Bayesian Updating

Consider first the case of coin tosses.

- you are trying to estimate p, the probability of heads
- you need a prior density for p, call it $\pi(p)$
- your data is k, the number of heads in n tosses
- you want the posterior density for p, $\pi(p|k)$

If you choose your prior to be a Beta(α, β) distribution:

$$\pi(p) = \frac{(\alpha + \beta - 1)!}{(\alpha - 1)!(\beta - 1)!} \; p^{\alpha-1}(1-p)^{\beta-1}$$

then your posterior has a Beta$(\alpha + k, \beta + n - k)$ distribution.

Updating is simple: you add the number of heads k to α, and the number of tails $n - k$ to β.

This gives you a density on [0,1], which describes your opinion of p after seeing the data.

If you want a single estimate for p, it makes sense to take the value where the density is highest, which happens to be

$$\frac{\alpha + k - 1}{\alpha + \beta + n - 2}$$

If you take various starting values for α and β, this gives you various estimates. If your prior is reasonable, the estimates are reasonable. In some cases, they may be better than the "obvious" estimator k/n.

The posterior probability that p falls in a certain interval is just the relevant area under the density curve. Hence you can get confidence intervals for p.

Bayesian Updating: The General Case

- you want to estimate some parameter θ
- first you need a prior $\pi(\theta)$
- then you observe the data X; you need to know its distribution conditional on θ, $f(x|\theta)$
- then the posterior is $\pi(\theta|x) = C \cdot \pi(\theta) \cdot f(x|\theta)$ where the constant C depends on x (not on θ) and is chosen so that the posterior integrates to 1

The hard part is usually computing $C = 1/\int \pi(\theta) \cdot f(x|\theta) d\theta$:

- sometimes you can choose your prior in such a way that C can be computed explicitly (e.g. beta distribution for Bernoulli trials)

- with a computer, you can try to do the integral by brute force

- a smarter way is a technique called MCMC (Markov chain Monte Carlo) that computes the posterior indirectly as a byproduct of running a Markov chain. This is a fashionable approach.