ICIC DATA ANALYSIS WORKSHOP

DAY 1 PROBLEMS

Simple problems:

0. We wish to estimate the mean μ of a population $\{x_i\}$, $i = 1 \ldots N$, each independent and with variance σ^2. Defining an estimator $\hat{\mu} = \sum_i w_i x_i$ for some weights w_i, show that for the estimator to be unbiased (i.e. $\langle \hat{\mu} \rangle = \mu$, we require $\sum_i w_i = 1$. By minimising the variance of $\hat{\mu}$, show that the minimum variance estimator is $\hat{\mu} = (1/N) \sum_i x_i$, and show that its distribution for repeated trials is a gaussian centred on μ with a variance is σ^2/N.

Hints: (a) for sums of independent variables, variances add, (b) you will need a Lagrange multiplier.

Note: this looks very similar to the posterior for μ given a dataset, but the interpretation is very different. Here you have computed the distribution of the estimator $\hat{\mu}$ for repeated trials drawn from a population with expectation value μ; the Bayesian calculation gives the posterior for the true value, μ, given the data. $p(\mu|x)$ and $p(\hat{\mu}|\mu)$ are not the same, even if the expressions look somewhat similar.

1. The distribution of flux densities of extragalactic radio sources is a power-law with slope $-\alpha$, say, so the likelihood to measure a source flux S is $p(S|\alpha) \propto S^{-\alpha}$, above some (known) instrumental limiting flux density of S_0. In a non-evolving Euclidean universe $\alpha = 3/2$ and departure of α from the value 3/2 is evidence for cosmological evolution of radio sources (we assume measurement errors are negligible). This was the most telling argument against the steady-state cosmology in the early 1960s (even though they got the value of a wrong by quite a long way).

- Given observations of radio sources with flux densities S, what is the most probable value of α, assuming a uniform prior? (Hint: in this case you will have to normalise $p(S|\alpha)$).
- Show that if a single source is observed, and the flux is $2S_0$, that the most probable value of α is 2.44.
- By examining the second derivative of the posterior, estimate the error on α to be 1.44.
- Plot out the posterior of α. How good is the second derivative as a guide to the uncertainty?

2. Solve the 'Monty Hall' problem given in the lectures, using Bayes' theorem.

More involved problems:

3. An astronomical source emits photons with a Poisson distribution, at a rate of λ per second. A telescope detects the photons independently, with probability p. In time t, the source emits M photons, and N are detected. Show that the joint probability of N and M is

$$P(M,N) = \frac{\mu^M}{M!}\, e^{-\mu}\, \frac{M!}{N!(M-N)!} p^N q^{M-N} \tag{1}$$

where $\mu = \lambda t$ and $q = 1 - p$.

Marginalise over M to show that

$$P(N) = \frac{p^N q^{-N} e^{-\mu}}{N!} \sum_{M=N}^{\infty} \frac{(q\mu)^M}{(M-N)!} \tag{2}$$

Sum the series[1] to show that N has a Poisson distribution with expectation value $p\mu$. Why could this have been anticipated?

Calculate the probability that the source has emitted M photons given that N have been detected, $P(M|N)$, for $M \geq N$, and deduce that $M - N$ also has a Poisson distribution, and compute the expectation value for $M - N$.

[1] Remember that $\sum_{i=0}^{\infty} \frac{x^i}{i!} = e^x$.