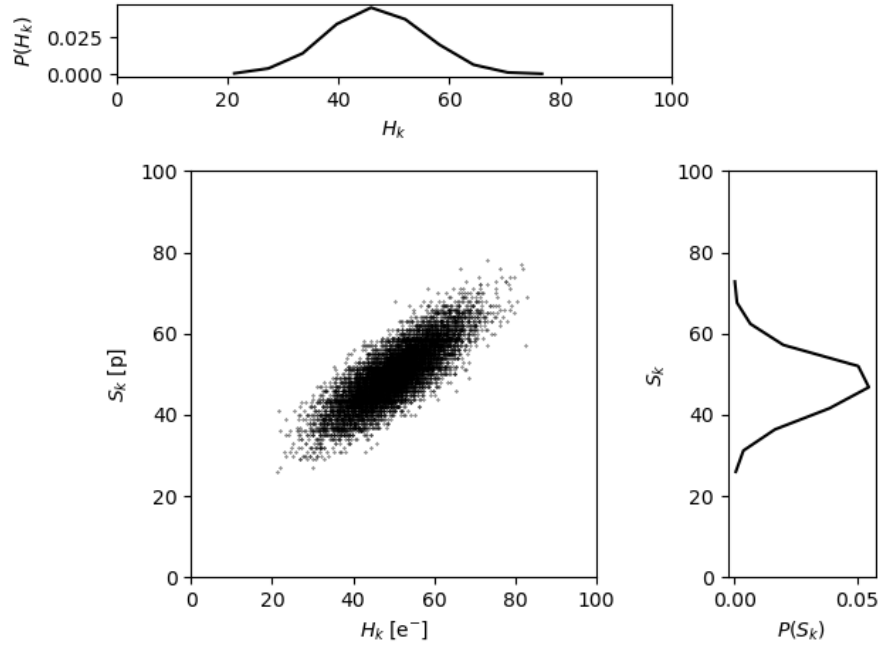
(b) Simulation of $P(S_k, H_k)$ and marginal distributions

3

$$P(H_k) = \frac{1}{Z} \sum_{n_k=0}^{\infty} \frac{\exp\left(-\lambda_k\right)\lambda_k^{n_k}}{n_k!} \exp\left(-\frac{(H_k - g_k n_k)^2}{\sigma_k^2}\right)$$

Notice that when $\mu_k = \lambda_k \Delta t$ is large, $P(S_k)$ is well-approximated by a Gaussian distribution. It turns out the for $\mu > 20$ the Poisson distribution is approximately

$$S_k \sim \mathcal{N}(\lambda_k \Delta t, \lambda_k \Delta t)$$

which turns out to be a very useful approximation for simplifying $P(H_k)$. It can be shown that the product of two Gaussian PDFs is

$$\mathcal{N}(\mu_1, \sigma_1^2) \times \mathcal{N}(\mu_2, \sigma_2^2) = \mathcal{N}\left(\frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}\right)$$

a property which we can apply to the above equation to get

$$P(H_k) = \frac{1}{Z} \sum_{n_k=0}^{\infty} \exp\left(-\frac{1}{2}\left(H_k - \frac{g_k n_k (\lambda_k \Delta t)^2 + (\lambda_k \Delta t)\sigma_k^2}{(\lambda_k \Delta t)^2 + \sigma_k^2}\right)\left(\frac{1}{(\lambda_k \Delta t)^2} + \frac{1}{\sigma_k^2}\right)\right)$$

2 An array detector

Most detectors used for imaging have many elements (pixels) so that we can record an image projected onto the detect by a lens or system of lenses. Due to diffraction, a point emitter will be registered as a diffraction limited spot. We will adopt the Gaussian PSF approximation for such a spot:

$$q(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{2\sigma^2}\right)$$

The function $q(x, y)$ can be thought of an the intensity profile of the fluorophore emission, although this is not exactly correct. The wave particle-duality in quantum mechanics says that observation of a wave produces a particle, and our detector is nothing but an observer. The EM field that is waving is observed and then we record a particle. However, the statistical nature of quantum mechanics says that $q(x, y)$ is actually a probability distribution and our photons are samples from it. The detector array is binning those samples in a 2D histogram that we call an image. All that said, the Poisson rate at every pixel is not the same and we need to re-evaluate $P(S_k)$ for each pixel. This is fairly simple, since the rate at every bin is just the probability assigned to that bin, which is an integral of $q(x, y)$ over the bin

$$\lambda_k = \gamma \cdot \int_{C_k} q(x, y) dx dy$$

3 Fisher information and the Cramer-Rao bound

3.1 Example: Fisher information for the normal distribution

Suppose the we have a normally distributed random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ where μ and σ^2 are both unknown parameters. We are also given a batch of N samples $\{X_i\}_{i=1}^N$ and we would like to know how much information those samples carry about our parameters $\theta = (\mu, \sigma)$. In principle there is some probability distribution $P(\theta) = P(\mu, \sigma)$ which can be initially taken to be uniform (which has maximum entropy). A natural question that follows is, how much does the batch of samples reduce the entropy of $P(\theta)$ and thus our uncertainty about θ?

$$I_{ij}(\theta) = \mathop{\mathbb{E}}_{x \sim p} \left[\frac{\partial}{\partial \theta_i} \left(\ell(x|\theta) \right) \frac{\partial}{\partial \theta_j} \left(\ell(x|\theta) \right) \right]$$

For the 1D normal distribution, we only have two parameters, so the Fisher information matrix has a simple form

$$\mathbf{I}(\theta) = \begin{pmatrix} \partial_\mu^2 \ell & \partial_\mu \partial_\sigma \\ \partial_\sigma \partial_\mu & \partial_\sigma^2 \ell \end{pmatrix}$$

It is straightforward to show that the log-likelihood of the data for this model is given by

$$\ell(x|\theta) = -\frac{1}{2} \log 2\pi\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\partial_\mu^2 \ell = -\partial_\mu^2 \frac{(x-\mu)^2}{2\sigma^2}$$

3.2 The array detector again

When there are many parameters, the Fisher Information (second moment of the score) is a covariance matrix

$$I_{ij}(\theta) = \mathop{\mathbb{E}}_{x \sim p} \left[\frac{\partial}{\partial \theta_i} \left(\ell(x|\theta) \right) \frac{\partial}{\partial \theta_j} \left(\ell(x|\theta) \right) \right]$$

We have shown that the model for the number of photoelectrons at a pixel is

$$P(H_k) = \frac{1}{Z} \sum_{s=0}^{\infty} \frac{\exp\left(-\lambda_k\right) \lambda_k^{n_k}}{n_k!} \exp\left(-\frac{(H_k - g_k n_k)^2}{\sigma_k^2} \right)$$

As we have noted above, λ_k is dependent on $\int q(x,y)dxdy$ and therefore the PSF parameters $\theta = (\mu_x, \mu_y, \sigma)$. These can be plugged into the following Fisher information matrix

$$
\begin{aligned}
I_{ij}(\theta) &= \underset{H \sim P(H)}{\mathbb{E}} \left[\frac{\partial}{\partial \theta_i} \log P(H_k) \frac{\partial}{\partial \theta_j} \log P(H_k) \right] \\
&= \underset{H \sim P(H)}{\mathbb{E}} \left[\frac{\partial}{\partial \theta_i} \log P(H_k) \frac{\partial}{\partial \theta_j} \log P(H_k) \right]
\end{aligned}
$$

where $H \sim P(H)$ is perhaps a small patch of an image containing a diffraction limited spot. We need to then calculate the derivatives with respect to each of the parameters $\theta = (x_0, y_0, \sigma)$.

$$
\begin{aligned}
\frac{\partial}{\partial x_0} \log P(H_k) &= \frac{1}{P(H_k)} \frac{\partial P(H_k)}{\partial x_0} \\
&= \frac{1}{ZP(H_k)} \sum_{n_k=0}^{\infty} \frac{\partial}{\partial x_0} \frac{\exp\left(-\lambda_k\right) \lambda_k^{n_k}}{n_k!} \exp\left(-\frac{(H_k - g_k n_k)^2}{\sigma_k^2} \right)
\end{aligned}
$$

Let's consider just the parts relevant to the derivative $\frac{\partial}{\partial x_0} \exp\left(-\lambda_k\right) \lambda_k^{n_k}$. This is far too difficult to compute analytically. Instead, we need to compute this infinite sum numerically, using Monte Carlo integration to compute λ_k on a 3D lattice of points in the parameter space θ. In total, we have to for each pixel (1) compute this sum over the 3D parameter space numerically (2) take the logarithm of all values (3) compute the Hessian matrix (4) average the Hessians over pixels to get the Fisher information matrix.

We can then derive the Cramer-Rao bound by inverting that matrix and use that bound to filter out particles which cannot be well estimated by other means. It may be the case that the MLE estimator is an optimal estimator.