Astronomical Statistics: Tutorial Solutions 3

John Peacock
C20 Royal Observatory; jap@roe.ac.uk

1. Data: $(v_1, r_1) = (0, 1)$ and $(v_2, r_2) = (1, 2)$, with errors $\sigma_i = 1$ on r_1 and r_2. Errors on v_1 and v_2 are negligible.

   (i) Test Hypothesis that $r_i = 0$. Method: compute χ^2, and calculate probability that χ^2 would be larger than observed value by chance (χ^2 is on the high side).

$$\chi^2 = \sum_{i=1}^{2} \frac{r_i^2}{\sigma_i^2} = 5. \tag{1}$$

   There are TWO degrees of freedom, $\nu = n - n_p = 2$, since there are $n_p = 0$ free parameters in the model, and n=2 data points. Hence

$$p(\chi^2|\nu) = \frac{e^{-\chi^2/2}}{2\Gamma(1)} = \frac{e^{-\chi^2/2}}{2}. \tag{2}$$

   Probability that $\chi^2 > 5$ is

$$p(\chi^2 > 5|\nu = 2) = \frac{1}{2}\int_5^\infty d\chi^2 e^{-\chi^2/2} = e^{-5/2} = 0.08. \tag{3}$$

   Can exclude the hypothesis with 92% confidence (or 84% if you accept that a very low value of χ^2 is just as unlikely). This is not very strong - generally results with less than 99% confidence are regarded with suspicion, and some people insist on more).

   (ii) Parameter Estimation. If we assume uniform prior for m and c (i.e. joint $p(m, c)$ =constant), then $p(m, c|D) \propto p(D|m, c)$, the likelihood, L. ($D$=the data). For *gaussian* errors,

$$L \propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{2}\frac{[r_i - (mv_i + c)]^2}{\sigma_i^2}\right\}. \tag{4}$$

   Multiplying out the exponent, we get

$$\begin{aligned} -2\ln L &= \textit{constant} + (1-c)^2 + [2-(m+c)]^2 \\ &= \textit{constant} + 2c^2 - 6c + 2mc + m^2 - 4m + 5. \end{aligned} \tag{5}$$

   The most probably values m and c are the *maximum likelihood* estimate, given by $\partial \ln L/\partial c = 0$ and $\partial \ln L/\partial m = 0$, which give two simultaneous linear equations

$$\begin{aligned} 4c - 6 + 2m &= 0 \\ 2m - 4 + 2c &= 0 \end{aligned} \tag{6}$$

with solution $\underline{m = c = 1}$.

Assuming the likelihood has a gaussian shape, the *conditional* error on c (i.e. keeping m fixed at its maximum likelihood value $m = 1$ is given by

$$\sigma_c^{-2} = - \left.\frac{\partial^2 \ln L}{\partial c^2}\right|_{c=1} = 2 \Rightarrow \sigma_c = \frac{1}{\sqrt{2}} \tag{7}$$

Similarly, the conditional error on m is 1. This expression comes from a Taylor expansion of $\ln L$ about the maximum (so the first derivative term vanishes):

$$\begin{aligned} \ln L &= constant + \frac{1}{2} \left.\frac{\partial^2 \ln L}{\partial c^2}\right|_{c=c_{max}} (c - c_{max})^2 + \ldots \\ \Rightarrow L &\propto \exp\left(\frac{1}{2} \left.\frac{\partial^2 \ln L}{\partial c^2}\right|_{c=c_{max}} (c - c_{max})^2 \right) \end{aligned} \tag{8}$$

and comparison with the standard form of the gaussian allows us to identify the conditional error in c as above.

For the *marginal* error, we allow for the fact that m is *not* fixed at $m = 1$, and this variation inevitably increases the error on c. We need to invert the (negative) second-derivative matrix (see notes),

$$H_{ij} = -\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \Rightarrow H^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \tag{9}$$

and take:

$$\sigma_c = (H^{-1})_{11}^{1/2} = 1; \sigma_m = (H^{-1})_{22}^{1/2} = \sqrt{2}. \tag{10}$$

Note that it is almost always correct to quote the *marginal* error, not the conditional error.

(iii) Model selection.

The Bayesian evidence ration is

$$B_{12} \equiv \frac{p(M_1|D)}{p(M_2|D)} = \frac{\int d\theta_1 p(D|\theta_1 M_1) p(\theta_1|M_1)}{\int d\theta_2 p(D|\theta_2 M_2) p(\theta_2|M_2)} \frac{p(M_1)}{p(M_2)} \tag{11}$$

where we note that model 1 has only one parameter (c), whereas model 2 has 3 (m and c).

Now, for model selection (as opposed to parameter estimation) we have to pay some attention to the normalisation of the priors, even if flat. If we assume that c and m are assumed to have flat priors within finite ranges Δc and Δm, then the prior on the single parameter c in model 1 is $p(c|M_1) = 1/\Delta c$, so that it integrates properly to unity. In model 2, the prior on m and c is $p(m, c|M_2) = (\Delta c \Delta m)^{-1}$.

Thus, assuming flat priors on the models, $p(M_1) = p(M_2)$,

$$\begin{aligned} B_{12} &= \Delta m \frac{\int dc \exp\left\{ -\frac{1}{2} \sum_{i=1}^2 (r_i - c)^2 \right\}}{\int dc\, dm \exp\left\{ -\frac{1}{2} \sum_{i=1}^2 (r_i - [mv_i + c])^2 \right\}} \\ &= \Delta m \frac{\int dc e^{-\frac{1}{2}\left[(1-c)^2 + (2-c)^2\right]}}{\int dc\, dm e^{-\frac{1}{2}[(1-c)^2 + (2-\{m+c\})^2]}}. \end{aligned} \tag{12}$$

The denominator is just 2π, since it is integrated easily by substituting $u = 2 - (m + c)$ and $v = 1 - c$. The jacobian of the transformation is unity, and the top is just $\int du\, dv e^{-(u^2+v^2)/2}$, which factorises. Note the prior on c has cancelled out.

The denominator is computed by completing the square $(c-3/2)^2 - 1/4$ and substituting $z = c - 3/2$. The resulting integral

$$\int_{-\infty}^{\infty} dz e^{-z^2} = \sqrt{\pi}. \tag{13}$$

Hence for $\Delta m = 5$, $B_{12} = 5\frac{e^{-1/4}}{2\sqrt{\pi}} = 1.1$. So the first model is very slightly favoured. Note that this reveals an unsatisfactory aspect of Bayesian methods - our conclusion depends heavily on the prior for m, which we may have few grounds to set.

2. We have to consider the joint probability of x, y and z, and marginalise over z to get $p(x,y|m,c) = \int dz p(x,y,z|m,c)$. We need to decide how to split the probability of x, y and z. A moment's thought tells us that x depends on z (its true value), via a gaussian error distribution, and y depends on the value of z through the linear relation $y = mz + c$ (plus noise), so it makes sense to split as follows (this is just the *product rule*; the m and c are common to all the probabilities):

$$p(x,y,z|m,c) = p(x,y|z,m,c)p(z|m,c). \tag{14}$$

If we take a flat prior for z, then the likelihood is

$$p(x,y|m,c) \propto \int_{-\infty}^{\infty} dz p(x|z,m,c)p(y|z,m,c) \tag{15}$$

which is, for gaussian errors with rms σ_x and σ_y,

$$p(x,y|m,c) = \frac{1}{2\pi\sigma_x\sigma_y}\int_{-\infty}^{\infty} dz \exp\left[-\frac{(x-z)^2}{2\sigma_x^2}\right]\exp\left[-\frac{(y-\{mz+c\})^2}{2\sigma_y^2}\right]. \tag{16}$$

With $\sigma_x = \sigma_y = 1$, the probability of m and c, given a single data point, is proportional to the likelihood, assuming uniform priors. (We get the full probability of the data*set* by multiplying the probabilities of each *individual* data point, assuming they are independent).

Letting $u \equiv x - z$:

$$p(m,c|x,y) \propto \exp\left(-\frac{u^2}{2}\right)\exp\left\{-\frac{[y-m(x-u)-c]^2}{2}\right\}. \tag{17}$$

Multiplying the second bracket out to isolate the u^2 term, and combining it with the first gives

$$p(m,c|x,y) \propto \exp\left[-\frac{u^2(1+m^2)}{2}\right]\exp\left\{-\frac{[y-mx-c]^2}{2}\right\}\exp\left[-\frac{1}{2}(y-mx-c)2um\right]. \tag{18}$$

Letting $v = y - mx - c$, and completing the square for the u integration,

$$p(m,c|x,y) \propto \exp\left[-\frac{1}{2}\left(u\sqrt{1+m^2} + \frac{vm}{\sqrt{1+m^2}}\right)^2\right]\exp\left\{-\left[\frac{v^2m^2}{2(1+m^2)}\right]\right\}\exp\left(-\frac{v^2}{2}\right). \tag{19}$$

Integrating over u (change variable again...), and including all the points:

$$p(m,c|\{x_i,y_i\}) \propto \prod_i (1+m^2)^{-1/2}\exp\left[-\frac{(y_i-mx_i-c)^2}{2(1+m^2)}\right]. \tag{20}$$

3. Malmquist Bias arises when we have a sample of objects which are selected, or detected, on their apparent luminosity. In the case of stars or galaxies there is a limiting luminosity below which we do not see the object as it is too faint. Since the apparent luminosity, L, is given by $L = 4\pi S D^2$, where S is the flux of the object and D is its distance, this correlates luminosity and distance. As this happens independently of the wavelength, the luminosities at each wavelength will all become correlated with the distance and consequently with each other. One way to avoid this is to correlate fluxes (rather than luminosities) in different wavelengths instead.

4. Use the product rule:
$$P(M, N) = P(M)P(N|M) \tag{21}$$
and the first term is a Poisson distribution (given), with parameter (=mean) $\mu = \lambda t$. The second term is just a binomial - the probability of detecting N photons from M, if the individual probability is p.

Thus
$$P(M, N) = \frac{\mu^M}{M!} e^{-\mu} \; \frac{M!}{N!(M-N)!} p^N q^{M-N} \tag{22}$$

To get $P(N)$, marginalise over M, noting that M is *discrete*, so it is a sum, not an integral:
$$P(N) = \sum_{M=N}^{\infty} P(M, N) \tag{23}$$

where we note that at least M photons must be emitted if N are detected... Substituting for $P(M, N)$, and letting $i = M - N$,
$$P(N) = \sum_{M=N}^{\infty} P(M, N) = \frac{p^N q^{-N} e^{-\mu} (\mu q)^N}{N!} \sum_{i=0}^{\infty} \frac{(\mu q)^i}{i!}. \tag{24}$$

The sum is just $e^{\mu q}$ (make sure you can spot these!), hence
$$P(N) = \frac{(\mu p)^N e^{-p\mu}}{N!}, \tag{25}$$

which is just a Poisson distribution with mean μp.

The last part follows similar lines, and is left for you to work out. If you can't do it, come and see me.