Lindley’s paradox

Consider the result $x$ of some experiment, with two possible explanations, hypotheses $H_0$ and $H_1$, and some prior $\pi$ representing uncertainty as to which hypothesis is more accurate before taking into account $x$. Lindley’s paradox occurs when

1. The result is ”significant” by a frequentist test of $H_0$, indicating sufficient evidence to reject $H_0$, say, at the 5% level, and,

2. the posterior probability of $H_0$ given $x$ is high, indicating strong evidence that $H_0$ is in better agreement with $x$ than $H_1$.

These results can occur at the same time when $H_0$ is very specific, $H_1$ more diffuse, and the prior distribution does not strongly favor one or the other, as seen below.
An example

Suppose we flipped a coin and obtained 49,581 heads and 48,870. We assume that the number of heads is a binomial variable with parameter $\theta$. We are interested in testing whether $\theta$ is 0.5 or some other value. That is, our null hypothesis is $H_0 : \theta = 0.5$ and the alternative is $H_1 : \theta \neq 0.5$. 
An example

> binom.test(49581, 49581 + 48870)

   Exact binomial test

data:  49581 and 49581 + 48870
number of successes = 49581, number of trials = 98451,  p-value = 0.02365
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
  0.5004826 0.5067390
sample estimates:
  probability of success
    0.5036109
An example

We have no reason to favor one hypothesis over the other: \( \text{prob}(H_0|I) = \text{prob}(H_1|I) = 0.5 \). The posterior for \( H_0 : \theta = 0.5 \) is

\[
\text{prob}(H_0|k, I) = \frac{\text{prob}(k|H_0, I) \times \text{prob}(H_0|I)}{\text{prob}(k|H_0, I) \times \text{prob}(H_0|I) + \text{prob}(k|H_1, I) \times \text{prob}(H_1|I)}
\]

The likelihood functions can be calculated by

\[
\text{prob}(k|H_0, I) = \binom{n}{k}(0.5)^k(1 - 0.5)^{n-k} \approx 1.95 \times 10^{-4}
\]

\[
\text{prob}(k|H_1, I) = \int_0^1 \binom{n}{k} u^k(1 - u)^{n-k} du
\]

\[
= \binom{n}{k} \text{B}(k + 1, n - k + 1) \approx 1.02 \times 10^{-5},
\]

i.e. \( \text{prob}(H_0|k, I) \approx 0.95 \)
The lighthouse problem

A lighthouse is somewhere off a piece of straight coastline at a position $\alpha$ along the shore and a distance $\beta$ out at sea. It emits a series of short highly collimated flashed at random intervals and hence at random azimuths. These pulses are intercepted on the coast by photo-detectors that record the fact that a flash has occurred, but not the angle from which it came. $N$ flashes have so far been recorded at positions $\{x_k\}$. Where is the lighthouse?
The lighthouse problem

Fig. 2.7 A schematic illustration of the geometry of the lighthouse problem.
The lighthouse problem

Prior information $I$

- Random azimuths $\theta$

$$\text{prob}(\theta_k|\alpha, \beta, I) = \begin{cases} \frac{1}{\pi} & \text{if } -\frac{\pi}{2} \leq \theta_k \leq +\frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases}$$

- BUT: we do not know the azimuth $\theta$, all we know is $x_k$, the position along the coast line
The lighthouse problem

Fig. 2.7 A schematic illustration of the geometry of the lighthouse problem.
The lighthouse problem

Prior information $I$

- Random azimuths $\theta$

\[
\text{prob}(\theta_k | \alpha, \beta, I) = \begin{cases} 
\frac{1}{\pi} & \text{if } -\frac{\pi}{2} \leq \theta_k \leq +\frac{\pi}{2} \\
0 & \text{otherwise}
\end{cases}
\]

- BUT: we do not know the azimuth $\theta$, all we know is $x_k$, the position along the coast line

\[
\beta \tan \theta_k = x_k - \alpha
\]
Changing variables

Given that $Y = f(X)$, how is the pdf $\text{prob}(X|I)$ related to $\text{prob}(Y|I)$?
Changing variables

Given that $Y = f(X)$, how is the pdf $\text{prob}(X|I)$ related to $\text{prob}(Y|I)$?

**Fig. 3.13** Changing variables in one dimension: the function $f$ maps the point $X^*$ to $Y^* = f(X^*)$ and the small interval $\delta X$ to the corresponding region $\delta Y$. 
Changing variables

Given that \( Y = f(X) \), how is the pdf \( \text{prob}(X|I) \) related to \( \text{prob}(Y|I) \)?

\[
\text{prob}(X = X^*|I) \delta X = \text{prob}(Y = Y^*|I) \delta Y
\]

\[\text{prob}(X|I)\]

\[\text{prob}(Y|I)\]

Fig. 3.13 Changing variables in one dimension: the function \( f \) maps the point \( X^* \) to \( Y^* = f(X^*) \) and the small interval \( \delta X \) to the corresponding region \( \delta Y \).
Changing variables

Given that $Y = f(X)$, how is the pdf $\text{prob}(X|I)$ related to $\text{prob}(Y|I)$?

$$\text{prob}(X|I) = \text{prob}(Y|I) \left| \frac{dY}{dX} \right|$$
Changing variables

The lighthouse problem

\[
\text{prob}(\theta|\alpha, \beta, I) = \begin{cases} 
\frac{1}{\pi} & \text{if } -\frac{\pi}{2} \leq \theta \leq +\frac{\pi}{2} \\
0 & \text{otherwise}
\end{cases}
\]

and

\[\beta \tan \theta = x - \alpha\]
Changing variables

Lets differentiate $\beta \tan \theta = x - \alpha$ on both sides with respect to $x$:

$$\beta(1 + \tan^2 \theta) \frac{d\theta}{dx} = 1$$

i.e. with $\tan \theta = \frac{x - \alpha}{\beta}$

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

Thus, we arrive at

\[
\text{prob}(x|\alpha, \beta, I) = \text{prob}(\theta|\alpha, \beta, I) \left| \frac{d\theta}{dx} \right|
\]

\[
= \frac{\beta}{\pi \left[ \beta^2 + (x - \alpha)^2 \right]}
\]
Fig. 2.8 The Cauchy, or Lorentzian, distribution. It is symmetric with respect to the maximum, at \( x = \alpha \), and has a FWHM of twice \( \beta \).
The lighthouse problem

A lighthouse is somewhere off a piece of straight coastline at a position $\alpha$ along the shore and a distance $\beta$ out at sea. It emits a series of short highly collimated flashed at random intervals and hence at random azimuths. These pulses are intercepted on the coast by photo-detectors that record the fact that a flash has occurred, but not the angle from which it came. $N$ flashes have so far been recorded at positions $\{x_k\}$. Where is the lighthouse?
The lighthouse problem

\[ \text{prob}(\alpha, \beta|\{x_k\}, \beta, I) \propto \text{prob}(\{x_k\}|\alpha, \beta, I) \times \text{prob}(\alpha|\beta, I) \times \text{prob}(\beta|I) \]
The lighthouse problem

\[
\text{prob}(\alpha|\beta, I) = \begin{cases} A & \alpha_{\text{min}} \leq \alpha \leq \alpha_{\text{max}} \\ 0 & \text{otherwise} \end{cases}
\]

\[
\text{prob}(\beta|I) = \begin{cases} B & 0 < \beta \leq \beta_{\text{max}} \\ 0 & \text{otherwise} \end{cases}
\]

\[
\text{prob}\{x_k\}|\alpha, \beta, I) = \prod_{k=1}^{N} \text{prob}(x_k|\alpha, \beta, I)
\]
The lighthouse problem

\[ L = \ln [\text{prob}(\alpha, \beta|\{x_k\}, I)] = \text{constant} + \sum_{k=1}^{N} \{\ln(\beta) - \ln [\beta^2 + (x_k - \alpha)^2]\} \]
The lighthouse problem

\[ L = \ln \left[ \text{prob}(\alpha, \beta | \{x_k\}, I) \right] \]

\[ = \text{constant} + \sum_{k=1}^{N} \left\{ \ln(\beta) - \ln \left[ \beta^2 + (x_k - \alpha)^2 \right] \right\} \]

Case 1: \( \beta \) known, \( \beta = 1 \text{km} \)

\[ L = \ln \left[ \text{prob}(\alpha, \beta = 1 | \{x_k\}, I) \right] \]

\[ = \text{constant} - \sum_{k=1}^{N} \ln \left[ \beta^2 + (x_k - \alpha)^2 \right] \]
N = 1

\[ \text{prob}(\alpha|\beta, I) \]

N = 2

\[ \text{prob}(\alpha|\beta, I) \]

N = 3

\[ \text{prob}(\alpha|\beta, I) \]

N = 8

\[ \text{prob}(\alpha|\beta, I) \]

N = 64

\[ \text{prob}(\alpha|\beta, I) \]

N = 1024

\[ \text{prob}(\alpha|\beta, I) \]
The lighthouse problem

\[ L = \ln \left[ \text{prob}(\alpha, \beta|\{x_k\}, I) \right] \]

\[ = \text{constant} + \sum_{k=1}^{N} \left\{ \ln(\beta) - \ln \left[ \beta^2 + (x_k - \alpha)^2 \right] \right\} \]

Case 2: both \( \alpha \) and \( \beta \) unknown
The central limit theorem

Suppose \( \{X_1, X_2, \ldots \} \) is a sequence of i.i.d random variables with \( E[X_i] = \mu \) and \( Var[X_i] = \sigma^2 < \infty \). Then as \( n \) approaches infinity, the random variables \( \sqrt{n}(S_n - \mu) \) converge in distribution to a normal \( N(0, \sigma^2) \), with \( S_n = \frac{X_1 + \cdots + X_n}{n} \).
The central limit theorem
\[ \text{prob}(\alpha_{xk} | \beta, I) \]

**N = 1**

![Graph for N = 1](image)

**N = 2**

![Graph for N = 2](image)

**N = 3**

![Graph for N = 3](image)

**N = 8**

![Graph for N = 8](image)

**N = 64**

![Graph for N = 64](image)

**N = 1024**

![Graph for N = 1024](image)
The central limit theorem

Suppose \( \{X_1, X_2, \ldots \} \) is a sequence of i.i.d random variables with \( E[X_i] = \mu \) and \( Var[X_i] = \sigma^2 < \infty \). Then as \( n \) approaches infinity, the random variables \( \sqrt{n}(S_n - \mu) \) converge in distribution to a normal \( N(0, \sigma^2) \), with \( S_n = \frac{X_1 + \cdots + X_n}{n} \).
For a Cauchy distribution $\sigma^2$ is infinite and $\mu$ is not defined!