Solutions 9: Maximum likelihood

BIO-369

Prof. Anne-Florence Bitbol
EPFL

1 Estimating a diffusion coefficient

a) Because the N measurements of displacements Δx, Δy of the particle are considered independent, the joint probability distribution function of the dataset $(\Delta x_1, \Delta y_1, \Delta x_2, \Delta y_2, \ldots, \Delta x_N, \Delta y_N)$ at a given V is the product of the probability distributions of each $(\Delta x, \Delta y)$:

$$p(\Delta x_1, \Delta y_1, \Delta x_2, \Delta y_2, \ldots, \Delta x_N, \Delta y_N|V) = \frac{1}{(2\pi V)^N} \exp\left[-\frac{1}{2V}\sum_{i=1}^{N}\left(\Delta x_i^2 + \Delta y_i^2\right)\right], \qquad (1)$$

It is the probability of the data given the model (Gaussian distribution with variance V), and thus it can be interpreted as a likelihood function. Its logarithm reads:

$$\log\left[p(\Delta x_1, \Delta y_1, \Delta x_2, \Delta y_2, \ldots, \Delta x_N, \Delta y_N|V)\right] = -N\log(2\pi V) - \frac{1}{2V}\sum_{i=1}^{N}\left(\Delta x_i^2 + \Delta y_i^2\right), \qquad (2)$$

It is easier to work with it because it does not involve an exponential.

b) To obtain the maximum-likelihood estimate of V, the log-likelihood can be differentiated with respect to V (the data being fixed):

$$\frac{\partial \log\left[p(\Delta x_1, \Delta y_1, \Delta x_2, \Delta y_2, \ldots, \Delta x_N, \Delta y_N|V)\right]}{\partial V} = -\frac{N}{V} + \frac{1}{2V^2}\sum_{i=1}^{N}\left(\Delta x_i^2 + \Delta y_i^2\right), \qquad (3)$$

and then, setting it to zero yields

$$V = \frac{\sum_{i=1}^{N}\left(\Delta x_i^2 + \Delta y_i^2\right)}{2N}. \qquad (4)$$

c) See Jupyter notebook. The computation with the real data yields $\sigma = \sqrt{V} = 5.75$ μm.

d) Here $D = V/(2T) = 0.55$ $\mu\text{m}^2/\text{s}$.

2 Counting fluorescent molecules

a) For the Poisson distribution, the variance is equal to the mean λ. See Jupyter notebook. Here the variance (16) is not far from the mean (22), they are not the same but there is not much data so the discrepancy is not sufficient to exclude a Poisson distribution.

b) Assuming that measurements are independent, the likelihood of the data $(n_1, n_2, \ldots, n_K)$ given the model (Poisson distribution) with mean λ reads

$$P(n_1, \ldots, n_K|\lambda) = e^{-K\lambda}\,\frac{\lambda^{\sum_i n_i}}{\prod_i n_i!}, \qquad (5)$$

where i ranges from 1 to K. The log-likelihood reads:

$$\log\left[P(n_1,\ldots,n_K|\lambda)\right] = -K\lambda + \left(\sum_i n_i\right)\log\lambda - \log\left(\prod_i n_i!\right)\,, \tag{6}$$

and we will focus on the reduced log-likelihood

$$\mathcal{L}(n_1,\ldots,n_K|\lambda) = -K\lambda + \left(\sum_i n_i\right)\log\lambda\,, \tag{7}$$

because the last term in Eq. 6 does not depend on λ and is heavy to compute. Differentiating the log-likelihood (or the reduced log-likelihood) with respect to λ yields

$$\frac{\partial \log\left[P(n_1,\ldots,n_K|\lambda)\right]}{\partial\lambda} = -K + \frac{\sum_i n_i}{\lambda}\,, \tag{8}$$

and setting this to zero gives the maximum-likelihood estimate of λ as

$$\lambda = \frac{\sum_i n_i}{K}\,, \tag{9}$$

which is the regular expectation for the mean of the data.

c) The maximum value taken by the reduced log-likelihood is obtained from Eq. 7 for λ in Eq. 9, yielding

$$\max\mathcal{L}(n_1,\ldots,n_K|\lambda) = \left(\sum_i n_i\right)\left[\log\left(\frac{\sum_i n_i}{K}\right) - 1\right]\,. \tag{10}$$

See Jupyter notebook. Given the plot, the range of credible values of λ is approximately 20 to 24, since the likelihood is much smaller than its maximum value outside of this range.

3 Luria-Delbrück experiment, revisited

a) Here the data is a set of K numbers m of observed mutants, one per replicate culture. Since replicates are independent, the likelihood of the data given a model reads

$$P(m_1,\ldots,m_K|\text{model}) = \prod_{i=1}^{K} P(m_i|\text{model})\,, \tag{11}$$

and its logarithm reads

$$\begin{aligned}\log\left[P(m_1,\ldots,m_K|\text{model})\right] &= \sum_{i=1}^{K}\log\left[P(m_i|\text{model})\right] = \sum_m \sum_{i\text{ such that }m_i=m}\log\left[P(m|\text{model})\right]\\ &= \sum_m n(m)\log\left[P(m|\text{model})\right]\,,\end{aligned} \tag{12}$$

where we have separated experimental replicates i according to the number m of resistant cells that were observed in m and then introduced $n(m)$, the number of replicates where m resistant cells are observed. Thus, if we now consider two different models, the logarithm of the ratio of likelihoods for these two models can be expressed as

$$\log\left[\frac{P(\text{data}|\text{model})}{P(\text{data}|\text{model}')}\right] = \sum_m n(m)\left[\log P(m|\text{model}) - \log P(m|\text{model}')\right]\,, \tag{13}$$

where $n(m)$ is the number of replicates of the experiment where m resistant bacteria were observed.

b) For each value of m, and using the logarithm with base 10, $\left[\log P(m|\text{model}) - \log P(m|\text{model}')\right]$ can be read directly from the plot by looking at the height between the gray and the red dots, for instance when m is between 11 and 20, the gray dot is around 10^{-10} and the red one around 10^{-2}, so the logarithms in base 10 of the probabilities according to each models are -10 and -2, and their difference is 8. There are 6 replicates that fall in this category, so already the contribution of this point to the log-likelihood ratio is 48. This is a lower bound for the logarithm of the ratio of likelihoods for the two models, we can make it larger by considering other data points, but this is sufficient to see that the likelihood ratio is huge (more than 10^{48}).

c) Even if we initially thought the Lamarckian hypothesis was five times more probable than the Darwinian, the likelihood ratio completely overwhelms the prior ratio, and we conclude in favor of the Darwinian hypothesis after the experiment.

4 Additional problem: Luria-Delbrück experiment again

a) Each bacteria exposed to phage has a probability μ of becoming resistant and $1-\mu$ of remaining sensitive to phage under the Lamarckian hypothesis. The probability distribution of the random variable X characterizing whether one bacteria is resistant or not is a Bernoulli distribution with parameter μ.

b) In a culture containing N bacteria that is exposed to phage, the probability that m of them become resistant under the Lamarckian hypothesis is given by the binomial distribution:

$$P(m) = \binom{N}{m} \mu^m (1-\mu)^{N-m} \,. \tag{14}$$

c) If $\mu \ll 1$ and $N \gg 1$ but $N\mu = \lambda$ is of order one, this distribution becomes the Poisson distribution with parameter $\lambda = N\mu$,

$$P(m) = e^{-\lambda} \frac{\lambda^m}{m!} \,. \tag{15}$$

The mean value of m reads:

$$\langle m \rangle = \sum_{m=0}^{\infty} m \frac{\lambda^m e^{-\lambda}}{m!} = \sum_{m=1}^{\infty} m \frac{\lambda^m e^{-\lambda}}{m!} = \sum_{m=1}^{\infty} \frac{\lambda^m e^{-\lambda}}{(m-1)!} = \lambda e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^{m-1}}{(m-1)!} \,. \tag{16}$$

Introducing $k = m-1$, we obtain

$$\langle m \rangle = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda \,, \tag{17}$$

where we have used the series expansion of the exponential function

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \,. \tag{18}$$

In addition, for a Poisson distribution, the variance is equal to the mean, so it is also equal to λ.

d) Under the Darwinian hypothesis where phage resistance mutations can be acquired randomly at any time during growth of the bacterial cultures (before exposure to phage), if a mutation occurs early in the growth process, then the culture will comprise many mutants as all descendants of the mutant are mutant too. Thus, the number of colonies with large numbers of mutants (albeit small) is much larger under the Darwinian hypothesis than under the Lamarckian one. These events are called "jackpots".

e) If $\mu_D = 0$, there is no Darwinian mutation, so we recover the Lamarckian model. If $\mu_L = 0$, there is no Lamarckian mutation, so we recover the Darwinian model. Since both the Lamarckian and the Darwinian model are included in the Combined model for specific values of the parameters, the Combined model cannot fit the data less well than the pure Darwinian and the pure Lamarckian model.

f) The link is given by Bayes theorem:

$$P(\text{model}|\text{data}) = P(\text{data}|\text{model}) \times \frac{P(\text{model})}{P(\text{data})}\,, \tag{19}$$

where $P(\text{model}|\text{data})$ is the posterior, $P(\text{data}|\text{model})$ is the likelihood and $P(\text{model})$ is the prior.

g) To compare two models with a fixed dataset, one should in principle compare the posterior of each model given that data. We have

$$\frac{P(\text{model 1}|\text{data})}{P(\text{model 2}|\text{data})} = \frac{P(\text{data}|\text{model 1})}{P(\text{data}|\text{model 2})} \times \frac{P(\text{model 1})}{P(\text{model 2})}\,, \tag{20}$$

Since the data is the same in both cases, the term $P(\text{data})$ has vanished here. If the two models have the same prior, then the ratio of the posteriors is equal to the ratio of the likelihoods, and this reduces to a maximum likelihood analysis. One should choose the model that has the largest likelihood among the 2 possible models. More generally, under the uniform prior, finding the model that maximizes the posterior is equivalent to finding the model that maximizes the likelihood.

h) The drawback of the Combined model is that it has 2 free parameters that will have to be fitted to the data, μ_D and μ_L. It is thus a more complex models than the other two, that aims to explain the same data. If we want to penalize the Combined model for its complexity, we could possibly decrease its prior.

i) In a maximum likelihood analysis, having more replicates can make the likelihoods more distinct and allows to better distinguish between models. Indeed, the product of the likelihood ratios of each replicate will yield the overall likelihood ratio. If for instance there are multiple small terms there, then the overall likelihood ratio will be smaller.

[Detailed analysis below not required.] Technically, if the data is a set of K numbers m of observed mutants, one per replicate culture. Since replicates are independent, the likelihood of the data given a model reads

$$P(m_1, \ldots, m_K|\text{model}) = \prod_{i=1}^{K} P(m_i|\text{model})\,, \tag{21}$$

and if comparing two models, we should focus on

$$\frac{P(m_1, \ldots, m_K|\text{model 1})}{P(m_1, \ldots, m_K|\text{model 2})} = \prod_{i=1}^{K} \frac{P(m_i|\text{model 1})}{P(m_i|\text{model 2})}\,, \tag{22}$$

which for instance becomes very small if many of the terms of the product are small.

j) The figure shows that the fit represents well the data, with only some fluctuations of one with respect to the other but no visible systematic difference. The jackpot events with large m in particular are well captured by the model (as shown in the inset). This seems to indicate that there is some Lamarckian component in the process, since the best fit value for μ_L is not zero. However, we see that the adjusted values satisfy $\mu_L < \mu_D$, meaning that the Darwinian mechanism appears dominant from the analysis of this experiment.

k) Here the fit is far less good, and in particular the jackpot events do not appear to be well-captured by the fit, since we see that the experimental data features higher frequencies for large values of m than the fit. We see that the fit gave $\mu_L = 0$ here (no Lamarckian component),and we know that jackpots arise from Darwinian mutations. Here, even the purely Darwinian hypothesis does not yield large enough jackpots to explain the data. There could have been an additional phenomenon at play in this particular experiment, for instance a contamination by a resistant strain.

l) From these two results, we see that in one case the best fit has a nonzero μ_L, but with $\mu_L < \mu_D$. In the other case the best fit is purely Darwinian, but it explains the data less well then in the first case. Overall, based on these results alone, we cannot exclude the Combined or the Darwinian models, but it seems reasonable to exclude the Lamarckian one because in both cases, $\mu_L < \mu_D$ and μ_D is nonzero. To gain more insight, it would be good to perform additional experiments with more replicates.