1. Start w/ Poisson distribution:

$$P = \frac{\lambda^x e^{-\lambda}}{x!}$$

using Stirling's approximation:

$$n! \sim n^n e^{-n} \sqrt{2\pi n}$$

re-write factorial:

$$P = \frac{\lambda^x e^{-\lambda}}{x^x e^{-x} \sqrt{2\pi x}} = e^{(x-\lambda)} \left(\frac{\lambda}{x}\right)^x \frac{1}{\sqrt{2\pi x}}$$

$$P = e^{(x-\lambda)} \left(e^{\ln(\lambda/x)}\right)^x \frac{1}{\sqrt{2\pi x}} = e^{(x-\lambda)} e^{x \ln\left(\frac{\lambda}{x}\right)} \frac{1}{\sqrt{2\pi x}}$$

let $\alpha = x - \lambda$

$x = \alpha + \lambda$

$$P = e^{\alpha} e^{(\alpha+\lambda)\ln\left(\frac{\lambda}{\alpha+\lambda}\right)} \frac{1}{\sqrt{2\pi(\alpha+\lambda)}}$$

using identity of $\ln\left(\frac{1}{x}\right) = -\ln(x)$:

$$P = e^{\alpha} e^{-(\alpha+\lambda)\ln\left(\frac{\alpha+\lambda}{\lambda}\right)} \frac{1}{\sqrt{2\pi(\alpha+\lambda)}}$$

further simplify:

$$-(\alpha+\lambda)\ln\left(\frac{\alpha}{\lambda}+1\right)$$

using log expansion $\ln(x+1) \sim x - \frac{1}{2}x^2$ for $-1 < x \leq 1$

since $\frac{\alpha}{\lambda}$ ~~is very small with large λ, this is valid~~

$$-(\alpha+\lambda)\ln\left(\frac{\alpha}{\lambda}+1\right) \sim -(\alpha+\lambda)\left(\frac{\alpha}{\lambda} - \frac{1}{2}\left(\frac{\alpha}{\lambda}\right)^2\right)$$

$$= \lim_{\lambda\to\infty}\left(-\frac{\alpha^2}{\lambda} + \cancelto{0}{\frac{1}{2}\frac{\alpha^3}{\lambda^2}} - \alpha + \frac{1}{2}\frac{\alpha^2}{\lambda}\right)$$

$$= -\frac{1}{2}\frac{\alpha^2}{\lambda} - \alpha$$

$$P = e^{\alpha} e^{-\left(\frac{1}{2}\frac{\alpha^2}{\lambda} + \alpha\right)} \underbrace{\frac{1}{\sqrt{2\pi(\alpha+\lambda)}}}$$

$$\lim_{\lambda \to \infty} \lambda\left(\cancel{\frac{\alpha}{\lambda}}^{\nearrow 0} + 1\right) = \lambda$$

$$P = e^{-\frac{1}{2}\frac{\alpha^2}{\lambda}} \frac{1}{\sqrt{2\pi\lambda}}$$

plug back in

true gaussian:

$$\boxed{P \approx e^{-(x-\lambda)^2/2\lambda} \frac{1}{\sqrt{2\pi\lambda}}}$$

$$G = e^{-(x-\mu)^2/2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}}$$

we thus recover the gaussian distribution and see that the mean $\mu = \lambda$ and $\sigma^2 = \lambda$ which we'll use in problem 2

3. Starting w/ a weighted mean:

$$\mu = \frac{\sum_{i=1}^{n} \omega_i x_i}{\sum_{i=1}^{n} \omega_i} \qquad \text{where} \qquad \omega_i = \frac{1}{\sigma_i^2}$$

if σ is the same for all points we can find the variance of μ:

$$\sigma_\mu^2 = var(\mu) = var\left(\frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i}{n/\sigma^2}\right)$$

property of variance

$var(kX) = k^2 var(x)$

if k is constant

$$= \left(\frac{1}{n/\sigma^2}\right)^2 \frac{1}{\sigma^4} var\left(\sum x_i\right)$$

$$= \left(\frac{\sigma^4}{n^2}\right)\left(\frac{1}{\sigma^4}\right)(n\sigma^2)$$

$$\boxed{\sigma_\mu^2 = \frac{\sigma^2}{n}}$$

$\rightarrow$ a result we expect from error propagation

if half the points' variance is off by 2 the result changes to:

$$\mu = \frac{\sum_{i=1}^{n/2} \frac{1}{\sigma_i^2} x_i + \sum_{n/2+1}^{n} \frac{1}{2\sigma_i^2} x_i}{\frac{n}{2}\frac{1}{\sigma^2} + \frac{n}{2}\left(\frac{1}{2\sigma^2}\right)}$$

$$\sigma_\mu^2 = var(\mu) = \left(\frac{n}{2}\frac{1}{\sigma^2} + \frac{n}{2}\left(\frac{1}{2\sigma^2}\right)\right)^{-2}\left[\frac{1}{\sigma^4} var\left(\sum_{i=1}^{n/2} x_i\right) + \frac{1}{4\sigma^4} var\left(\sum_{n/2+1}^{n} x_i\right)\right]$$

$$var\left(\sum x_i\right) = \sum var(x_i)$$

$$\sigma^2 = \left(\frac{n}{2}\frac{1}{\sigma^2} + \frac{n}{2}\left(\frac{1}{2\sigma^2}\right)\right)^{-2} \underbrace{\left(\frac{1}{\sigma^4}\left(\frac{n}{2}\right)\sigma^2 + \frac{1}{4\sigma^4}\left(\frac{n}{2}\right)2\sigma^2\right)}_{\text{this simplifies to}}$$

$$\left(\frac{n}{2}\frac{1}{\sigma^2} + \frac{n}{2}\frac{1}{2\sigma^2}\right)$$

$$\sigma_M^2 = \left(\frac{n}{2}\frac{1}{\sigma^2} + \frac{n}{2}\left(\frac{1}{2\sigma^2}\right)\right)^{-1}$$

$$= \left(\frac{2n}{4\sigma^2} + \frac{n}{4\sigma^2}\right)^{-1}$$

$$\boxed{\sigma_M^2 = \frac{4\sigma^2}{3n}}$$

If we got all the data right, we would have gotten 1 versus $\frac{4}{3}$ of $\frac{\sigma^2}{n}$, so we see a larger variance if we are off on the error

If we underweight 1% of data by ~100:

$$\mu = \left(\frac{\sum \frac{1}{\sigma^2} x_i + \sum \frac{1}{100\sigma^2} x_i}{\frac{1}{\sigma^2}\frac{99}{100}n + \frac{1}{100\sigma^2}\left(\frac{n}{100}\right)}\right)$$

$$\sigma_M^2 = \mathrm{var}(\mu) = \left(\frac{1}{\sigma^2}\frac{99}{100}n + \frac{1}{100\sigma^2}\left(\frac{n}{100}\right)\right)^{-2} \underbrace{\left[\frac{1}{\sigma^4}\left(\frac{99}{100}n\sigma^2\right) + \frac{1}{100^2\sigma^4}\left(\frac{n}{100}100\sigma^2\right)\right]}_{= \left(\frac{99}{100}n\frac{1}{\sigma^2} + \frac{n}{100^2}\frac{1}{\sigma^2}\right)}$$

$$\sigma_M^2 = \left(\frac{1}{\sigma^2}\frac{99}{100}n + \frac{n}{100^2}\frac{1}{\sigma^2}\right)^{-1}$$

$$= \left(\frac{9{,}900n}{100^2\sigma^2} + \frac{n}{100^2\sigma^2}\right)^{-1}$$

$$\boxed{\sigma_M^2 = \frac{100^2\sigma^2}{9{,}901n} = \frac{100}{99.01}\left(\frac{\sigma^2}{n}\right)}$$

if we overweighting 1% of data by ~100:

$$\mu = \left(\frac{\sum \frac{1}{\sigma^2} x_i + \sum' \frac{100}{\sigma^2} x_i}{\frac{1}{\sigma^2}\frac{99}{100} n + \frac{100}{\sigma^2}\left(\frac{n}{100}\right)} \right)$$

$$\text{var}(\mu) = \left(\frac{1}{\sigma^2}\frac{99}{100}n + \frac{1}{\sigma^2}n\right)^{-2} \underbrace{\left(\frac{1}{\sigma^4}\left(\frac{99}{100}n\sigma^2\right) + \frac{100^2}{\sigma^4}\left(\frac{n}{100}\frac{\sigma^2}{100}\right)\right)}_{=\left(\frac{1}{\sigma^2}\frac{99}{100}n + \frac{1}{\sigma^2}n\right)}$$

$$\sigma_M^2 = \left(\frac{1}{\sigma^2}\frac{99}{100}n + \frac{1}{\sigma^2}n\right)^{-1}$$

$$= \left(\frac{99n}{100\sigma^2} + \frac{100n}{100\sigma^2}\right)^{-1}$$

$$\boxed{\sigma_M^2 = \frac{100}{199}\left(\frac{\sigma^2}{n}\right)}$$

You should be more worried about ~~over~~ weighting 1% of your data rather than underweighting.

It is also worse to get a small error on more of the data than a large error on a small % of the data.

5. Find the following:

$$\hat{N}_{ij} = \langle \hat{n}_i \hat{n}_j \rangle$$

Starting w/ $\hat{N}_{ij}$

introducing the new variable $\hat{N}$:

$\hat{N} = SNS^T$ where S is an invertible matrix and N is the diagonal noise matrix

breaking into components for inner product:

$$\hat{N}_{ij} = \sum_k (SN)_{ik} S^T_{kj}$$

Summed over some dummy variable k

We can take the transpose of S_{kj}:

$$S^T_{kj} = S_{jk}$$

$$\hat{N}_{ij} = \sum_k (SN)_{ik} S_{jk}$$

and $(SN)_{ik} = S_{ik} \sigma_k^2$

$N = \begin{pmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \ddots \end{pmatrix}$

this is because N is a diagonal matrix so for a given row or column there is a single entry of the variance, or σ^2. Therefore the only index necessary is the dummy variable k.

$$\hat{N}_{ij} = \sum_k S_{ik} S_{jk} \sigma_k^2$$

next we can look at $\langle \hat{n}_i \hat{n}_j \rangle$:

$\hat{n} = Sn$ we can take the same approach as before since n isn't a matrix like N

$$\hat{n}_i = \sum_k S_{ik} n_k$$

$$\hat{n}_j = \sum_l S_{jl} n_l$$

$$\langle \hat{n}_i \hat{n}_j \rangle = \left\langle \sum_k S_{ik} n_k \sum_l S_{jl} n_l \right\rangle$$

unless $k=l$ all the cross terms are 0:

$$\langle \hat{n}_i \hat{n}_j \rangle = \left\langle \sum_k S_{ik} S_{jk} n_k^2 \right\rangle = \sum_k S_{ik} S_{jk} \sigma_k^2$$

we can make this jump because the expectation of the noise squared approaches the variance

now we see that:

$$\langle \hat{n}_i \hat{n}_j \rangle = \sum_k S_{ik} S_{jk} \sigma_k^2 = \hat{N}_{ij}$$

$$\hat{N}_{ij} = \langle \hat{n}_i \hat{n}_j \rangle \checkmark$$

So our expression for χ^2 is still valid:

$$\chi^2 = (d-Am)^T N^{-1} (d-Am)$$

even for correlated noise (and non-liner modus $A(m)$)