Homework #1

Observational Techniques of Modern Astrophysics — PHYS 641

Fall 2018

Bridget Andersen

September 19, 2018

Problem 1

Show that a Poisson distribution converges to a Gaussian in the limit of large λ. Hints: use Stirlings approximation plus use the first two terms in a log expansion.
Solution: The Poisson distribution is given by:

$$P_P = \frac{e^{-\lambda}\lambda^k}{k!} \tag{1}$$

where λ is the expected number of events per interval and k is the actual number of events you get in the given interval. Note that the mean and the variance of the Poissonian are given by $\mu = \sigma^2 = \lambda$. Therefore, we will try to show that, in the limit of large λ, the Poisson distribution becomes the Gaussian:

$$P = \frac{e^{-(k-\lambda)^2/2\lambda}}{\sqrt{2\pi\lambda}} \tag{2}$$

Now let's set about doing this. In the limit of large λ, k essentially becomes a continuous variable, and we can write $k = \lambda\,(1+\delta)$ where δ is the variation from the mean. In this case, $\delta \ll \lambda$ and $k \approx \lambda$. First, we can use the Stirling Approximation for the factorial in the Poisson denominator:

$$n! \approx n^n e^{-n}\sqrt{2\pi n} \tag{3}$$

So we have:

$$P = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-\lambda}\lambda^k}{\sqrt{2\pi k}e^{-k}k^k} \tag{4}$$

Now we plug in $k = \lambda\,(1+\delta)$ and reduce:

$$\begin{aligned} P &= \frac{e^{-\lambda}\lambda^{\lambda(1+\delta)}}{\sqrt{2\pi\lambda\,(1+\delta)}e^{-\lambda(1+\delta)}\left(\lambda\,(1+\delta)\right)^{\lambda(1+\delta)}} \\ &= \frac{\cancel{e^{-\lambda}}\,\cancel{\lambda^{\lambda(1+\delta)}}}{\sqrt{2\pi\lambda}\cancel{e^{-\lambda}}e^{-\lambda\delta}\cdot\cancel{\lambda^{\lambda(1+\delta)}}\,(1+\delta)^{\lambda(1+\delta)+\frac{1}{2}}} \\ &= \frac{e^{\lambda\delta}\,(1+\delta)^{-\lambda(1+\delta)-\frac{1}{2}}}{\sqrt{2\pi\lambda}} \end{aligned}$$

Now, to clarify the use of the log expansion, we take the log of both sides of this equation:

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

Now, since δ is very small compared to λ, we can use the Taylor series expansion:

$$\ln(1+x) \approx x - \frac{x^2}{2} + \mathcal{O}\left(x^3\right) \tag{5}$$

And so we have:

$$\begin{aligned}
\ln(P) &= \ln\left(\frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}}\right) - \left(\lambda(1+\delta) + \frac{1}{2}\right)\left(\delta - \frac{\delta^2}{2}\right) \\
&= \ln\left(\frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}}\right) - \left(\lambda + \lambda\delta + \frac{1}{2}\right)\left(\delta - \frac{\delta^2}{2}\right)
\end{aligned}$$

Keeping only the $\mathcal{O}\left(\lambda\delta^2\right)$ terms:

$$\begin{aligned}
\ln(P) &= \ln\left(\frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}}\right) - \left(\lambda\delta - \frac{\lambda\delta^2}{2} + \lambda\delta^2\right) \\
&= \ln\left(\frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}}\right) - \left(\lambda\delta + \frac{\lambda\delta^2}{2}\right)
\end{aligned}$$

Now take the exponential of both sides:

$$\begin{aligned}
P &= e^{\ln\left(\frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}}\right) - \left(\lambda\delta + \frac{\lambda\delta^2}{2}\right)} = \frac{\cancel{e^{\lambda\delta}}}{\sqrt{2\pi\lambda}} \cdot \cancel{e^{-\lambda\delta}} \cdot e^{\frac{-\lambda\delta^2}{2}} = \frac{e^{-\cancel{\lambda}(k-\lambda)^2/2\lambda\cancel{^2}}}{\sqrt{2\pi\lambda}} \\
&\implies \boxed{P_G = \frac{e^{-(k-\lambda)^2/2\lambda}}{\sqrt{2\pi\lambda}}}
\end{aligned}$$

So we have shown that, in the limit of large λ, the Poisson distribution converges to a Gaussian with a mean and variance of λ.

Problem 2

The gold standard for a believable result is usually 5σ. Lets define the Gaussian approximation as "good enough" if it agrees with the Poisson to within a factor of 2. How large does n need to be for the Gaussian to be good enough at 5σ? How about at 3σ?

Solution: If λ (the expected number of events) is large enough, then $\lambda \sim n$ (the number of observed events). We can use our formalism from Problem 1 to set $P_P = 2P_G$ and solve for λ when $k = \mu + \sigma'$, where μ is the mean and σ' is the deviation that we want to test (either $\sigma' = 3\sigma$ or 5σ). Note that, in the limit of large λ, the variance of both the Poissonian and Gaussian is $\sigma^2 = \lambda$ and the mean is $\mu = \lambda$. Then the standard deviation is $\sigma = \sqrt{\lambda}$. Therefore, we have:

$$P_P = 2P_G \implies \frac{e^{-\lambda}\lambda^k}{k!} = 2\frac{e^{-(k-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

Now we plug in $\mu = \lambda$, $k = \mu + \sigma' = \lambda + \sigma'$, and $\sigma^2 = \lambda$:

$$\frac{e^{-\lambda}\lambda^{\lambda+\sigma'}}{(\lambda+\sigma')!} = 2\frac{e^{-(\lambda+\sigma'-\lambda)^2/2\lambda}}{\sqrt{2\pi\lambda}} \implies \frac{e^{-\lambda}\lambda^{\lambda+\sigma'}}{(\lambda+\sigma')!} = 2\frac{e^{-\sigma'^2/2\lambda}}{\sqrt{2\pi\lambda}} \tag{6}$$

Finally we can plug in $\sigma' = 3\sigma = 3\sqrt{\lambda}$:

$$\frac{e^{-\lambda}\lambda^{\lambda+3\sqrt{\lambda}}}{\left(\lambda+3\sqrt{\lambda}\right)!} = 2\frac{e^{-\left(3\sqrt{\lambda}\right)^2/2\lambda}}{\sqrt{2\pi\lambda}} \implies \frac{e^{-\lambda}\lambda^{\lambda+3\sqrt{\lambda}}}{\left(\lambda+3\sqrt{\lambda}\right)!} = 2\frac{e^{-9/2}}{\sqrt{2\pi\lambda}} \tag{7}$$

Solving this numerically, we find: $\boxed{n = \lambda = 8}$. Now, plugging in $\sigma' = 5\sigma = 5\sqrt{\lambda}$:

$$\frac{e^{-\lambda}\lambda^{\lambda+5\sqrt{\lambda}}}{\left(\lambda+5\sqrt{\lambda}\right)!} = 2\frac{e^{-25/2}}{\sqrt{2\pi\lambda}} \tag{8}$$

Solving this numerically, we find: $\boxed{n = \lambda = 575}$.

Problem 3

Lets say we have n Gaussian-distributed data points with identical standard deviations σ, and identical but unknown mean. What is the error on the maximum-likelihood estimate of the mean? Now lets say we got the errors on half the data wrong by a factor of $\sqrt{2}$ (so the variance is off by a factor of 2). What is the true error on the new non-optimal mean, and how does it compare to the maximum-likelihood you could have gotten had you gotten the noises right? How about if you underweight 1% of the data by a factor of ~ 100? And if you overweight 1% of the data by a factor of 100? What type of mistake in weighting your data should you be most concerned about?

Solution: So, first we assume that the each of the data points comes from a Gaussian probability distribution with the same standard deviation σ and the same mean μ. The probability P of observing all of the n data points is found by multiplying all of their individual probability distributions P_i together:

$$P = \prod_{i=1}^{n} P_i = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2\right) \tag{9}$$

The most probable mean μ is the one that gives the maximum value for the probability P. To maximize this probability, we simply need to minimize the argument of the exponent above:

$$X = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$\frac{dX}{d\mu} = \frac{d}{d\mu}\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2\right) = -\frac{1}{2}\sum_{i=1}^{n}\frac{d}{d\mu}\left(\frac{x_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma^2}\right) = 0$$

$$\implies \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\mu = 0 \implies N\mu = \sum_{i=1}^{n} x_i \implies \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Which is just the standard equation for the mean. To find the error on this estimate, we simply propagate the original uncertainty:

$$
\begin{aligned}
\sigma_r^2 &= \sum_{i=1}^{n}\left[\sigma^2\left(\frac{\partial\mu}{\partial x_i}\right)^2\right] = \sum_{i=1}^{n}\left[\sigma^2\left(\frac{\partial}{\partial x_i}\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\right)^2\right] \\
&= \sum_{i=1}^{n}\sigma^2\left(\frac{1}{n}\right)^2 = \cancel{n}\sigma^2\left(\frac{1}{n^{\cancel{2}}}\right) = \frac{\sigma^2}{n} \implies \boxed{\sigma_\mu = \frac{\sigma}{\sqrt{n}}}
\end{aligned}
$$

Now we consider the case where we got the errors on part of the data wrong. To determine this, we can rederive the error on the maximum likelihood in the more general case where each data point comes from a distribution with a different standard deviation σ_i. In this case, we have the same equation for the joint probability (equation 9), except that our original constant σ becomes σ_i. Then, to find the new equation for the maximum-likelihood estimate of the mean, we minimize the exponent:

$$
\begin{aligned}
X &= -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma_i}\right)^2 \\
\frac{dX}{d\mu} &= \frac{d}{d\mu}\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma_i}\right)^2\right) = -\frac{1}{2}\sum_{i=1}^{n}\frac{d}{d\mu}\left(\frac{x_i-\mu}{\sigma_i}\right)^2 = \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma_i^2}\right) = 0 \\
&\implies \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} = \sum_{i=1}^{n}\frac{\mu}{\sigma_i^2} \implies \mu = \frac{\sum_{i=1}^{n}\left(x_i/\sigma_i^2\right)}{\sum_{i=1}^{n}\left(1/\sigma_i^2\right)}
\end{aligned}
$$

This is just the weighted average. To find the error, we use the same procedure as before. In this case, the derivative of the mean is:

$$
\begin{aligned}
\frac{\partial\mu}{\partial x_i} &= \frac{\partial}{\partial x_i}\left(\frac{\sum_{i=1}^{n}\left(x_i/\sigma_i^2\right)}{\sum_{i=1}^{n}\left(1/\sigma_i^2\right)}\right) = \frac{1/\sigma_i^2}{\sum_{i=1}^{n}\left(1/\sigma_i^2\right)} \\
\sigma_\mu^2 &= \sum_{i=1}^{n}\left[\sigma_i^2\left(\frac{\partial\mu}{\partial x_i}\right)^2\right] = \sum_{i=1}^{n}\left[\sigma_i^{\cancel{2}}\cdot\frac{1/\sigma_i^{\cancel{4}\,2}}{\left(\sum_{i=1}^{n}\left(1/\sigma_i^2\right)\right)^2}\right] = \frac{\cancel{\sum_{i=1}^{n}1/\sigma_i^2}}{\left(\sum_{i=1}^{n}1/\sigma_i^2\right)^{\cancel{2}}} = \frac{1}{\sum_{i=1}^{n}1/\sigma_i^2}
\end{aligned}
$$

Now we let half of the errors be wrong by a factor of $\sqrt{2}$. Then we can write, $\sigma_i^2 \to \sigma_i^2/2$:

$$
\sigma_w^2 = \frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}} = \frac{1}{\sum_{i=1}^{n/2}\frac{1}{\sigma^2}+\sum_{i=n/2}^{n}\frac{2}{\sigma^2}} = \frac{1}{\frac{n}{2}\left(\frac{1}{\sigma^2}+\frac{2}{\sigma^2}\right)} = \frac{2}{n}\cdot\frac{\sigma^2}{3} = \frac{2}{3}\frac{\sigma^2}{n}
$$

Then the ratio of the "wrong" variance to the "right" variance is:

$$
\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{r}}^2} = \frac{2\sigma^2/3n}{\sigma^2/n} = \frac{2}{3} = 0.67 \tag{10}
$$

So the variance is down by 33% from what we would expect if we did everything correctly. Now we underweight 1% of the data by a factor of 100. By "weight" here we mean $w_i = 1/\sigma_i^2$. So underweighting would mean that $w_i \to w_i/100$ so that $\sigma_i^2 \to 100\sigma_i^2$:

$$
\begin{aligned}
\sigma_w^2 &= \frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}} = \frac{1}{\sum_{i=1}^{n/100}\frac{1}{100\sigma^2}+\sum_{i=n/100}^{n}\frac{1}{\sigma^2}} = \frac{1}{\frac{n}{100}\left(\frac{1}{100\sigma^2}\right)+\frac{99n}{100}\left(\frac{1}{\sigma^2}\right)} \\
&= \frac{1}{\frac{n}{10000\sigma^2}+\frac{99n}{100\sigma^2}} = \frac{1}{\frac{n+9900n}{10000\sigma^2}} = \frac{1}{\frac{9901n}{10000\sigma^2}} = \frac{10000\sigma^2}{9901n}
\end{aligned}
$$

Then the ratio of the "wrong" variance to the "right" variance is:

$$\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{r}}^2} = \frac{10000\sigma^2/9901n}{\sigma^2/n} = \frac{10000}{9901} = 1.01 \tag{11}$$

So the variance is up by 1% from what we would expect if we did everything correctly. Now we overweight 1% of the data by a factor of 100 so that $\sigma_i^2 \to \sigma_i^2/100$:

$$\begin{aligned}\sigma_w^2 &= \frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}} = \frac{1}{\sum_{i=1}^{n/100}\frac{100}{\sigma^2} + \sum_{i=n/100}^{n}\frac{1}{\sigma^2}} = \frac{1}{\frac{n}{100}\left(\frac{100}{\sigma^2}\right) + \frac{99n}{100}\left(\frac{1}{\sigma^2}\right)} \\ &= \frac{1}{\frac{100n+99n}{100\sigma^2}} = \frac{1}{\frac{199n}{100\sigma^2}} = \frac{100\sigma^2}{199n}\end{aligned}$$

Then the ratio of the "wrong" variance to the "right" variance is:

$$\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{r}}^2} = \frac{100\sigma^2/199n}{\sigma^2/n} = \frac{100}{199} = 0.50 \tag{12}$$

So the variance is down by 50% from what we would expect if we did everything correctly! Clearly, you should be most worried about overweighting your data.

Problem 4

In linear least-squares, our estimate for fit parameters $\hat{m}$ is unbiased if $\langle \hat{m} \rangle = m_{true}$. If our model is correct, $\langle d \rangle = Am$, then show that the least-squares solution is unbiased. Show that this result does not depend on our noise matrix N actually being the noise in the data.

Solution: If our model is truly correct, then we would have $\langle d \rangle = Am_{true}$. From our class derivations, the least-squares maximum likelihood estimation of the parameters is given by:

$$\hat{m} = \left(A^T N^{-1} A\right)^{-1} A^T N^{-1} d \tag{13}$$

Therefore, we can calculate the expectation of our predicted parameters to be:

$$\begin{aligned}\langle \hat{m} \rangle &= \langle \left(A^T N^{-1} A\right)^{-1} A^T N^{-1} d \rangle = \left(A^T N^{-1} A\right)^{-1} A^T N^{-1} \langle d \rangle \\ &= \cancel{\left(A^T N^{-1} A\right)^{-1}}\, \cancel{A^T N^{-1} A} m_{true} = \boxed{m_{true}}\end{aligned}$$

Therefore, if our model is correct, then the least-squares solution is unbiased.

Problem 5

The preceding statement comes with an important caveat, namely that our noise estimate is not correlated with any residual signal in the data. Write a computer program that generates random Gaussian noise (`numpy.random.randn` may come in handy here), and adds a template (possibly a Gaussian as would be typical for a source seen by a telescope with a finite-resolution beam, but the details arent important) to it. Estimate the noise by assuming its constant and equal to the scatter in the observed data, which has the template added to it. Show that the least-squares estimate for any individual chunk is unbiased, but that the least-squares estimate for many data chunks analyzed jointly is biased low. Basically, your program should fit an amplitude and error to each individual chunk, then use that to get an overall amplitude/error. How might you go about mitigating this bias? Note that this is an extremely common situation when you say observed the same field/source several times and want to make your best estimate of what you have seen.

Solution:

Problem 6 - bonus

In class it was asserted that adding orthogonal matrices into the expression for χ^2 let us work with correlated data. In particular, show that:

$$\chi^2 = \delta^T V^T V N^{-1} V^T V \delta \tag{14}$$

where $\delta_i = d_i - \langle d_i \rangle$, is equivalent to:

$$\chi^2 = \tilde{\delta}^T \tilde{N}^{-1} \tilde{\delta} \tag{15}$$

and that:

$$\tilde{N}_{ij} = \langle \tilde{\delta}_i \tilde{\delta}_j \rangle \tag{16}$$

Solution: