1. Consider samples x and y and let M_1 denote the model considering X and Y independent with binomial distributions

$$P(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad P(Y = y) = \binom{n}{y}\eta^y(1-\eta)^{n-y},$$

where $0 \leq \theta, \eta \leq 1$ are both unknown.

Let similarly M_2 denote the model where $\theta = \eta$ and everything else is as for M_1.

(a) Calculate the maximized log-likelihood ratio statistic $D = -2\log\Lambda$ for comparing the two models;

Under M_1,

$$\hat{\theta} = x/n, \quad \hat{\eta} = y/n$$

whereas under M_2,

$$\hat{\theta}_2 = \hat{\eta}_2 = (x+y)/(2n)$$

so

$$\Lambda = \frac{L(\hat{\theta}_2)}{L(\hat{(\theta)}L(\hat{\eta})} = \frac{\{(x+y)/(2n)\}^{x+y}\{1-(x+y)/(2n)\}^{2n-x-y}}{(x/n)^x(1-x/n)^{n-x}(y/n)^y(1-y/n)^{n-y}},$$

yielding the familiar (?) expression

$$\begin{aligned} D &= -2\log\Lambda = 2\sum \text{OBS}\log\frac{\text{OBS}}{\text{EXP}} \\ &= 2x\log\frac{2x}{x+y} + 2y\log\frac{2y}{x+y} \\ &\quad +(2n-x)\log\frac{2n-2x}{2n-x-y} + 2(n-x)\log\frac{2n-2x}{2n-x-y}. \end{aligned}$$

(b) For uniform prior distributions on θ, η, calculate the Bayes factor for comparing M_1 to M_2;

We have

$$\int_0^1 \theta^x(1-\theta)^{n-x}\,d\theta = \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)} = \frac{x!(n-x)!}{(n+1)!}$$

and thus get

$$B_{12} = \frac{f(x,y\,|\,M_1)}{f(x,y\,|\,M_2)} = \frac{\binom{2n+1}{n+1}}{\binom{x+y}{x}\binom{2n-x-y}{n-x}}.$$

(c) Find the BIC approximation to the Bayes factor, with or without including all terms, and comment on its accuracy;
M_1 has 2 parameters and M_2 has 1 so basic BIC is

$$\Delta\text{BIC} = \text{BIC}_1 - \text{BIC}_2 = (D - \log n)/2.$$

With correction factor it will be

$$2\Delta\text{BIC}^* = D - \log(2\pi n) - \log\frac{n}{x(n-x)} - \log\frac{n}{y(n-y)} + \log\frac{2n}{(x+y)(2n-x-y)}.$$

(d) Find the AIC for the two models

$$2\Delta\text{AIC} = D + 2$$

(e) Compare the model determination procedures using the three criteria for large values of n.
If M_2 holds, the deviance D will be approximately $\chi^2(1)$ and the BIC (as well as the Bayes factor) will favour M_2. Otherwise D will grow at a rate of n and BIC will favour M_1.

2. Consider regression data $(x, y) = ((x_1, y_1), \ldots, (x_n, y_n))$ with x considered fixed and the responses Y_i being independent with

$$Y_i \sim \mathcal{N}\{\mu_k(x_i), \phi\},$$

where μ_k is determined by model M_k as

$$M_1 : \mu_1(x_i) = \alpha; \quad M_2 : \mu_2(x_i) = \beta x_i; \quad M_3 : \mu_3(x_i) = \gamma x_i^2.$$

(a) For (improper) prior distributions $\pi_i(\eta, \phi) \propto \phi^{-1}$, where either $\eta = \alpha$, $\eta = \beta$, or $\eta = \gamma$, calculate expressions for the Bayes factor for comparing any pair of these models;
Write $\mu(x_i) = \theta_j z_{ij}$ where z_{ij} is either 1, x_i, or x_i^2. Next, partition the sum of squares as

$$\sum(y_i - \theta_j z_{ij})^2 = \sum(y_i - \hat{\theta}_j z_{ij})^2 + (\hat{\theta}_j - \theta)^2 \sum_i z_{ij}^2 = s_j + n_j^*(\hat{\theta}_j - \theta)^2.$$

where $\hat{\theta}_j = (\sum y_i z_{ij})/(\sum z_{ij}^2)$ and $n_j^* = \sum_i z_{ij}^2$. This yields the posterior density as

$$\pi_j(\theta_j, \phi) \propto \phi^{-(n+1)/2} \exp\{-\frac{s_j}{2\phi} - \frac{n_j^*(\theta_j - \hat{\theta}_j)^2}{2\phi}\}.$$

Integrating yields

$$B_{jl} \propto \sqrt{\frac{n_j^* s_j^{n-1}}{n_l^* s_l^{n-1}}}$$

and the Bayes factor favours the model with a small residual error, but corrects for the variation in the explanatory variable. The controversial issue is that it is generally not clear that different θ_j are meaningfully considered to be on the same scale.

(b) Find expressions for the BIC approximation to these models;
Since the MLE for (θ_j, ϕ) is $(\hat{\theta}_j, s_j/n)$ the maximizes log-likelihood is

$$l(\hat{\theta}_j, s_j/n) = \frac{n}{2} \log s_j + \text{constant}$$

so the BIC is equivalent to

$$\text{BIC} = \frac{n}{2} \log s_j + \log n$$

Since the models all have the same number of parameters, the BIC simply favours the model with the smallest residual error.

(c) Find expressions for the AIC for these models;

$$\text{AIC} = \frac{n}{2} \log s_j + 1$$

same happens here

(d) Find expressions for Mallows' C_p.
Mallows C_p is per definition

$$C_p = s_j + 2(1 - n)\sigma^2$$

and thus ranks the models by their residual sum of squares.

(e) Compare the model determination procedures.
Has been done more or less above. The only critical issue is the Bayesian correction using n_j^*. If we from the beginning normalize z such as to have $n_j^* = 1$ for all j, this factor does not enter. If this is not done, the "uniform" prior on θ means different things in the three cases and this is very problematic. Also uniform distributions can have different scales. .