Bayesian parameter estimation

September, 2023

This session is designed to familarize you with Bayesian probabilistic reasoning, and to give you some experience in applying Bayes' theorem.

As a reminder, Bayes' theorem is:

$$P(\theta|D) = \frac{P(\theta) \times P(D|\theta)}{P(D)} \tag{1}$$

where θ represents the hypothesis that you are testing, and D represents the data. The left-hand side, $P(\theta|D)$ is the posterior probability, and on the right-hand side $P(\theta)$ is the prior and $P(D|\theta)$ is the likelihood. The denominator on the right-hand side is a normalisation constant. If θ represents a choice of several possibilities θ_i, then we have $P(D) = \sum_i P(\theta_i)P(D|\theta_i)$.[1]

[1] This expression is for discrete probabilities; everything here is identical for the probability density of continuous variables except that sums become integrals.

Hint: you may assume Gaussian distributions, but think about why that is sensible.

You have a measurement device with a known root-mean-square error of σ. It measures the value of some quantity θ. Discuss some examples in your groups, with reasonable numerical values for the the errors which you should compare to typical values of the measurement θ.

1. If you make a measurement and the display readout shows μ, what likelihood function might you write down for this case? (You may need to make some assumptions — discuss and justify them in your groups.)

2. If you have very little prior information about the possible value of θ, we might consider a uniform prior, $P(\theta) = \text{const.}$[2] Discuss how this represents some sort of "ignorance". What is the resulting posterior distribution? How would you report this as an estimate and error (e.g., $\hat{\theta} \pm \hat{\sigma}$)?

3. Now imagine that you are informed of the results of a previous experiment which has provided its results in the form of a posterior distribution (for *that experiment*) showing that the value of θ is probably ϕ, with a root-mean-square spread of Δ. What is $P(\theta)$ if we encode this as a Gaussian distribution?

4. Now consider these two experiments together. How would you use Bayes' theorem to combine their results? What is the prior? What is the likelihood?

[2] This is technically an *improper* prior, since it cannot integrate to one for any value of the constant. You can think of it as being constant over some range that is much, much wider than any of the values of θ where the likelihood is non-negligible.

For the case in which both can be described by Gaussian distributions, the resulting posterior will also be a Gaussian in θ:

$$P(\theta|\mu) \propto \exp\left[-\frac{1}{2}(\theta-\bar{\theta})^2/\Sigma^2\right] , \tag{2}$$

with mean and variance given by

$$\begin{aligned}\bar{\theta} &= \frac{\mu/\sigma^2 + \phi/\Delta^2}{1/\Delta^2 + 1/\sigma^2} \\ \Sigma^2 &= \frac{\Delta^2\sigma^2}{\Delta^2+\sigma^2} = \frac{1}{1/\Delta^2 + 1/\sigma^2} \, .\end{aligned} \tag{3}$$

Note that the mean is just a weighted average of the prior and likelihood means, with the weights given by the *inverses* of the variances, and that the *inverse* of the posterior variance is the sum of the *inverses* of the likelihood and prior variances.

5. Derive these expressions for $\bar{\theta}$ and Σ.[3]

[3] One way to calculate $\bar{\theta}$ and variance Σ^2 is to "complete the square".

6. The previous experiment has given a value of $\phi = 10$ with error $\Delta = 1$. Consider several different possible results for your experiment:

μ	σ
9.0	1.0
2.0	1.0
11.0	0.05
20.0	0.05
8.0	10.0
50.0	10.0

   For each of these, separately graph the prior, likelihood, and posterior on the same axes. How does the posterior respond to the prior in each case?[4]

   What happens in the limit that $\Delta \ll \sigma$? What happens in the limit $\sigma \ll \Delta$?[5]

[4] You will make six graphs, each with three curves. This might be a good exercise for your programming/plotting skills, or you can use this colab jupyter notebook.

[5] You can probably answer these with what you've learned from the graphs, but also show it mathematically.

7. Further questions for discussion: What if you were measuring a parameter that you knew to be strictly positive? Or knew it to fall between known upper and lower bounds? What if your prior information was better expressed in terms of a function of θ (e.g., $\ln\theta$ or θ^2)? Perhaps one or more of these was already true for your examples. In which cases might these details *not* matter in terms of your numerical answers?