In-class exercises

Instructions

- **Don't look at the solutions yet!** This is for your benefit.
- These exercises must be submitted within 48 hours of the lecture in which it was given.
- As long as you do the exercises on time, you get full credit—your performance does not matter.
- For each exercise, without looking at the solution, take the amount of time shown to work on the exercise.
- Note: These exercises are more involved than usual, so you're not expected to get a complete solution in the time allotted.
- Submit your work on the course website.

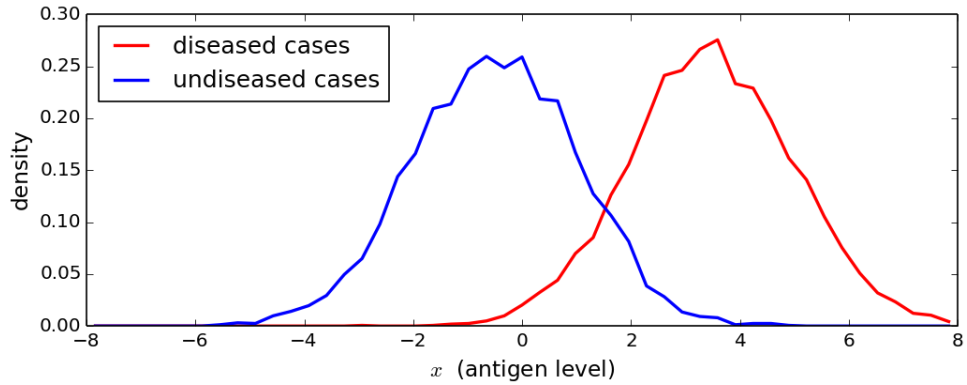


Figure 1: Empirical distribution of x for diseased and undiseased cases.

Exercise 1: Disease diagnosis

The enzyme-linked immunosorbent assay (ELISA) is a chemical test for the presence of an antigen of interest. It is very widely used for the detection of diseases such as malaria, HIV, West Nile virus, and celiac disease, and can also be used to detect food allergens such as nuts, milk, and eggs.

Suppose you are using an ELISA test to detect a disease. The test produces a single number, say x, measuring the amount of antigen. From many previous tests, you essentially know the distribution of x in each case, that is, you know $p(x|D)$ in the diseased cases (D) and $p(x|U)$ in the undiseased cases (U). See Figure 1.

You need to diagnose each future case as diseased or undiseased, based on the observed x. How would you make this diagnosis?

(Think for 1 minute about roughly how you might approach this.)

The normal (a.k.a. Gaussian) distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2 has p.d.f.

$$\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2 \Big).$$

Assume that

$$p(x|D) = \mathcal{N}(x \mid \mu_D, \sigma^2)$$
$$p(x|U) = \mathcal{N}(x \mid \mu_U, \sigma^2)$$

where μ_D, μ_U, and σ are all known.

(Work on this for 10 minutes.)

Solution

This is a clear case of a decision problem with two possible actions: diagnosed as diseased (D) or undiseased (U). A Bayesian approach is as follows. Determine the prior probability of disease $p(D)$ (and in turn, $p(U) = 1 - p(D)$); this might be possible to determine based on the previous cases, but beware of biased sampling. Determine the loss function ℓ:

		Diagnosis D	Diagnosis U
Truth	D	$\ell(D, D)$	$\ell(D, U)$
	U	$\ell(U, D)$	$\ell(U, U)$

Using Bayes' theorem, compute the posterior probability of disease for a given individual,

$$p(D|x) = \frac{p(x|D)p(D)}{p(x|D)p(D) + p(x|U)p(U)}.$$

Make the diagnosis ($a = D$ or $a = U$) that minimizes the posterior expected loss,

$$\rho(a, x) = \ell(D, a)p(D|x) + \ell(U, a)p(U|x).$$

It turns out that this decision rule takes a nice and simple form in which we just have to compare x to a "cut-off" value. Assume $\mu_D > \mu_U$ (the mean antigen level for diseased is higher than undiseased), $\ell(D, U) > \ell(D, D)$ (the loss for misdiagnosing diseased cases is higher than correctly diagnosing them), and $\ell(U, D) > \ell(U, U)$ (the loss for misdiagnosing undiseased cases is higher than correctly diagnosing them). It can be shown (with a couple pages of calculations) that the Bayes procedure is to diagnose as "diseased" whenever

$$x > c$$

and "undiseased" otherwise, where

$$c = \frac{\sigma^2}{\mu_D - \mu_U}\left(\frac{\mu_D^2 - \mu_U^2}{2\sigma^2} + \log\frac{p(U)}{p(D)} + \log\frac{\ell(U, D) - \ell(U, U)}{\ell(D, U) - \ell(D, D)}\right).$$

Note: If the variances of $p(x|D)$ and $p(x|U)$ are not equal, then the Bayes procedure is a little more complicated.

Exercise 2: The speed of light

In 1882, Simon Newcomb made 66 measurements of the amount of time it took for light to travel from his laboratory on the Potomac River, to a mirror placed at the base of the Washington Monument, and back. Converted to speed, in units of $\times 10^8$ meters/second, his data was as follows:

2.9974	2.9976	2.9968	2.9979	2.9966	3.0061
2.9975	2.9988	2.9959	3.0010	2.9973	2.9981
2.9979	2.9982	2.9977	2.9971	2.9980	2.9973
2.9970	2.9985	2.9979	2.9983	2.9964	2.9969
2.9964	2.9974	2.9977	2.9982	2.9974	2.9973
2.9963	2.9977	2.9974	2.9976	2.9971	2.9969
2.9964	2.9976	2.9971	2.9981	2.9964	2.9980
2.9975	2.9975	2.9974	2.9975	2.9970	2.9975
2.9976	2.9968	2.9976	2.9969	2.9969	2.9979
2.9960	2.9974	2.9979	2.9977	2.9969	2.9977
2.9973	2.9975	2.9974	2.9973	2.9988	2.9980

If, like Newcomb, you were interested in determining the speed of light, how would you analyze this data?

(Think for 1 minute about roughly how you might approach this.)

The normal (a.k.a. Gaussian) distribution $\mathcal{N}(\theta, \lambda^{-1})$ with mean θ and precision (inverse variance) $\lambda = 1/\sigma^2$ has p.d.f.

$$\sqrt{\frac{\lambda}{2\pi}} \exp\Big(-\frac{\lambda}{2}(x-\theta)^2\Big).$$

Let θ be the speed of light, and consider a normal model for Newcomb's data:

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \lambda^{-1}).$$

For simplicity, assume the precision λ is known from extensive calibration testing. Consider the "improper" uniform prior: for all θ,

$$p(\theta) = 1.$$

Using Bayes' theorem in the usual way, what is the resulting posterior distribution $p(\theta|x_{1:n})$? Give your answer in terms of $x_1, \ldots, x_n$ (in other words, you don't need to plug in the actual values above).

(Work on this for 10 minutes.)

Solution

One way of analyzing this data is to use a normal model, as described above. To derive the posterior, first note that

$$\begin{aligned} p(x_i|\theta) &= \sqrt{\frac{\lambda}{2\pi}} \exp\Big(-\frac{\lambda}{2}(x_i-\theta)^2\Big) \\ &\propto \exp\Big(-\frac{\lambda}{2}(-2x_i\theta+\theta^2)\Big) \\ &\propto \exp\big(\lambda x_i\theta - \tfrac{1}{2}\lambda\theta^2\big). \end{aligned}$$

Since we are using the flat prior $p(\theta) = 1$, the posterior is

$$\begin{aligned} p(\theta|x_{1:n}) &\propto p(x_{1:n}|\theta)p(\theta) \\ &= \prod_{i=1}^{n} p(x_i|\theta) \\ &\propto \exp\big(\lambda\theta\textstyle\sum x_i - \frac{1}{2}n\lambda\theta^2\big). \end{aligned} \tag{0.1}$$

Since the exponent is quadratic in θ, this is proportional to a normal distribution. To find the parameters, consider a generic normal distribution on θ:

$$\begin{aligned} \mathcal{N}(\theta|\mu, L^{-1}) &= \sqrt{\frac{L}{2\pi}} \exp\Big(-\frac{L}{2}(\theta-\mu)^2\Big) \\ &\propto \exp\big(-\tfrac{1}{2}L\theta^2 + L\theta\mu\big). \end{aligned}$$

This will be equal to Equation 0.1 if

$$\begin{aligned} L &= n\lambda \qquad \text{and} \\ L\mu &= \lambda\textstyle\sum x_i. \end{aligned}$$

Solving for μ and L gives us

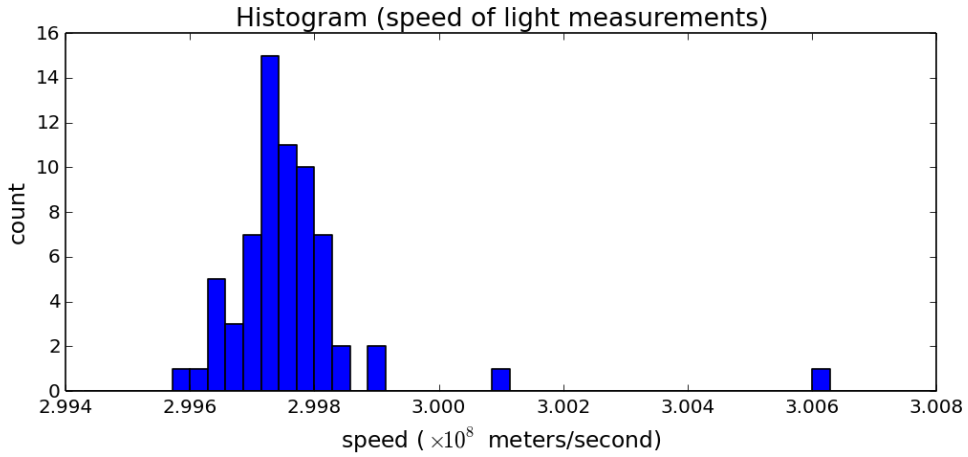
$$\begin{aligned} \mu &= \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \\ L &= n\lambda. \end{aligned}$$

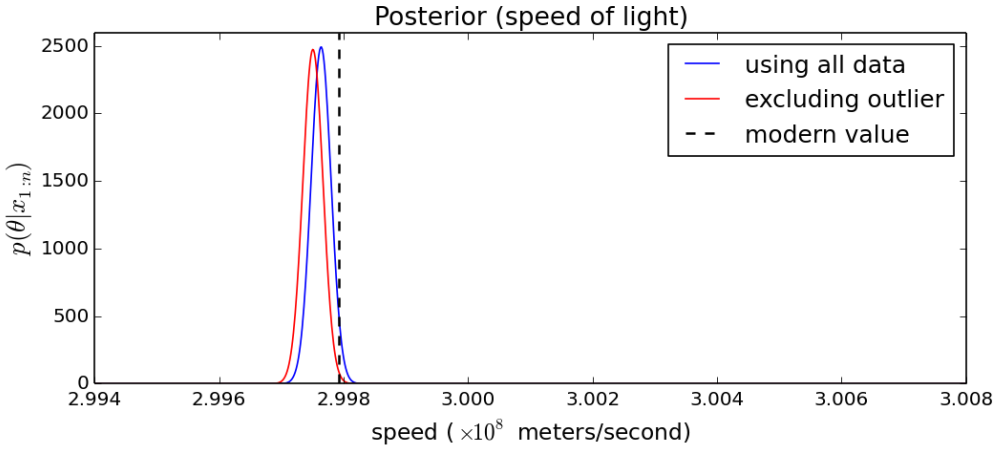
Therefore,

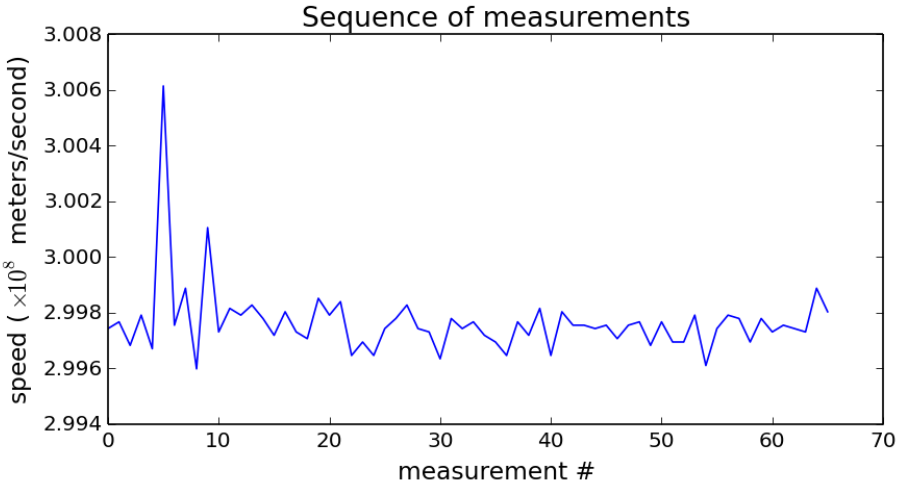
$$p(\theta|x_{1:n}) = \mathcal{N}(\theta \mid \bar{x}, (n\lambda)^{-1}).$$

Outliers

This dataset has an outlier (or maybe two outliers) relatively far away from the majority of the observations. This is difficult to see from the data alone, but apparent from the histogram. It's always important to visualize your data—it is said that the statistician's best tool is the eye.


Histogram (speed of light measurements)
count
speed ($\times 10^8$ meters/second)


Posterior (speed of light)
using all data
excluding outlier
modern value
$p(\theta|x_{1:n})$
speed ($\times 10^8$ meters/second)


Sequence of measurements
speed ($\times 10^8$ meters/second)
measurement #

If the extreme outlier (3.0061) is removed, then the posterior is noticeably shifted to the left. This illustrates that a normal model is sensitive to outliers. In fact, even one outlier—if sufficiently far away—can have an arbitrarily large effect on the posterior. To deal with this, sometimes people use a generating distribution with "heavier tails", such as Student's t-distribution or the Laplace (a.k.a. double exponential) distribution, instead of the normal distribution. However, these are not as easy to use, computationally.

Improper priors

An ***improper prior*** is a nonnegative function $p(\theta)$ which is not a p.d.f. because it does not have a finite integral, in other words, $\int p(\theta)d\theta = \infty$. The choice of $p(\theta) = 1$, for $\theta \in \mathbb{R}$, is an example of an improper prior.

Even though an improper prior is not actually a prior at all (since it does not correspond to a probability distribution), we can still try to plug it into Bayes' theorem as though it were—in other words, we can define the posterior to be proportional to $p(x|\theta)p(\theta)$. It turns out that in many cases (but not always) this is normalizable, and thus, results in a "proper" posterior.

References and supplements

- Youden, W. J. (1972). Enduring values. Technometrics, 14(1), 1-11.