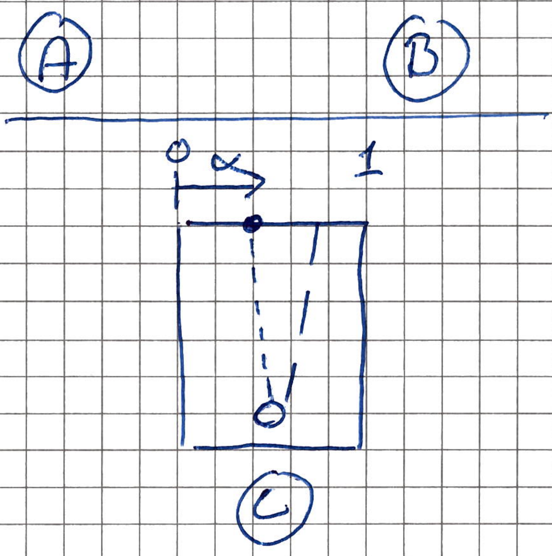
The Bayesian billiard game



Rules:

- 1st ball gives α
- Subsequent balls either on A:s or B:s side.
- All hits are equally probable.
- First to six points win.

Assume eight rolls:

Alice has five points

Bob has three

What is the probability that Bob will win?

- Frequentist approach (naive):
  - Needs an intermediate estimate
    $$\hat{\alpha} = \frac{5}{8}$$
  - $p(B) = (1-\hat{\alpha})^3$ — Needs three straight wins.
    $$\simeq 0.05$$

- Bayesian approach

  We will consider the joint probability
  $$p(B, \alpha \mid D, I)$$

and then marginalize over α

$$p(B|D,I) = \int d\alpha \, \underbrace{p(B,\alpha|D,I)}_{= p(B|\alpha,D,I)\, p(\alpha|D,I)}$$

Bayes' rule

$$p(\alpha|D,I) = \frac{p(D|\alpha,I)\, p(\alpha|I)}{p(D|I)}$$

We usually ignore the denominator in parameter estimation problems, where we seek a pdf. But here we actually seek a probability.

We can rewrite

$$p(D|I) = \int d\alpha \, p(D|\alpha,I)\, p(\alpha|I)$$

To summarize

$$p(B|D,I) = \frac{\int d\alpha \, p(B|\alpha,D,I)\, p(D|\alpha,I)\, p(\alpha|I)}{\int d\alpha \, p(D|\alpha,I)\, p(\alpha|I)}$$

We can compute all of these:

$$p(B|\alpha,D,I) = (1-\alpha)^3$$

$$p(D|\alpha,I) \propto \alpha^5(1-\alpha)^3 \quad \text{from Binomial dist.}$$

$$p(\alpha|I) = \begin{cases} 1 & \text{if } \alpha \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

We find

$$p(B|D,I) = \frac{\int_0^1 (1-\alpha)^6 \alpha^5 \, d\alpha}{\int_0^1 (1-\alpha)^3 \alpha^5 \, d\alpha}$$

these are special cases of the Beta function

$$B(n,m) = \int_0^1 (1-t)^{n-1} t^{m-1} \, dt$$

$$\Rightarrow p(B|D,I) = 0.09$$

Marginal posterior

We have $p(\theta_0, \theta_1 | D, I)$

but are interested in $p(\theta_1 | D, I)$

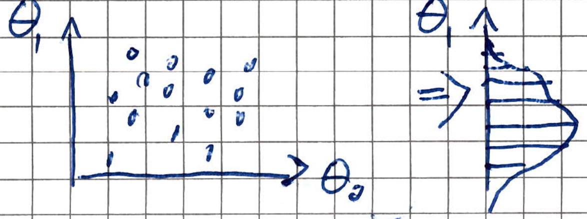
This marginal posterior can be obtained via marginalization

$$p(\theta_1 | D, I) = \int d\theta_0 \, p(\theta_0, \theta_1 | D, I)$$

Assuming that we are happy with N samples of the pdf on the lhs : $\{\theta_{1,i}\}_{i=1}^{N}$; we can get those from N samples $\{(\theta_0, \theta_1)_i\}_{i=1}^{N}$ of the joint distribution by just ignoring the θ_0 column

$$\begin{pmatrix} \theta_{0,1} & \theta_{1,1} \\ \theta_{0,2} & \theta_{1,2} \\ \vdots & \vdots \\ \theta_{0,N} & \theta_{1,N} \end{pmatrix} \rightsquigarrow \begin{pmatrix} \theta_{1,1} \\ \theta_{1,2} \\ \vdots \\ \theta_{1,N} \end{pmatrix}$$

or



Systematic error example

(See notebook if short on time)

Hubble's law

$$v = H_0 x \,,$$ where x is the distance to e.g. a galaxy and v its recessional velocity

Assume $v_{measured} = (100 \pm 5) \times 10^3$ km/sec

Assume H_0 is known from previous analysis

$\Rightarrow$ Determine posterior pdf for distance x to this galaxy assuming:

Case 1: H_0 is (supposedly) known exactly

Case 2: H_0 is known through $p(H_0 | I)$

We assume that

$$v_{measured} = v_{theory} + \delta v_{exp}$$

where $v_{theory} = H_0 x$

$$\delta v_{exp} \sim N(0, \sigma_v)$$ [Here $\sigma_v = 5 \cdot 10^3$ km/sec]

We also assume that δv_{exp} is uncorrelated with the error in H_0.

Case 1:

$$p(x|D,I) \propto p(D|x,I)\, p(x|I)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_v} \exp\left[-\frac{(v_{measured} - v_{theory})^2}{2\sigma_v^2} \right] p(x|I)$$

$$= \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left[-\frac{(v_{measured} - H_0 x)^2}{2\sigma_v^2}\right] & \text{if } x \in [x_{min}, x_{max}] \\ 0 & \text{otherwise} \end{cases}$$

where we have assumed a uniform prior in $[x_{min}, x_{max}]$

Case 2:

Here we know $p(H_0 | I)$ either via some functional form, or via a finite number of samples $\{H_i\}_{i=1}^{N}$ generated by a MCMC sampler

We will obtain the desired pdf $p(x | D, I)$ via marginalization of the joint distribution $p(x, H_0 | D, I)$

$$p(x | D, I) = \int dH_0 \underbrace{p(x, H_0 | D, I)}_{\left\{ \begin{array}{l} \propto p(D | x, H_0, I)\, \underbrace{p(x, H_0 | I)} \\ = \underbrace{p(H_0 | x, I)}_{= p(H_0 | I)}\, p(x | I) \end{array} \right\}}$$

$$\propto p(x | I) \int_{-\infty}^{\infty} dH_0\, p(H_0 | I)\, p(D | x, H_0, I)$$

We have expressed the quantity that we seek in terms of quantities that we know

- $p(D|X, H_0, I)$ is the same as in Case 1 (where H_0 was part of I)
- $p(H_0|I)$

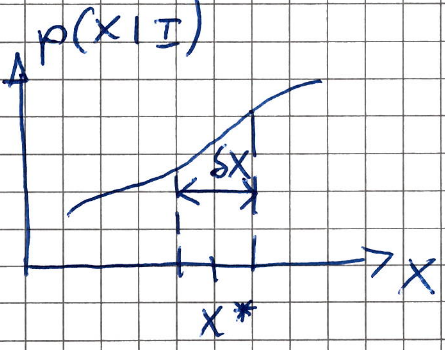
Marginalization with finite number of samples

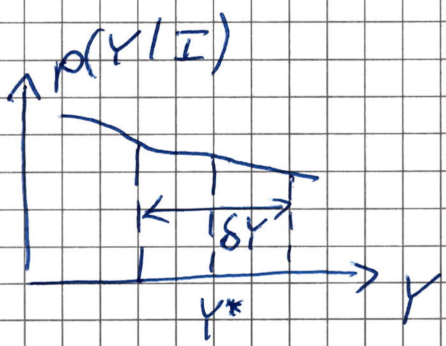
$$p(X|D, I) \propto \int dH_0 \, p(H_0|I) \, p(D|X, H_0, I)$$
$$\approx \frac{1}{N} \sum_{i=1}^{N} p(D|X, H_i, I)$$

(where we have used $p(X|I) \propto 1$)

Error propagation

Assume there is an uncertainty in parameter X, how does that translate to an uncertainty in quantity $Y = f(X)$?





$$p(X=X^*|I)\,\delta X = p(Y=Y^*|I)\,\delta Y$$

where $Y^* = f(X^*)$

must be true for all X^*

$$\Rightarrow p(X|I) = p(Y|I) \times \left|\frac{dY}{dX}\right|$$

in the limit $\delta X, \delta Y \to 0$

- A very usual (and useful) shortcut:

  Assuming that we are interested in small deviations $\delta X = X - X_0$ from an optimal solution X_0 that is well described by a Gaussian

Then

$$\langle \delta X^2 \rangle = \sigma_X^2 \quad ; \quad \langle \delta Y^2 \rangle = \sigma_Y$$

Example: square root

Assume that the amplitude of a signal peak (Bragg peak) has been measured / estimated using a least-squares MLE

$$A = A_0 \pm \sigma_A$$

A is related to the complex structure factor

$$A = |F|^2 \equiv f^2$$

We want to extract

$$f = f_0 \pm \sigma_f$$

Obviously $f_0 = \sqrt{A_0}$

Differentiate, square, expectation

$$\underbrace{\langle \delta A^2 \rangle}_{=\sigma_A^2} = 4 f_0^2 \langle \delta f^2 \rangle = 4 A_0 \underbrace{\langle \delta f^2 \rangle}_{=\sigma_f^2}$$

$$\rightarrow \quad f = \sqrt{A_0} \pm \frac{\sigma_A}{2\sqrt{A_0}}$$

What happens if the best fit gives $A_0 < 0$?