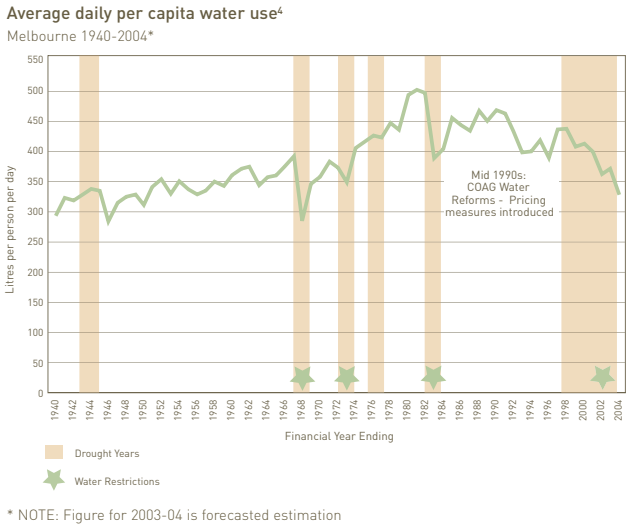
MATH3871/MATH5960
Bayesian Inference and Computation

Lab 1 Exercises

These exercises provide some practice in performing basic Bayesian analyses. There is no requirement to do the exercises in order. Outline solutions are available in a separate file.

1) Water consumption



In the Melbourne average daily per capita water use analyis, we modelled the discrete observations $x_1, \ldots, x_n$ as independent draws from a Poisson(θ) distribution. Assuming a Gamma(α, β) prior, which has a density function of

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta), \qquad \text{for } \alpha, \beta, \gamma > 0,$$

we computed the posterior as a Gamma $(\alpha + \sum_{i=1}^n x_i, \beta + n)$ distribution.

(a) Given that $n = 65, \sum_i x_i = 24,890$ and with prior parameters $\alpha = 1, \beta = 0.01$, compute a point estimate (i.e. the posterior mean) and a 95% central credible interval for θ. Note, you will need to compute the credible interval numerically in `R` (hint: use the `R` command `qgamma`).

(b) Draw a sample of size $N = 500$ directly from the posterior distribution (see the `R` command `rgamma`), and obtain Monte Carlo estimates of the lower and upper values of the 95% credible interval for θ.

Repeat this 250 times, and produce a histogram for the distribution of each interval endpoint. Superimpose a point corresponding to the true interval endpoints. How accurate is the Monte Carlo estimate? (`R` commands: `hist`, `points`).

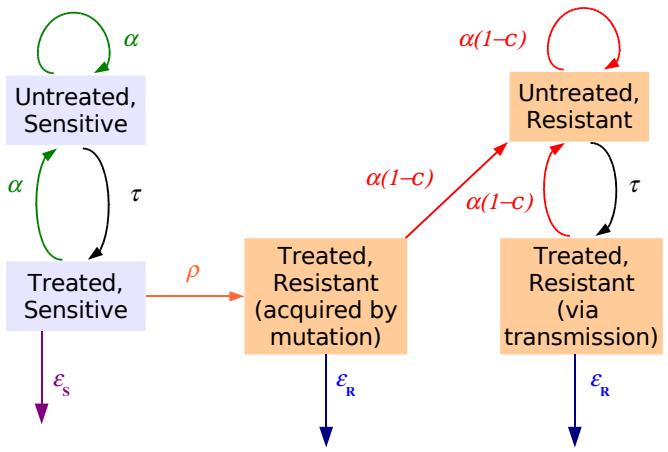
Produce another pair of histograms, but this time use $N = 5000$ samples. How is the precision of the Monte Carlo estimates affected?

How many samples, N, are needed for the spread (i.e. min - max) of the Monte Carlo estimates for each interval endpoint to be less than 0.15?

(c) In lectures it was stated that the predictive distribution for a future observation, y, is $\text{NegBin}\left(y \mid \alpha + \sum_i x_i, \frac{1}{\beta+n+1}\right)$. Draw samples directly from the posterior distribution. Use these to obtain samples from the posterior predictive distribution, and plot this via a histogram. Superimpose the density of the algebraically computed negative binomial predictive distribution (`R` command: `dnbinom`). Do the distributions coincide?

(d) What are the advantages/disadvantages of performing statistical analyses using the algebraically exact approach, and the Monte Carlo approximations?

2) Estimating evolutionary fitness of tuberculosis



Luciani et al. (2009) developed a stochastic model (above) to estimate epidemiological parameters relating to the development of drug resistance in *mycobacterium tuberculosis*. Unknown model parameters included the transmission rate (α), the marker mutation rate (μ), the mutation rate of drug resistance (ρ) and the transmission cost due to resistance (c). The rates of cure due to treatment for sensitive (ε_S) and resistant (ε_R) strains, and the detection and treatment rate (τ) are held fixed.

Samples from the posterior distribution when analysing the *IS6110* marker from Cuban data can be found in the file `tuberculosis.txt` (the rows correspond to α, c, ρ and μ in order).

(a) Read the posterior into `R` (using the command `read.table`). Produce marginal posterior histograms of each parameter, and scatterplots of the 6 bivariate distributions (e.g. (α, c), $\ldots$, (ρ, μ)).

(b) The *relative fitness* of the drug-resistant strains based on the model of Luciani et a. (2009) can be expressed as

$$\Phi = (1-c)\frac{\frac{1}{\tau} + \frac{1}{\delta+\varepsilon_R}}{\frac{1}{\tau} + \frac{1}{\delta+\varepsilon_S+\rho}}.$$

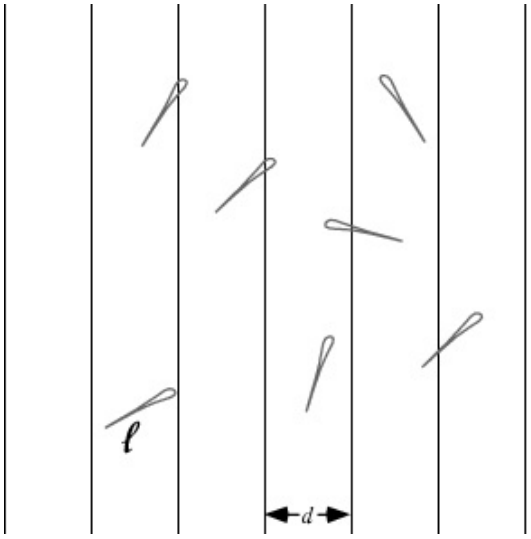
If $\delta = \tau = \varepsilon_S = 0.52$ and $\varepsilon_R = 0.202$ are fixed, then produce a histogram of the posterior distribution of Φ. Calculate $\Pr(\Phi < 1)$, the posterior probability that $\Phi < 1$. Is there any evidence that the resistant strain is any less evolutionarily fit than the susceptible strain? (i.e. is there any evidence that $\Phi < 1$?)

(c) A *central* 95% credible interval for ρ can be estimated by discarding the lower and upper 2.5% of the posterior samples. Obtain such a 95% interval for ρ and comment on its length.

(d) A 95% credible interval is *any* interval for which the interval contains 95% of the posterior density. For example, $(q_{0.01}, q_{0.96})$, where q_x is the x-th quantile of a parameter, is also a 95% credible interval. As there are (in theory) an infinite number of 95% credible intervals for any parameter, convention a useful strategy is to take the *shortest* one.

Based on the posterior sample of length 5000, compute 250 unique 95% credible intervals for ρ, and identify the shortest. How does this compare to using the central 95% credible interval? Under what circumstances is the central 95% credible interval likely to be close to the shortest length?

3) Buffon's Needle



One of the most famous simulation experiments is Buffon's Needle, designed to calculate (not very efficiently!) an estimate of π. Imagine a grid of parallel lines with spacing d, on which a needle of length $\ell \leq d$ is dropped. We repeat this experiment n times and count the proportion of times, $\hat{p}$, that the needle intersects with a line.

The rationale behind this is that if x is the distance from the centre of the needle to the leftmost line, and if θ is the angle from the vertical, then under the assumption of random needle throwing, we would have $x \sim U(0, d)$, and $\theta \sim U(0, \pi)$. Hence

$$\begin{aligned} p &= \Pr(\text{needle intersects line}) \\ &= \frac{1}{\pi}\int_0^{\pi} \Pr(\text{needle intersects}|\theta = \phi)d\phi \\ &= \frac{1}{\pi}\int \left(\frac{2}{d} \times \frac{\ell}{2}\sin\phi\right) d\phi \\ &= \frac{2\ell}{\pi d}. \end{aligned}$$

Hence, an estimate of π is $\hat{\pi} = \frac{2\ell}{\hat{p}d}$.

(a) Produce some code to simulate the Buffon's Needle experiment, given the lengths ℓ and d, and produce an estimate of π. Plot the estimate of π as the number of simulations, n, increases.

(b) A natural question is how to optimise the relative sizes of ℓ and d. Consider the variability of $1/\hat{\pi}$.

Now $n\hat{p} \sim \text{Bin}(n, p)$, so $\text{Var}(\hat{p}) = p(1-p)/n$. Show that $\text{Var}(1/\hat{\pi}) = \text{Var}(\hat{p}d/2\ell) = \ldots = \frac{1}{n\pi^2}\left(\frac{\pi}{2\rho} - 1\right)$ where $\rho = \ell/d$. When is this minimised (for $0 \leq \rho \leq 1$)?

(c) By computing the estimate of π 1000 times and computing the standard deviation, for a range of values of $\rho = \ell/d$, empirically demonstrate that your optimal value of ρ leads to the smallest variability for $\hat{\pi}$.

There are a number of things which may (or may not!) improve the efficiency of this experiment, including:

- using a grid of rectangles or squares;
- using a cross or other shape instead of a needle
- using a needle of length greater than the grid separation.

The point is: simulation can be used to answer many interesting problems, but careful design may be needed to achieve even moderate efficiency.