Exercise 1

For Gaussian models,

a). AIC statistic: $AIC = -2\ln f(y|\hat{\theta}_k) + 2k$

Mallow's $C_p$: $C_p = \frac{RSS}{\hat{\sigma}_0^2} + 2d - n$ $\quad \hat{\sigma}_0$: estimator of $\sigma_0^2$

$$AIC = -2\ln f(y|\hat{\theta}_k) + 2k$$

$$= n\ln\hat{\sigma}^2 + n(\ln 2\pi + 1) + 2(d+1), \quad \hat{\sigma} = \frac{RSS}{n}: \text{MLE of } \sigma^2$$

minimize AIC: $n\ln\hat{\sigma}^2 - n\ln\hat{\sigma}_0^2 + 2d = n\ln\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right) + 2d$

$$n\ln\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right) \approx n\ln(1) + \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} - 1\right)n$$

$$= \frac{n\hat{\sigma}^2}{\hat{\sigma}_0^2} - n$$

$$= \frac{RSS}{\hat{\sigma}_0^2} - n.$$

$$AIC = n\ln\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right) + 2d \approx \left[\frac{RSS}{\hat{\sigma}_0^2} - n\right] + 2d = C_p$$

b). $C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2)$ $\quad$ want $BIC = \frac{1}{n}(RSS + \ln(n)\, d\hat{\sigma}^2)$

$$BIC = -2\ln f(y|\hat{\theta}_k) + k\cdot\ln n.$$

$$= n + n\ln 2\pi + n\ln\hat{\sigma} + \ln n(d+1)$$

minimize BIC: $n\ln\hat{\sigma}^2 - n\ln\hat{\sigma}_0^2 + \ln n = n\ln\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right) + \ln n \cdot d$

$$n\ln\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right) = n\ln(1) + \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} - 1\right)n$$

$$= \frac{n\hat{\sigma}^2}{\hat{\sigma}_0^2} - n$$

$$= \frac{RSS}{\hat{\sigma}_0^2} - n$$

$$BIC = n\ln\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right) + \ln(n)d = \left[\frac{RSS}{\hat{\sigma}_0^2} - n\right] + \ln(n)\, d$$

Since $C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}_0^2)$

$$= \left[\frac{RSS}{\hat{\sigma}_0^2} - n\right] + 2d,$$

$$BIC = \frac{1}{n}(RSS + \ln(n)\, d\,\hat{\sigma}_0^2)$$

This representation provides a more general applicability, because it is easy to calculate RSS and $\hat{\sigma}_0$ for a Gaussian model instead of $\hat{\sigma}$