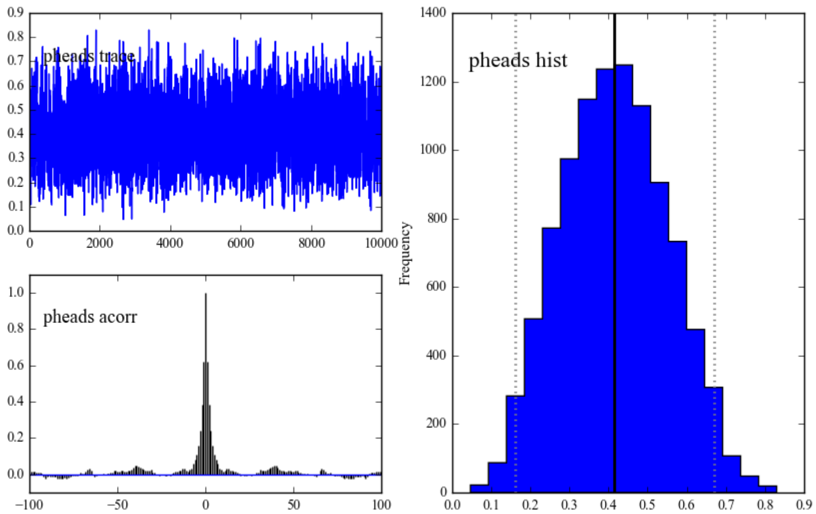


$$\hat{H} = 0.416 \pm 0.135 \qquad \text{95\% confidence interval: } [0.16, 0.67]$$

Figure 1: Coin model results with true $H = 0.38$, $n = 10$ coin tosses, an uniform prior, and a needlessly long $l = 10{,}000$ step chain. Since ten coin tosses is a very small dataset, the trace oscillates a lot, and the 95% confidence interval is large.

1 Coin tossing

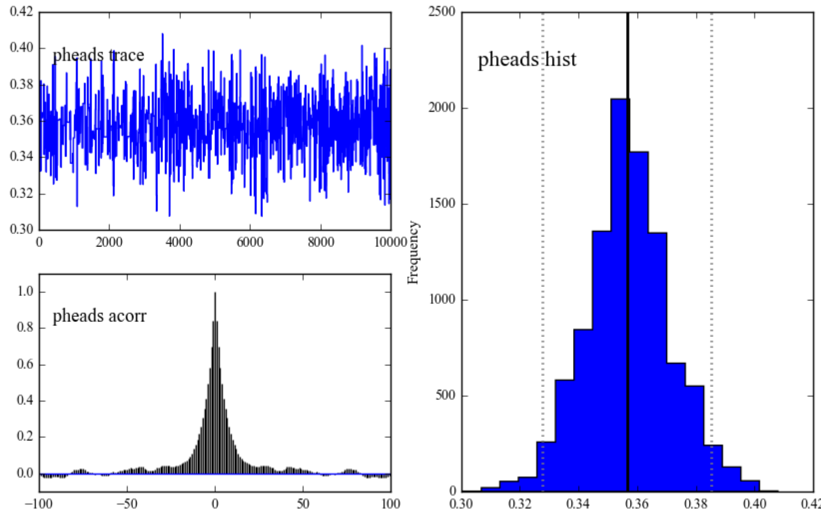
We consider a sequence of tosses of a biased coin. The coin can be modeled with a single parameter $H \in [0, 1]$, the probability of heads on a single toss. Thus our parameter set is $\mathbf{X} = \{H\}$.

The dataset $\mathbf{D}$ can be reduced to a tuple (n, h) where h is the number of heads in n tosses. There is no other prior information: $I = \emptyset$. The probability of obtaining some data given the parameter H is a binomial distribution:

$$\Pr(\mathbf{D}|\mathbf{X}, I) = \Pr\Big((n, h)|H\Big) = \binom{n}{h} H^h (1 - H)^{n-h},$$

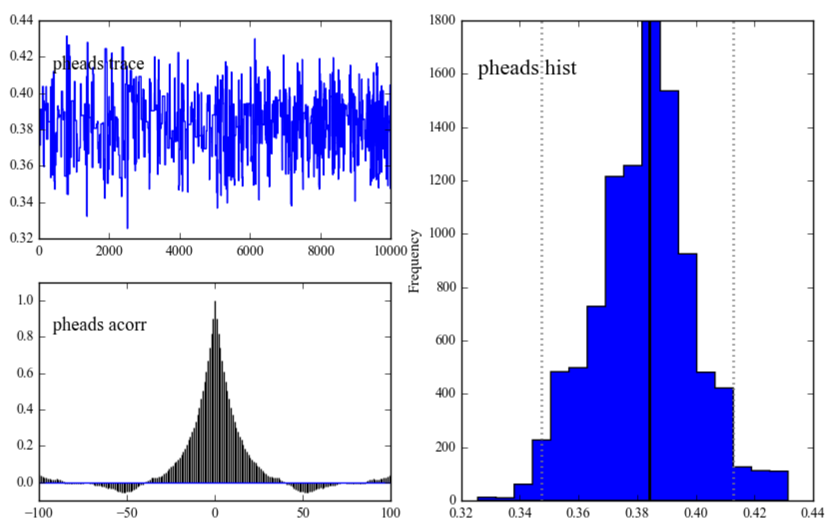
since there are $\binom{n}{h}$ ways of getting h heads (each with probability H) and $n - h$ tails (each with probability $1 - H$). In this simulation, we choose $H = 0.38$ and three prior distributions:

1. an uniform one,
2. a Gaussian with $\sigma = 0.2$ centered around $\mu = 0.5$, representing a belief that the coin is fair, and
3. a Gaussian with $\sigma = 0.05$ centered around $\mu = 0.5$, representing a stronger belief that the coin is fair.



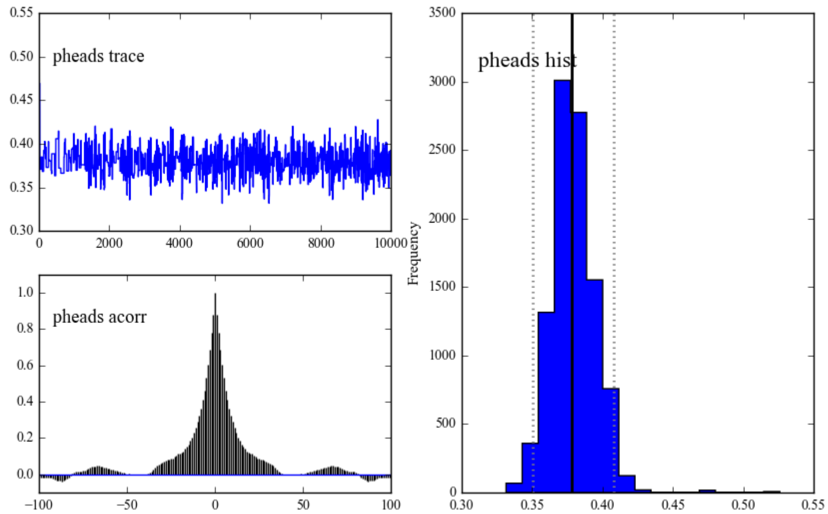
$\hat{H} = 0.357 \pm 0.014 \qquad$ 95% confidence interval: $[0.328, 0.385]$

Figure 2: Coin model results with true $H = 0.38$, $n = 1000$ coin tosses, an uniform prior, and a needlessly long $l = 10,000$ step chain. The larger dataset than in Figure 1 makes the confidence interval much narrower. The trace oscillation amplitude is much smaller, and the histogram shows a narrower peak.



$\hat{H} = 0.383 \pm 0.017 \qquad$ 95% confidence interval: $[0.347, 0.413]$

Figure 3: Coin model results with true $H = 0.38$, $n = 1000$ coin tosses, a wide Gaussian prior ($\mu = 0.5, \sigma = 0.2$), and a needlessly long $l = 10,000$ step chain. With 1000 tosses, the results from the data dominate over the effects from the prior.



$$\hat{H} = 0.379 \pm 0.016 \qquad \text{95\% confidence interval: } [0.35, 0.408]$$

Figure 4: Coin model results with true $H = 0.38$, $n = 1000$ coin tosses, a narrow Gaussian prior ($\mu = 0.5, \sigma = 0.05$), and a needlessly long $l = 10{,}000$ step chain. With 1000 tosses, the results from the data dominate over the effects from the prior – despite the intial narrowness of the peak. The trace shows that the algorithm starts near $H = 0.5$ but almost immediately reaches the desired region.

2 Lighthouse problem

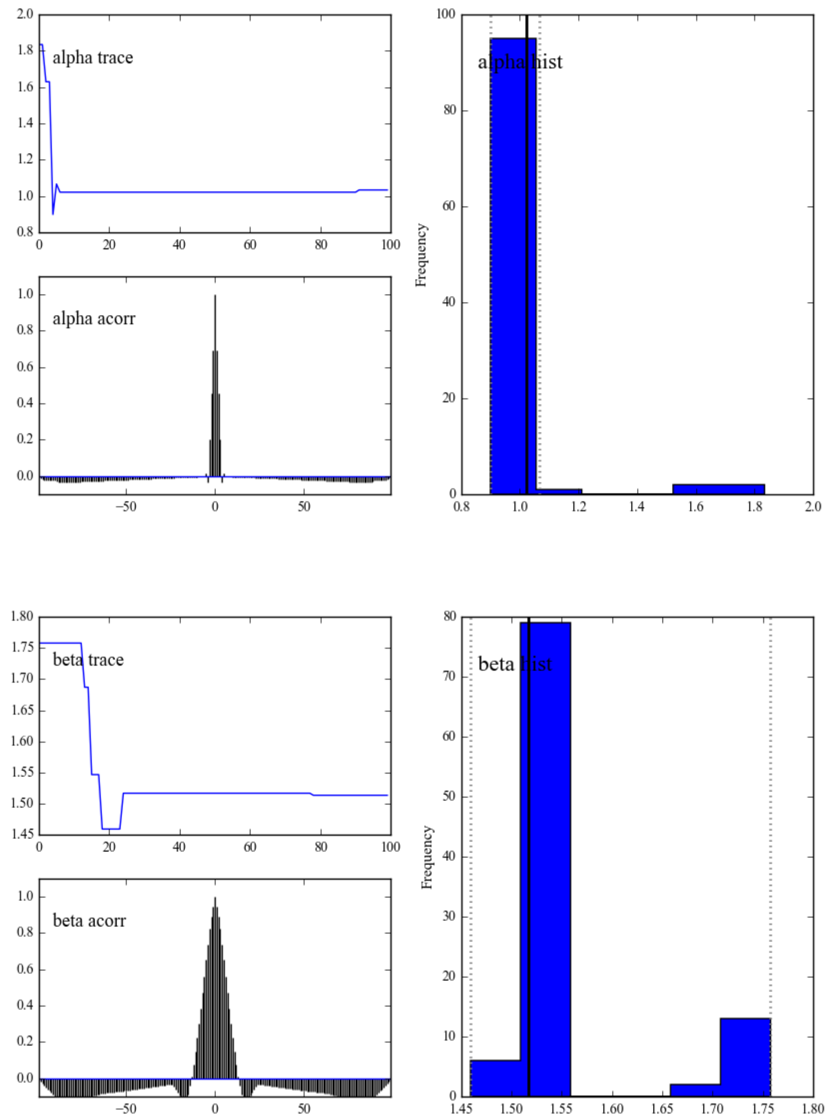
We have a lighthouse at a distance β from the shore and at a location α along the shore. The lighthouse emits flashes at random angles θ_k which arrive at the shore at locations x_k. In set 4, we derived the distribution of the x_k's: a Lorentzian. We manipulate it into the form required to use the Cauchy distribution implemented in `pymc`:

$$f_x(x) = \frac{\beta/2\pi}{(x-\alpha)^2 + \beta^2} = \frac{1}{2} \cdot \frac{1}{\pi\beta\left[1 + \left(\frac{x-\alpha}{\beta}\right)^2\right]}$$

For our study we choose "true" values $\alpha = 1.0$ and $\beta = 1.5$. We again use three prior distributions:

1. an uniform one,
2. a wide Gaussian centered at $\mu = 1.5$ with $\sigma = 0.3$,
3. and a narrow Gaussian centered at $\mu = 1.5$ with $\sigma = 0.1$.

Results for the one-dimensional case, where β is known but α is not, are shown in Figures 4-6. In the two-dimensional case, we use identical priors for each parameter. Results are shown in Figures 7-9.



$\hat{\alpha} = 1.051 \pm 0.141 \qquad \hat{\beta} = 1.548 \pm 0.086$

Figure 5: Lighthouse model results with true $(\alpha, \beta) = (1.0, 1.5)$, $n = 10000$ lighthouse flashes (so roughly 5000 data points on the shore), and a chain length of $l = 100$. The traces converge quickly – both are quite good by the 30th step. The `pymc` implementation of MCMC does not allow chain lengths of less than 100.