S2 Statistics for Data Science

CRSiD: tmb76

University of Cambridge

Contents

The Lighthouse Problem

The Setup

A lighthouse that is at a distance β from the coast is at position α along it. The lighthouse emits flashes at random angles θ, following a uniform distribution. The flashes can be considered narrow and, provided $\pi/2 < \theta < \pi/2$, intersect the coastline at a single point. Detectors on the coastline record only the flashes' locations x_k (where $k = 1, 2, \ldots, N$) for N flashes received. The setup is illustrated in Figure 1.

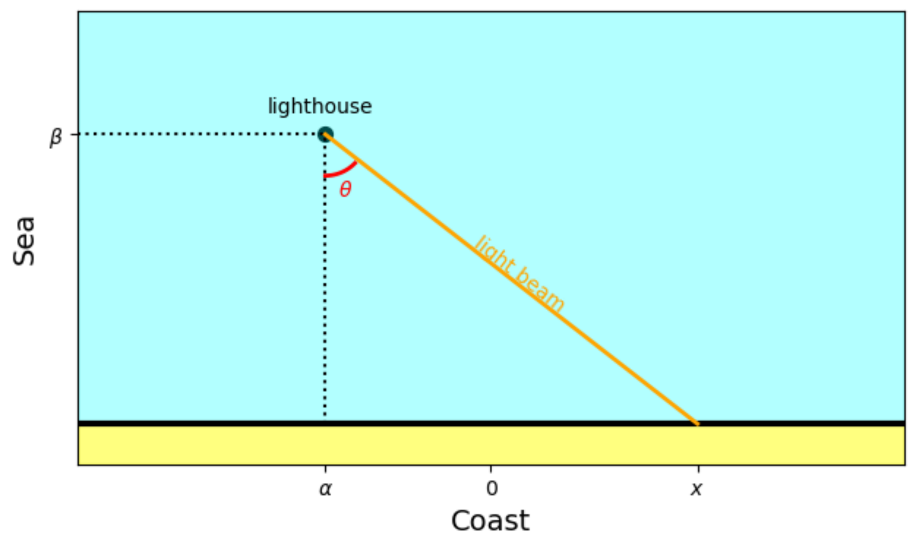

Figure 1: The lighthouse problem setup

The goal of this project is to find the location of the lighthouse from the recorded data x_k.

(i) The Geometry of the Problem

As stated above, only the distribution of the flashes' angle, θ, is known. However, the observed data obtained is the location of the flashes on the coastline, x_k. Thus, it is worth considering what the relationship is between the angle of the flash and the location of the flash on the coastline, expressed in terms of the unknown lightouse location parameters α and β. Using trigonometry, the following relationship can be derived:

$$x = \alpha + \beta \tan(\theta) \iff \theta = \tan^{-1}\left(\frac{x-\alpha}{\beta}\right) \tag{1}$$

(ii) The Likelihood Function

In addition to the relationship above, the probability distribution of the flashes' angles is known to be uniform. Thus, the probability of a single flash to have the angle θ is given by:

$$\theta \sim U(-\pi/2, \pi/2) \tag{2}$$

$$P(\theta) = \mathbb{1}_{(-\pi/2,\pi/2)}(\theta) \times \frac{1}{\pi} = \begin{cases} \frac{1}{\pi}, & \text{if } -\frac{\pi}{2} < \theta < \frac{\pi}{2} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Now, using the trigonometric relationship between θ and x derived in (1), a change of variable is conducted to find the probability distribution of the flashes' locations on the coastline, x. Here, this will be considered as the likelihood of a single flash to be observed at location x given the lighthouse location parameters α and β. For this 1-D case, the change of variable can be written as:

$$\mathcal{L}_x(x|\alpha,\beta) = \mathcal{L}_\theta(\theta) \times \left|\frac{d\theta}{dx}\right| \tag{4}$$

$\mathcal{L}_\theta(\theta)$ is the probability distribution of the flashes' angles, which is given in Equation (3). Since the angle θ must be such that $-\pi/2 < \theta < \pi/2$ for the flash to be observed on the coastline, $\mathcal{L}_\theta(\theta)$ can be set to $1/\pi$. The dereivative term is given by:

$$\frac{d\theta}{dx} = \frac{d}{dx}(\tan^{-1}\left(\frac{x-\alpha}{\beta}\right)) \tag{5}$$

Using the chain rule, the derivative can be found to be:

$$\frac{d\theta}{dx} = \frac{1}{1+\left(\frac{x-\alpha}{\beta}\right)^2} \times \frac{1}{\beta} \tag{6}$$

With some algebraic manipulation, this can be re-arranged as:

$$\frac{d\theta}{dx} = \frac{\beta}{\beta^2 + (x-\alpha)^2} \tag{7}$$

Thus the likelihood for a single flash to be observed at location x given the lighthouse location parameters α and β is:

$$\mathcal{L}_x(x|\alpha,\beta) = \frac{1}{\pi} \times \frac{\beta}{\beta^2 + (x-\alpha)^2} \tag{8}$$

as required.

(iii) Frequentist Claim

A frequentist colleague claims the most likely location, $\hat{x}$, for any flash to be received is given by the parameter α, the location of the lighthouse along the coastline. They also suggest using the sample mean to estimate the α parameter. This turns out to be a bad estimator. First, the $hatx = \alpha$ claim. hat_x is the location for which the likelihood in Eq. (8) is maximised. This is found by taking the derivative of the likelihood with respect to x and setting it to zero. This gives:

$$\left.\frac{d\mathcal{L}_x(x|\alpha,\beta)}{dx}\right|_{x=\hat{x}} = 0 \iff \left.\frac{d}{dx}\left(\frac{\beta}{\beta^2+(x-\alpha)^2}\right)\right|_{x=\hat{x}} = 0 \tag{9}$$

$$-\frac{1}{\pi}\beta(2\hat{x}-2\alpha)\frac{1}{\beta^4+2\beta^2(\hat{x}-\alpha)^2+(\hat{x}-\alpha)^4} = \tag{10}$$

This is satisfied when $\hat{x} = \alpha$, as claimed. Thus, the frequentist's claim is accurate for the value of the most likely flash location. However, estimating the value of α with the sample mean is not good. To show this, we compare it with the Maximum Likelihood Estimate (MLE) method. The aim is to find the value of α that maximises the likelihood derived in Eq. (8), based on the observed sample x_k $(k = 1, 2, \ldots, N)$. This means now having the likelihood being the product of individula likelihoods for all values of x_k. This is done by taking the derivative of the likelihood with respect to α and setting it to zero. However, for simplicity in derivation and computation, the log-likelihood is used instead.

$$\hat{\alpha}_{MLE} = \arg\max_{\alpha} \ln(\prod_{k=1}^{N} \mathcal{L}_x(x_k|\alpha,\beta)) = \arg\max_{\alpha} \sum_{k=1}^{N} \ln(\frac{1}{\pi} \times \frac{\beta}{\beta^2+(x_k-\alpha)^2}) \tag{11}$$

$$\hat{\alpha}_{MLE} = \arg\max_{\alpha} \sum_{k=1}^{N} \ln(\frac{\beta}{\pi}) + \ln(\frac{1}{(\beta^2+(x_k-\alpha)^2)}) \tag{12}$$

$$\left.\frac{d}{d\alpha}\left(\sum_{k=1}^{N} \ln(\frac{\beta}{\pi}) + \ln(\frac{1}{(\beta^2+(x_k-\alpha)^2)})\right)\right|_{\alpha=\hat{\alpha}_{MLE}} = 0 \tag{13}$$

$$\sum_{k=1}^{N} \frac{(2\alpha - 2x_k)}{(\beta^2 + (x_k - \alpha)^2)}\bigg|_{\alpha=\hat{\alpha}_{MLE}} = 0 \tag{14}$$

To get an estimate of $\hat{\alpha}_{MLE}$, the above equation needs to be rearranged to isolate α. However, this is difficult to do analytically. More importantly, the overarching issue is that the Cauchy distribution does not follow the Central Limit Theorem. This is shown in the fact that the distribution of the sample mean of the Cauchy distribution is itself a Cauchy distribution, no matter the sample size:

(iv) Priors for α and β

Uniform for α and β. cuz it makes sense.

(v) Stochastic Sampling