PHYS 641
Fall 2022

Assignment 1

Patrick Horlaville
260931235

[1]

Show that a Poisson distribution converges to a Gaussian in the limit of large λ

· Poisson distribution is $$P_p = \frac{e^{-\lambda}\lambda^k}{k!}$$

for λ = expected # of events in an interval
k = actual # of events in an interval

So effectively, we would like to show that for large λ:
$$P_p \to \frac{e^{-(k-\lambda)^2/2\lambda}}{\sqrt{2\pi\lambda}}$$

· For large λ, $k \sim$ continuous, so we can express it as
$k = \lambda(1+\delta)$ with a deviation from the mean $\delta \ll \lambda$

· Stirling's approximation: $n! \approx n^n e^{-n}\sqrt{2\pi n}$

$$\Rightarrow P_p = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-\lambda}\lambda^k}{\sqrt{2\pi k}\, e^{-k} k^k}$$

$$= \frac{e^{-\lambda}\lambda^{\lambda(1+\delta)}}{\sqrt{2\pi\lambda(1+\delta)}\, e^{-\lambda(1+\delta)}(\lambda(1+\delta))^{\lambda(1+\delta)}}$$

$$= \frac{\cancel{e^{-\lambda}}\,\cancel{\lambda^{\lambda(1+\delta)}}}{\sqrt{2\pi\lambda}\,\cancel{e^{-\lambda}}e^{-\lambda\delta}\,\cancel{\lambda^{\lambda(1+\delta)}}\,(1+\delta)^{\lambda(1+\delta)+\frac{1}{2}}}$$

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

• Take log expansion:

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

• For small δ, $\ln(1+\delta) \simeq \delta - \frac{\delta^2}{2} + \underbrace{O(\delta^3)}_{\text{negligible}}$

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

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

$$= \ln\left(\frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}}\right) - \left(\lambda\delta - \frac{\lambda\delta^2}{2} + \lambda\delta^2 \underbrace{- \frac{\lambda\delta^3}{2} + \frac{\delta}{2} - \frac{\delta^2}{2}}_{\text{negligible}}\right)$$

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

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

• Take exponential:

$$k = \lambda(1+\delta)$$
$$\Rightarrow \delta = \frac{k-\lambda}{\lambda}$$

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

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

$$= \frac{\cancel{e^{\lambda\delta}}}{\sqrt{2\pi\lambda}}\, \cancel{e^{-\lambda\delta}}\, e^{-\frac{\lambda\delta^2}{2}}$$

$$= \frac{e^{-\frac{\lambda\delta^2}{2}}}{\sqrt{2\pi\lambda}} \Rightarrow P_p = \frac{e^{-\lambda\left(\frac{k-\lambda}{\lambda}\right)^2/2}}{\sqrt{2\pi\lambda}}$$

mean = var ~ λ

$$\Rightarrow \boxed{P_p = e^{-(k-\lambda)^2/2\lambda} \Big/ \sqrt{2\pi\lambda}}$$

[2]

The gold standard for a believable result is usually 5σ
Let's 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 @ 5σ? 3σ?

- If λ is large enough, then the expected # of events ~ # of observed events
  i.e. $\lambda \sim n$
- We can set $P_P = 2 P_G$ with $k = \mu + \sigma^*$ & solve for λ
  with $\sigma^* = 3\sigma$ & 5σ

for which we would have $\mu \sim \lambda$
$\sigma^2 \sim \lambda \Rightarrow \sigma = \sqrt{\lambda}$

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

$$\Rightarrow \frac{e^{-\lambda}\lambda^{\lambda+\sigma^*}}{(\lambda+\sigma^*)!} = 2\,\frac{e^{-(\lambda+\sigma^*-\lambda)^2/2\lambda}}{\sqrt{2\pi\lambda}}$$

$$\Rightarrow \frac{e^{-\lambda}\lambda^{\lambda+\sigma^*}}{(\lambda+\sigma^*)!} = 2\,\frac{e^{-(\sigma^*)^2/2\lambda}}{\sqrt{2\pi\lambda}}$$

Take first $\sigma^* = 3\sigma = 3\sqrt{\lambda}$

$$\Rightarrow \frac{e^{-\lambda}\lambda^{\lambda+3\sqrt{\lambda}}}{(\lambda+3\sqrt{\lambda})!} = \frac{2\,e^{-(3\sqrt{\lambda})^2/2\lambda}}{\sqrt{2\pi\lambda}}$$

$$\Rightarrow \quad \frac{e^{-\lambda}\lambda^{\lambda+3\sqrt{\lambda}}}{(\lambda+3\sqrt{\lambda})!} = 2\,\frac{e^{-9/2}}{\sqrt{2\pi\lambda}}$$

We can plug e.g. in Wolfram & we find $\lambda = 8{,}217\ldots$

$\lambda \simeq 9 = n$

$\boxed{\text{So for } 3\sigma, \text{ we need } n \simeq 9}$

- Now take $\sigma^* = 5\sigma = 5\sqrt{\lambda}$

$$\Rightarrow \quad \frac{e^{-\lambda}\lambda^{\lambda+5\sqrt{\lambda}}}{(\lambda+5\sqrt{\lambda})!} = 2\,\frac{e^{-25/2}}{\sqrt{2\pi\lambda}}$$

Again with Wolfram, we find $\lambda = \ldots$

Actually, Wolfram cannot solve it. Use Mathematica instead!

we find $\lambda \sim 575\ldots$

$\boxed{\text{So for } 5\sigma, \text{ we need } n = 576}$

$\boxed{3}$

n Gaussian - distributed data points ; identical σ ; same μ
Find error on max likelihood estimate of the mean

- Probability of observing n data points is

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

- Best estimate of μ maximizes P. Minimize exponent :

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

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

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

$$\Rightarrow \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\mu = 0$$

$$\Rightarrow n\mu = \sum_{i=1}^{n} x_i$$

$$\Rightarrow \boxed{\mu = \frac{1}{n}\sum_{i=1}^{n} x_i}$$ Standard expression for the mean

Variance on μ is $\sigma_\mu^2 = \sum_{i=1}^{n}\sigma^2\left(\frac{\partial\mu}{\partial x_i}\right)^2$

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

$$= \sum_{i=1}^{n}\sigma^2\cdot\frac{1}{n^2}$$

$$= n\frac{\sigma^2}{n^2}$$

$$\Rightarrow \boxed{\sigma_\mu^2 = \frac{\sigma^2}{n} \quad \text{or} \quad \sigma_\mu = \frac{\sigma}{\sqrt{n}}}$$

Now we got the errors on $\frac{1}{2}$ the data wrong by a factor of $\sqrt{2}$
What is the error on the new mean estimate?

- We can find our new mean estimate from eq. (2) with independent errors σ_i:

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

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

$$\Rightarrow -\frac{1}{2}(-2)\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma_i^2} = 0$$

$$\Rightarrow \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} - \sum_{i=1}^{n}\frac{\mu}{\sigma_i^2} = 0$$

$$\Rightarrow \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2} = \sum_{i=1}^{n}\frac{\mu}{\sigma_i^2}$$

$$\Rightarrow \boxed{\mu = \frac{\sum_{i=1}^{n} x_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}}$$ (weighted average)

Varince on μ here is $\sigma_\mu^2 = \sum_{i=1}^{n}\sigma_i^2\left(\frac{\partial\mu}{\partial x_i}\right)^2$

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

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

$$= \sum_{i=1}^{n}\sigma_i^2\left(\frac{1/\sigma_i^4}{\left(\sum_{i=1}^{n} 1/\sigma_i^2\right)^2}\right)$$

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

$$\Rightarrow \boxed{\sigma_\mu^2 = \frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2} \quad \text{or} \quad \sigma_\mu = \sqrt{\frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2}}}$$

- Consider the "accurate" variance $\sigma_A^2 = \dfrac{1}{\sum_{i=1}^{n} 1/\sigma_i^2} = \dfrac{1}{n \cdot \frac{1}{\sigma^2}} = \dfrac{\sigma^2}{n}$

$$\Rightarrow \sigma_A = \frac{\sigma}{\sqrt{n}}$$

& the "off" variance $\sigma_B^2 = \dfrac{1}{\sum_{i=1}^{n/2} 1/\sigma_i^2 + \sum_{i=n/2}^{n} 1/2\sigma_i^2}$

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

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

$$\sigma_B^2 = \frac{4\sigma^2}{3n} \Rightarrow \sigma_B = \frac{2}{\sqrt{3}}\frac{\sigma}{\sqrt{n}}$$

- So the error on our estimate would be off by a factor of

$$\frac{\sigma_B}{\sigma_A} = \frac{2}{\sqrt{3}} \sim 1.155$$

- Weights $= w_i \equiv \dfrac{1}{\sigma_i^2}$

- What if 1% of the data is underweighted by a factor of 100?

i.e. $w_{i,A} \to w_{i,C} = \dfrac{w_{i,A}}{100} \Rightarrow \sigma_{i,C}^2 = 100\,\sigma_{i,A}^2$

$$\Rightarrow \sigma_C^2 = \frac{1}{\sum_{i=1}^{n/100} \frac{1}{100\sigma^2} + \sum_{i=n/100}^{n} \frac{1}{\sigma^2}}$$

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

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

$$\sigma_C^2 \approx 1.01\,\frac{\sigma^2}{n} = 1.01\,\sigma_A^2$$

- So the error on our estimate would be off by a factor of $\sqrt{1.01} \sim 1.005$

• What if 1% of the data is overweighted by a factor of 100?

i.e. $w_{i,A} \rightarrow w_{i,O} = 100 w_{i,A} \Rightarrow \sigma_{i,O}^2 = \frac{\sigma_{i,A}^2}{100}$

$$\Rightarrow \sigma_O^2 = \frac{1}{\sum_{i=1}^{n/100} \frac{100}{\sigma^2} + \sum_{i=\frac{n}{100}}^{n} \frac{1}{\sigma^2}}$$

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

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

$$\sigma_O^2 \approx \frac{1}{2}\frac{\sigma^2}{n} = \frac{1}{2}\sigma_A^2$$

• So the error on our estimate would be off by a factor of $\sqrt{1/2} \sim 0.71$

★ I should be more concerned about overweighting that underweighting my data! ★

$\boxed{4}$ c.f. Jupyter Notebook

$\boxed{5}$

Show that in the new rotated space, $\tilde{N}_{ij} = \langle \tilde{n}_i \tilde{n}_j \rangle$

- $\chi^2 = (d - A(m))^T V V^T N^{-1} V V^T (d - A(m))$

$$= (d^T V - A(m)^T V)(V^T N^{-1} V)(V^T d - V^T A(m))$$

- The noise matrix in our rotated space is $\tilde{N}^{-1} = V^T N^{-1} V$ so that

$$\tilde{N} = (V^T N^{-1} V)^{-1}$$

$$\tilde{N} = V^T N V$$ as $V^T V = I$ since V orthogonal

$$\therefore \tilde{N}_{ij} = V_{ij}^T N_{ij} V_{ij} \qquad (3)$$

- Our noise $n_i = d_i - d_{t_i}$ becomes $\tilde{n}_i = \tilde{d}_i - \tilde{d}_{ti}$

for which the observed & true data are

$$\tilde{d}_i = V_{ij}^T d_j$$

$$\tilde{d}_{t_i} = V_{ij}^T d_{t_j}$$

- The noise becomes

$$\tilde{n}_i = V_{ij}^T (d_j - d_{t_j})$$

$$\Rightarrow \langle \tilde{n}_i \tilde{n}_j \rangle = \langle V_{ij}^T (d_j - d_{t_j}) (V_{ji}^T (d_i - d_{t_i}))^T \rangle$$

$$= \langle V_{ij}^T (d_j - d_{t_j})(d_i - d_{t_i})^T V_{ji} \rangle$$

$$= \langle V_{ij}^T \rangle \langle (d_j - d_{t_j})(d_i - d_{t_i})^T \rangle \langle V_{ji} \rangle$$

$$= V_{ij}^T \langle n_i n_j \rangle V_{ij}$$

$$= V_{ij}^T N_{ij} V_{ij}$$

$$= \tilde{N}_{ij} \text{ from eq (3)}$$

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