1 Fisher Information for a Poisson variable

This is derivation of the fisher information for Poisson distributed variable X_k with mean λ.

$$X_k \sim \mathrm{Po}(x_k, \lambda_k(\theta)) = \frac{\lambda_k^{x_k} e^{-\lambda_k}}{x_k!}$$

1.1 Likelihood

Likelihood of the Poisson distributed variable:

$$l(\theta) = \prod_{k=1}^{K} l_k = \prod_{k=1}^{K} \frac{\lambda_k^{x_k}(\theta) e^{-\lambda_k(\theta)}}{x_k!} \tag{1}$$

Log-Likelihood with detection in K pixels:

$$\mathcal{L} = \sum_k \left(x_k \log \lambda_k - \lambda_k - \log x_k\right)$$

1.2 Fisher Information

Fisher information:

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \theta^2}\right] = \mathbb{E}\left[\left(\frac{\partial \mathcal{L}}{\partial \theta}\right)^2\right] = \mathbb{E}\left[\left(\sum_k \frac{\partial \log(l_k)}{\partial \theta}\right)^2\right] = \mathbb{E}\left[\left(\sum_k \frac{1}{l_k}\frac{\partial l_k}{\partial \theta}\right)^2\right] \tag{2}$$

$$\begin{aligned} I(\theta) &= \mathbb{E}\left[\left(\sum_k \frac{1}{l_k}\frac{\partial l_k}{\partial \theta}\right)\left(\sum_m \frac{1}{l_m}\frac{\partial l_m}{\partial \theta}\right)\right] \\ &= \mathbb{E}\left[\sum_k \frac{1}{l_k^2}\left(\frac{\partial l_k}{\partial \theta}\right)^2\right] + \mathbb{E}\left[\sum_k \sum_{m \neq k} \frac{1}{l_k}\frac{\partial l_k}{\partial \theta}\frac{1}{l_m}\frac{\partial l_m}{\partial \theta}\right] \end{aligned}$$

as x_k are iid then the second term can be expressed as

$$\mathbb{E}\left[\sum_k \sum_{m \neq k} \frac{1}{l_k}\frac{\partial l_k}{\partial \theta}\frac{1}{l_m}\frac{\partial l_m}{\partial \theta}\right] = \sum_k \sum_{m \neq k} \mathbb{E}_k\left[\frac{1}{l_k}\frac{\partial l_k}{\partial \theta}\right]\mathbb{E}_m\left[\frac{1}{l_m}\frac{\partial l_m}{\partial \theta}\right]$$

where

$$\mathbb{E}_k\left[f(x_k)\right] = \int l_k f(x_k) dx_k$$

But

$$\mathbb{E}_k\left[\frac{1}{l_k}\frac{\partial l_k}{\partial \theta}\right] = \sum_x l_k \frac{1}{l_k}\frac{\partial l_k}{\partial \theta} = \sum_x \frac{\partial l_k}{\partial \theta} = \frac{\partial \sum_x l_k}{\partial \theta} = 0$$

And so the Fisher Information can be expressed

$$\begin{aligned} I(\theta) &= \mathbb{E}\left[\sum_k \frac{1}{l_k^2}\left(\frac{\partial l_k}{\partial \theta}\right)^2\right] \\ &= \sum_{x \geq 0} \sum_{k=1}^{k} l_k \frac{1}{l_k^2}\left(\frac{\partial l_k}{\partial \theta}\right)^2 \\ &= \sum_{x \geq 0} \sum_{k=1}^{k} \frac{1}{l_k}\left(\frac{\partial l_k}{\partial \theta}\right)^2 \end{aligned}$$

Derivatives of likelihood Eq.(1):

$$\frac{\partial l_k}{\partial \theta} = \frac{l_k(x_k - \lambda_k)}{\lambda_k}\frac{\partial \lambda_k}{\partial \theta}$$

And we get:

$$\begin{aligned} I(\theta) &= \sum_{x\geq 0}\sum_{k=1}^{k} \frac{l_k(x_k-\lambda_k)^2}{\lambda_k^2}\left(\frac{\partial \lambda_k}{\partial \theta}\right)^2 \\ &= \sum_{k=1}^{k}\frac{1}{\lambda_k^2}\left(\frac{\partial \lambda_k}{\partial \theta}\right)^2 \mathbb{E}_k\left[x_k^2 - 2x_k\lambda_k + {\lambda_k}^2\right] \end{aligned}$$

but $\mathbb{E}_k\left[x_k^2\right] = \mathrm{var}(x_k) + \mathrm{mean}(x_k)^2$ and for Poisson $\mathrm{var}(x) = \mathrm{mean}(x) = \lambda$ gives

$$\mathbb{E}_k\left[x_k^2\right] = \lambda_k + \lambda_k^2$$

and $\mathbb{E}_k\left[2x_k\lambda_k\right] = 2\lambda_k^2$ so

$$\mathbb{E}_k\left[x_k^2 - 2x_k\lambda_k + {\lambda_k}^2\right] = \lambda_k + \lambda_k^2 - 2\lambda_k^2 + \lambda_k^2 = \lambda_k$$

and

$$I(\theta) = \sum_{k=1}^{K}\frac{1}{\lambda_k}\left(\frac{\partial \lambda_k}{\partial \theta}\right)^2$$

This is the pixelised version (detection of the photons in K detectors - CCD camera and $\lambda_k = \int_{C_k}\lambda(x)dx$ where C_k is an area of the pixels of the detector).

Non pixelised version [Ram et al., 2006]

$$I(\theta) = \int \frac{1}{\lambda(x)}\left(\frac{\partial \lambda(x)}{\partial \theta}\right)^2 dx$$

1.3 Time distribution of the intensities (blinking)

For likelihood with dependent on parameter Λ_t (T different time slicese)

$$l_T(\theta, \Lambda) = \prod_{k=1}^{K}\prod_{t=1}^{T} l_k(d, \Lambda_t)p(\Lambda_t)$$

$$\mathcal{L}_T(d, \Lambda) = \sum_{k=1}^{K}\sum_{t=1}^{T}\left[\log\left(l_k(d, \Lambda_t)\right) + \log\left(p(\Lambda_t)\right)\right]$$

as $p(\Lambda)$ is not dependent on d then

$$\frac{\partial^2 \mathcal{L}_T(d, \Lambda)}{\partial d^2} = \sum_{t=1}^{T}\frac{\partial^2 \mathcal{L}(d, \Lambda_t)}{\partial d^2}$$

but in the expectation equation Eq.(2) the time dependence appears as

$$\begin{aligned} I_T(\theta) &= -\mathbb{E}_T\left[\sum_{t=1}^{T}\frac{\partial^2 \mathcal{L}(d, \Lambda_t)}{\partial d^2}\right] = \sum_{t=1}^{T} -\mathbb{E}_T\left[\frac{\partial^2 \mathcal{L}(d, \Lambda_t)}{\partial d^2}\right] = \sum_{t=1}^{T}\mathbb{E}_T\left[\left(\frac{\partial \mathcal{L}(d, \Lambda_t)}{\partial d}\right)^2\right] \\ &= \sum_{t=1}^{T}\int_{\Lambda_t} p(\Lambda_t)I(\theta)d\Lambda_t = \sum_{t,k}\int_{\Lambda_t} p(\Lambda_t)\frac{1}{\lambda_k}\left(\frac{\partial \lambda_k}{\partial d}\right)^2 d\Lambda_t \end{aligned}$$

1.4 Two sources separated by a distance d

These are comment on Fisher Information estimation as described in [Ram et al., 2006].

For two sources separated by a distance d we have a mean value of the intensity:

$$\lambda = \Lambda_1 f_1 + \Lambda_2 f_2$$

where f_i and Λ_i is the response function and intensity, respectively, of the source i. For translationally invariant PSF and in-focus sources: $f_1 = q(x - \frac{d}{2})$ and $f_2 = q(x + \frac{d}{2})$

$$\lambda(d) = \Lambda_1 q(x - \frac{d}{2}) + \Lambda_2 q(x + \frac{d}{2})$$

where q is the PSF of the sources. For pixelised version (integral over pixel area C_k)

$$\lambda_k(d) = \Lambda_1 \int_{C_k} q(x - \frac{d}{2})dx + \Lambda_2 \int_{C_k} q(x + \frac{d}{2})dx$$

so we get (as described in[Ram et al., 2006])

$$I(d) = \frac{1}{4}\sum_{k=1}^{K} \frac{\left(\Lambda_1 \int_{C_k} \partial_x q(x - \frac{d}{2})dx - \Lambda_2 \int_{C_k} \partial_x q(x + \frac{d}{2})dx\right)^2}{\Lambda_1 \int_{C_k} q(x - \frac{d}{2})dx + \Lambda_2 \int_{C_k} q(x + \frac{d}{2})dx} \tag{3}$$

Limit $d = 0$ If $\Lambda_1 = \Lambda_2$ then $I(d = 0) = 0$ which means $\mathrm{var}(d = 0) \to \infty$. (This does not hold for $\Lambda_1 \neq \Lambda_2$).

Limit $d \to \infty$ When sources are far apart then the mixing term in nominator in (3) $\Lambda_1\Lambda_2 \partial_x q(x - \frac{d}{2})\partial q(d + \frac{d}{2}) = 0$ as the $\partial_x q(x - \frac{d}{2})$ ($q(x - \frac{d}{2})$) and $\partial_x q(x + \frac{d}{2})$ ($q(x + \frac{d}{2})$) do not have any overlap. The (3) then decomposes into two individual terms (sum of Fisher Information for localisation of individual sources.)

$$\begin{aligned} I(d) &= \frac{1}{4}\sum_{k=1}^{K} \left[\frac{\left(\Lambda_1 \int_{C_k} \partial_x q(x - \frac{d}{2})dx\right)^2}{\Lambda_1 \int_{C_k} q(x - \frac{d}{2})dx} + \frac{\left(\Lambda_2 \int_{C_k} \partial_x q(x + \frac{d}{2})dx\right)^2}{\Lambda_2 \int_{C_k} q(x + \frac{d}{2})dx} \right] \\ &= \frac{1}{4}\sum_{k=1}^{K} \frac{\left(\int_{C_k} \partial_x q(x)dx\right)^2}{\int_{C_k} q(x)dx} [\Lambda_1 + \Lambda_2] \end{aligned}$$

Limit $\Lambda_i = 0$ If $\Lambda_1 = 0$ or $\Lambda_2 = 0$ $I(d) \neq 0$. So the variance is finite even if one of the sources is not present.

1.5 An alternative way to derive Fisher information for two sources separated by d:

These is a suggestion how to fix the problems with limits for Fisher Information derived above. This gives finite variance even for $d = 0$ and infinite variance when one of the sources is no present.

For two sources $f_1 = q(x - c_1)$ and $f_2 = q(x - c_2)$ we have $\lambda = \Lambda_1 f_1 + \Lambda_2 f_2$. The distance between the two sources is $d = |c_1 - c_2|$ and

$$\mathrm{var}(d) = \mathrm{var}(c_1) + \mathrm{var}(c_2) = \frac{1}{I(c_1)} + \frac{1}{I(c_2)} \tag{4}$$

$$I(c_1) = \sum_{k=1}^{K} \frac{1}{\lambda_k}\left(\frac{\partial \lambda_k}{\partial c_1}\right)^2 = \sum_{k=1}^{K} \frac{\left(\Lambda_1 \int_{C_k} \partial_x q(x - c_1)dx\right)^2}{\Lambda_1 \int_{C_k} q(x - c_1)dx + \Lambda_2 \int_{C_k} q(x - c_2)dx}$$

if this keeps translational invariance (non-pixelised version does as $\int_{\mathbb{R}} g(x + c)dx = \int_{\mathbb{R}} g(x)dx$) then

$$I(c_1) = \sum_{k=1}^{K} \frac{\left(\Lambda_1 \int_{C_k} \partial_x q(x)dx\right)^2}{\Lambda_1 \int_{C_k} q(x)dx + \Lambda_2 \int_{C_k} q(x + d)dx}$$

where $d = c_1 - c_2$ and

$$I(c_2) = \sum_{k=1}^{K} \frac{\left(\Lambda_2 \int_{C_k} \partial_x q(x) dx\right)^2}{\Lambda_2 \int_{C_k} q(x) dx + \Lambda_1 \int_{C_k} q(x-d) dx}$$

For symmetrical PSF $q(x-d) = q(x+d)$ we have

$$I(c_i) = \sum_{k=1}^{K} \frac{\left(\Lambda_i \int_{C_k} \partial_x q(x) dx\right)^2}{\Lambda_i \int_{C_k} q(x) dx + \Lambda_j \int_{C_k} q(x-d) dx} \tag{5}$$

Limit $d = 0$ For $d = 0$ we have

$$I(c_i) = \frac{\Lambda_i^2}{\Lambda_i + \Lambda_j} \sum_{k=1}^{K} \frac{\left(\int_{C_k} \partial_x q(x) dx\right)^2}{\int_{C_k} q(x) dx} \neq 0$$

For non-pixelised version, $\Lambda_1 = \Lambda_2 = \Lambda$ and for Gaussian approximation of the PSF ($q(x-a) \propto \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$ (with $\sigma = \frac{\sqrt{2}}{2\pi}\frac{\lambda}{NA}$ [Zhang et al., 2007]) we have $\partial_x q(x) = \frac{1}{\sigma^2} x q(x)$ and for $d = 0$ we get

$$I(c_i) = \frac{1}{2}\frac{\Lambda}{\sigma^4}\int_{\mathbb{R}} x^2 q(x) dx = \frac{1}{2}\frac{\Lambda}{\sigma^4}(\mathrm{var}(q(x)) - \mathrm{mean}(q(x))^2) = \frac{1}{2}\frac{\Lambda}{\sigma^4}\sigma^2 = \frac{1}{2}\frac{\Lambda}{\sigma^2} \neq 0$$

Limit $d \to \infty$

$$I(c_i) = \Lambda \int \frac{(\Lambda_i \partial_x q(x))^2}{\Lambda_i q(x) + \Lambda_j q(x-d)} dx = \Lambda_i \int_{\mathbb{R}} \frac{(\partial_x q(x))^2}{q(x)} dx$$

as the PSF $q(x)$ (and also $\partial_x q(x)$) have a finite support, if d is big, $q(x-d)$ is outside the support of the $\partial_x q(x)$. They have no overlap so it doesn't have any effect in the denominator.

For non-pixelised version, $\Lambda_1 = \Lambda_2 = \Lambda$ and for Gaussian approximation of the PSF:

$$I(d \to \infty) = \frac{\Lambda}{\sigma^2} = 2I(d = 0)$$

Limit $\Lambda_i = 0,\ \Lambda_j \neq 0$ $I(c_i) = 0,\ I(c_j) \neq 0$ and for variance (4) $\mathrm{var}(d) = \frac{1}{I(c_1)} + \frac{1}{I(c_2)} \to \infty$.

References

[Ram et al., 2006] Ram, S., Ward, E. S., and Ober, R. J. (2006). Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4457–62.

[Zhang et al., 2007] Zhang, B., Zerubia, J., and Olivo-Marin, J. (2007). Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics*, 46(10):1819–1829.