PHYS 641 Assignment 1

Andrew V. Zwaniga

Due Sept 19 2018

1 Poisson distribution

Let X be a Poisson random variable with $\mathrm{E}[X] = \lambda$. We show that for large λ, the distribution of X tends towards a Gaussian distribution with $\mathrm{E}[X] = \lambda$ and $\mathrm{Var}[X] = \lambda$. The probability that X takes the value $X = x$ is given by

$$P(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{1}$$

A posteriori, for $\lambda >> 1$, $P(x, \lambda)$ is small when $|x - \lambda|$ is large. Therefore, the trick is to choose a parameter δ such that $|\delta| << 1$ and focus only on values of x such that

$$x = \lambda(1 + \delta) \iff \delta = \frac{x - \lambda}{\lambda} \tag{2}$$

From here we substitute into (1) and apply Stirling's approximation, which holds that for N large we have $N! \approx N^N e^N \sqrt{2\pi N}$.

$$\begin{aligned}
P &= \frac{\lambda^{\lambda(1+\delta)e^{-\lambda}}}{[\lambda(1+\delta)]!} \\
\implies P &\approx \frac{\lambda^{\lambda(1+\delta)} e^{-\lambda}}{[\lambda(1+\delta)]^{\lambda(1+\delta)} \exp[-\lambda(1+\delta)] \sqrt{2\pi\lambda(1+\delta)}} \\
&= \frac{\lambda^{\lambda(1+\delta)}}{\sqrt{2\pi}[\lambda(1+\delta)]^{\lambda(1+\delta)+\frac{1}{2}} \exp[-\lambda(1+\delta)]} \\
&= \frac{1}{\sqrt{2\pi\lambda}} \, \frac{\exp[-\lambda + \lambda + \lambda\delta]}{(1+\delta)^{\lambda(1+\delta)+\frac{1}{2}}} \\
\implies P &\approx \frac{1}{\sqrt{2\pi\lambda}} \, \exp[\lambda\delta](1+\delta)^{-\lambda(1+\delta)-\frac{1}{2}}
\end{aligned}$$

One more approximation is needed to complete the proof. The mysterious term $(1+\delta)^{-\lambda(1+\delta)-\frac{1}{2}}$ can be approximated for $\lambda >> 1$ and $|\delta| << 1$ by using the Taylor series expansion of $\log(1 + t)$ for $|t| < 1$:

$$\log(1+t) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} t^k = t - \frac{1}{2}t^2 + \frac{1}{3}t^3 - \frac{1}{4}t^4 + \cdots \approx t - \frac{1}{2}t^2 \tag{3}$$

We get

$$\begin{aligned}
\log[(1+\delta)^{-\lambda(1+\delta)-\frac{1}{2}}] &= [-\lambda(1+\delta) - \tfrac{1}{2}] \log(1+\delta) \\
&\approx [-\lambda - \lambda\delta - \tfrac{1}{2}](\delta - \tfrac{1}{2}\delta^2) \\
&= -\lambda\delta + \tfrac{1}{2}\lambda\delta^2 - \delta^2\lambda + \tfrac{1}{2}\lambda\delta^3 - \tfrac{1}{2}\delta + \tfrac{1}{4}\delta^2
\end{aligned}$$

Now, as λ is taken to be large and δ is taken to be small, in accordance with the second-order approximation of the logarithm $\log(1+\delta)$, we only take binomials from the above that have total degree at most 3; the other binomials are assumed to be vanishing. We get

$$\begin{aligned}
\log[(1+\delta)^{-\lambda(1+\delta)-\frac{1}{2}}] &\approx -\lambda\delta - \tfrac{1}{2}\lambda\delta^2 \\
\implies (1+\delta)^{-\lambda(1+\delta)-\frac{1}{2}} &\approx \exp(-\lambda\delta - \tfrac{1}{2}\lambda\delta^2) \\
&= \exp(-\lambda\delta)\exp(-\tfrac{1}{2}\lambda\delta^2)
\end{aligned}$$

Putting this into our latest result for P we get

$$\begin{aligned}
P &\approx \frac{1}{\sqrt{2\pi\lambda}} \exp(\lambda\delta)\exp(-\lambda\delta)\exp(-\tfrac{1}{2}\delta^2) \\
&= \frac{1}{\sqrt{2\pi\lambda}} \exp(-\tfrac{1}{2}\lambda\delta^2)
\end{aligned}$$

and at last by replacing our placeholder $\delta^2 = \frac{(x-\lambda)^2}{\lambda^2}$ we arrive at

$$P \approx \frac{1}{\sqrt{2\pi\lambda}} \exp\Big[-\frac{(x-\lambda)^2}{2\lambda}\Big] \tag{4}$$

In summary, we find that for a random variable X that follows a Poisson distribution with a large expectation value λ, the distribution is actually Gaussian with mean λ and variance λ.

2 Approximating Poisson with Gaussian

Let $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[-\frac{(x-\mu)^2}{2\sigma^2}\Big]$ and $g(x) = \frac{\mu^x e^{-\mu}}{x!}$. Since the Poisson distribution is tail-heavy, we search for solutions μ such that μ at $x = \mu - k\sigma$, for $k \in \mathbb{N}$, we have

$$\frac{f(\mu - k\sigma)}{g(\mu - k\sigma)} = 2 \tag{5}$$

Note that Eq. 5 is just the condition that the Poisson and Gaussian distributions agree to within a factor of 2 at a level of $k\sigma$. (We will look at the cases for $k = 3, 5$.) We have

$$\begin{aligned} 2 &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big[- \frac{(\mu - k\sigma - \mu)^2}{2\sigma^2} \Big] \frac{(\mu - k\sigma)! e^{\mu}}{\mu^{\mu - k\sigma}} \\ 2\sqrt{2\pi\sigma^2} \exp(\tfrac{k^2}{2}) &= \frac{(\mu \ - k\sigma)! e^{\mu}}{\mu^{\mu - k\sigma}} \end{aligned}$$

We approximate the factorial using Stirling's approximation.

$$\begin{aligned} (\mu - k\sigma)! &\approx \sqrt{2\pi(\mu - k\sigma)} (\mu - k\sigma)^{\mu - k\sigma} \exp(-\mu + k\sigma) \\ \implies \log[(\mu - k\sigma)!] &\approx \log[\sqrt{2\pi(\mu - k\sigma)}] + (\mu - k\sigma) \log(\mu - k\sigma) - \mu + k\sigma \\ \implies \log[2\sqrt{s\pi\sigma^2} \exp(\tfrac{k^2}{2})] &\approx \log[\sqrt{2\pi(\mu - k\sigma)}] + (\mu - k\sigma) \log(\mu - k\sigma) - \mu + k\sigma + \mu - (\mu - k\sigma) \log \mu \\ \implies \log[2\sqrt{s\pi\sigma^2} \exp(\tfrac{k^2}{2})] &\approx \log[\sqrt{2\pi(\mu - k\sigma)}], \qquad (\log \mu \approx \log(\mu - k\sigma) \\ \implies 2\sqrt{2\pi\sigma^2} e^{\frac{k^2}{2}} &\approx \sqrt{2\pi(\mu - k\sigma)} \end{aligned}$$

At last we obtain

$$\mu(k, \sigma) \approx 4\sigma^2 \exp(\tfrac{k^2}{2}) + k\sigma \tag{6}$$

As an example, $\mu(3, 1) \approx 363$ and $\mu(5, 1) \approx 1.07 \times 10^8$. So even for small σ, μ must be large for factor of 2 agreement at the 3σ level and much larger still for factor of 2 agreement at the 5σ level.

Since μ is proportional to $\exp(0.5k^2)$ then it is clear that in general the $k = 3$ the value of μ will be much smaller than for $k = 5$. Indeed, if we take the ratio of $\mu(3, \sigma)$ and $\mu(5, \sigma$ we get

$$\frac{\mu(3, \sigma)}{\mu(5, \sigma)} \approx \frac{e^{4.5} + 0.75\sigma^{-1}}{e^{12.5} + 1.25\sigma^{-1}} \tag{7}$$

which as $\sigma \to \infty$ gives $e^{-8} \approx 0.000034$.

3 n Gaussian-distributed RVs

Consider n random variables $(x_i)_{i=1}^{n}$ each Gaussian distributed with identical variance σ^2 and mean μ. If μ is unknown we can calculate the maximum likelihood estimate $\hat{\mu}$ as follows. Let L be the likelihood. We have

$$\log(L) = -\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2} \tag{8}$$

$$\frac{d}{d\mu}\log(L) = \frac{1}{L}\frac{dL}{d\mu} \tag{9}$$

$$\begin{aligned}
\implies \frac{dL}{d\mu} &= L\frac{dL}{d\mu} \\
&= L\frac{d}{d\mu}\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right] \\
&= \frac{L}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)
\end{aligned}$$

$$\implies \frac{dL}{d\mu} = \frac{1}{\sigma^2}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}\right]\sum_{j=1}^{n}(x_j-\mu) \tag{10}$$

In accordance with maximizing the likelihood, we find the value $\hat{\mu}$ such that

$$\left.\frac{dL}{d\mu}\right|_{\mu=\hat{\mu}} = 0 \tag{11}$$

Only the last term on the right in Eq. 10 can be zero. Thus we get

$$\sum_{i=1}^{n}(x_i-\hat{\mu}) = 0 \iff \sum_{i=1}^{n}x_i - n\hat{\mu} = 0 \iff \hat{\mu} = \frac{\sum_{i=1}^{n}x_i}{n} \tag{12}$$

This shows that the estimated mean is just the arithmetic average of the data points. To find the error $\sigma_{\hat{\mu}}$ we start with the variance:

$$\begin{aligned}
\mathrm{Var}[\hat{\mu}] &= \mathrm{Var}\left[\sum_{i=1}^{n}\frac{x_i}{n}\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[x_i] \\
&= \frac{1}{n^2}n\sigma^2 \\
\mathrm{Var}[\hat{\mu}] &= \frac{\sigma^2}{n} \\
\implies \sigma_{\hat{\mu}} &= \frac{\sigma}{\sqrt{n}}
\end{aligned}$$

Now suppose we have $2n$ data points and we erroneously get half of the errors wrong by a factor of $\sqrt{2}$. If the weights are $\mathbf{w} = (w_i)_{i=1}^{2n} = (\frac{1}{\sigma^2})_{i=1}^{2n}$ then we have

$$\begin{aligned}
\mathrm{Var}[\hat{\mu}] &= \frac{\sum_{i=1}^{2n} w_i^2 \mathrm{Var}[x_i]}{(\sum_{i=1}^{2n} w_i)^2} \\
&= \frac{\sum_{i=1}^{n} \frac{1}{\sigma^4}\sigma^2 + \sum_{j=n+1}^{2n} \frac{1}{(\sqrt{2}\sigma)^4}\sigma^2}{(\sum_{i=1}^{n} \frac{1}{\sigma^2} + \sum_{j=n+1}^{2n} \frac{1}{(\sqrt{2}\sigma)^2})^2} \\
&= \frac{\sigma^{-2}}{\sigma^{-4}} \frac{\sum_{i=1}^{n} 1 + \frac{1}{4}\sum_{j=n+1}^{2n} 1}{(\sum_{i=1}^{n} 1 + \frac{1}{2}\sum_{j=n+1}^{2n} 1)^2} \\
&= \sigma^2 \frac{n + \frac{1}{4}}{(n + \frac{1}{2}n)^2} \\
&= \frac{5}{9}\frac{\sigma^2}{n}
\end{aligned}$$

so in this case our false judgment of the errors leads to a lower variance by a factor of 1.8.

Now suppose we have $100n$ data points and we underestimate 1% of the errors by a factor of 100. If $\mathbf{w} = (\frac{1}{\sigma^2})_{i=1}^{100n}$ we get

$$\begin{aligned}
\mathrm{Var}[\hat{\mu}] &= \sigma^2 \frac{\sum_{i=1}^{99n} 1 + 10^8 \sum_{j=99n+1}^{100n} 1}{(\sum_{i=1}^{99n} 1 + 10^4 \sum_{j=99n+1}^{100n} 1)^2} \\
&= \frac{99 + 10^8}{(99 + 10^4)^2}\frac{\sigma^2}{n} \\
&\approx 0.980 \frac{\sigma^2}{n}
\end{aligned}$$

so in this case the misjudged errors lead to a lower variance by only 2%. The case where we overestimate 1% of the errors by a factor of 100 gives

$$\begin{aligned}
\mathrm{Var}[\hat{\mu}] &= \sigma^2 \frac{\sum_{i=1}^{99n} 1 + 10^{-8} \sum_{j=99n+1}^{100n} 1}{(\sum_{i=1}^{99n} 1 + 10^{-4} \sum_{j=99n+1}^{100n} 1)^2} \\
&= \frac{99 + 10^{-8}}{(99 + 10^{-4})^2}\frac{\sigma^2}{n} \\
&\approx 0.01 \frac{\sigma^2}{n}
\end{aligned}$$

In this final case, the variance is underestimated by a factor of about 100. In conclusion, we can say that erring on the side of larger errors has a much smaller effect! It would be safer therefore to overestimate one's uncertainties than to underestimate.

4 Unbiased mean estimate in LLS

Suppose $\mathbf{d} = (d_i)_{i=1}^{n_d}$ represents data points collected and $\mathbf{m} = (m_i)_{i=1}^{n_p}$ is a collection of model parameters. We show that if $\langle \mathbf{d} \rangle = A\mathbf{m}$ for an $n_d \times n_p$ matrix A then the estimate $\hat{\mathbf{m}}$ is unbiased, i.e. $\langle \hat{\mathbf{m}} \rangle = \mathbf{m}$

The maximum-likelihood estimate for

$$\chi^2 = (\mathbf{d} - A\mathbf{m})^{\mathrm{T}} N^{-1} (\mathbf{d} - A\mathbf{m}) \tag{13}$$

is found by solving $\nabla \chi^2 = 0$ and yields

$$\hat{\mathbf{m}} = (A^{\mathrm{T}} N^{-1} A)^{-1} A^{\mathrm{T}} N^{-1} \mathbf{d} \tag{14}$$

Let $\langle \mathbf{d} \rangle = A\mathbf{m}$. We explore the two possibilities: (1) A is non-singular and square; (2) A is singular and possibly rectangular.

1. Let A be non-singular and square. Then A^{-1} is unique. The proof is a straightforward evaluation of Eq. 14.

$$\begin{aligned}
\langle \hat{\mathbf{m}} \rangle &= \langle (A^{\mathrm{T}} N^{-1} A)^{-1} A^{\mathrm{T}} N^{-1} \mathbf{d} \rangle \\
&= \langle A^{-1} N (A^{\mathrm{T}})^{-1} A^{\mathrm{T}} N^{-1} \mathbf{d} \rangle \\
&= \langle A^{-1} N N^{-1} \mathbf{d} \rangle \\
&= \langle A^{-1} \mathbf{d} \rangle \\
&= A^{-1} \langle \mathbf{d} \rangle, \qquad \text{(by linearity)} \\
&= A^{-1} (A\mathbf{m}), \qquad \text{(by hypothesis)} \\
&= \mathbf{m}
\end{aligned}$$

In summary, N could have been an arbitrary $n_d \times n_d$ non-singular matrix since it vanished.

2. Let A be singular and possibly rectangular. By the singular-value decomposition theorem, there exists an orthogonal rectangular matrix U, a diagonal square matrix S and an orthogonal square matrix V such that $A = USV^{\mathrm{T}}$. This gives $A^{\mathrm{T}} = VSU^{\mathrm{T}}$.

We proceed as for case (1):

$$\begin{aligned}
\langle \hat{\mathbf{m}} \rangle &= (A^{\mathrm{T}} N^{-1} A)^{-1} A^{\mathrm{T}} N^{-1} \langle \mathbf{d} \rangle \\
&= (A^{\mathrm{T}} N^{-1} A)^{-1} (A^{\mathrm{T}} N^{-1} A) \mathbf{m} \\
&= [VS(U^{\mathrm{T}} N^{-1} U) SV^{\mathrm{T}}]^{-1} [VS(U^{\mathrm{T}} N^{-1} U) SV^{\mathrm{T}}] \mathbf{m} \\
\langle \hat{\mathbf{m}} \rangle &= [VSQSV^{\mathrm{T}}]^{-1} [VSQSV^{\mathrm{T}}] \mathbf{m}, \qquad Q = U^{\mathrm{T}} N^{-1} U
\end{aligned}$$

But notice $Q\mathrm{T} = (U^{\mathrm{T}}N^{-1}1)^{\mathrm{T}} = U^{\mathrm{T}}N^{-1}U = Q$ so Q is orthogonal and hence the product $VSQSV^{\mathrm{T}}$ is orthogonal (these matrices all have the same dimension and orthogonal matrices form a subgroup of $\mathrm{GL}_n(\mathbb{C})$.) Therefore the last line above evaluates to

$$\langle \hat{\mathbf{m}} \rangle = \mathbf{m} \tag{15}$$

as required. In summary, we again see that N could have been an arbitrary $n_d \times n_d$ matrix.

5 Fitting a Gaussian template

Please see the Python code and discussion on the next page generated from a Jupyter notebook.