Exercises for MA 413 – Statistics for Data Science

This sheet will be handed out on the lecture 23/11/2020, and is quite long- it is more than a week's worth of problems.

1. In a genetics experiment, a sample of size n is found to contain genotypes GG, Gg and gg each with frequency n_1, n_2 and n_3 respectively, where $n_1 + n_2 + n_3 = n$. The denotation GG corresponds to an individual having obtained two G's, where the event of obtaining each G is an independent Bernoulli trial. The population frequency of G is $\theta/(\theta+1)$. Assuming the population size is large and form the likelihood of this experiment, i.e. calculate the probability of observing n_1 of genotype GG, n_2 of Gg (equivalent to gG) and n_3 of gg, and denote the likelihood $L(\theta)$. Find the maximum likelihood estimate of θ, by differentiating the $\log(L(\theta))$.

2. A political researcher believes that the fraction p_1 of republicans in favour of the death penalty is greater than the fraction p_2 of Democrats in favour of the death penalty. He acquired independent random samples of 200 republicans and 200 Democrats and found 46 Republicans and 34 Democrats favouring the death penalty.

   (a) Estimate p_1 and p_2 from the data using maximum likelihood.

   (b) Denote the variances of any observation from either of the two populations by σ_1^2 and σ_2^2, respectively. Estimate the variances by writing σ_i^2 as a function of p_i and plug in $\widehat{p}_i$. Can you use the invariance property of maximum likelihood estimators to justify this procedure?

   (c) Write down the null and alternative hypothesis corresponding to a test of the researcher's belief.

   (d) Determine a pivotal quantity for $p_1 - p_2$. Base this on $\widehat{p}_1 - \widehat{p}_2$ and use the central limit theorem to approximate the distribution, treating the variances σ_1^2 and σ_2^2 as if they were known.

   (e) Treat your estimates of the variances of the two populations, *i.e.* $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$, as if they corresponded to the true unknown variances and test if the evidence provide statistical support for the researcher's belief. Use $\alpha = 0.05$.

3. The closing prices of two common stocks were recorded for a period of 15 days. The observed sample means and variances were $\overline{y}_1 = 40.33$, $\overline{y}_2 = 42.54$, $s_1^2 = 1.54$ and $s_2^2 = 1.69$.

   (a) Do these data present sufficient evidence to indicate a difference in means of closing prices of the two stocks for the populations associated with the two samples? Use $\alpha = 0.02$.

   (b) If you instead considered whether there was sufficient evidence to indicate stock 1 has a smaller mean than stock 2, would your deductions change? Use the same value of $\alpha = 0.02$.

   State all the assumptions you make regarding the distribution of the observations carefully, and test any assumed equivalences of parameters before using them, if you wish to assume any parameters of the distributions are equal.

4. A vice president in charge of sales for a large corporation claims that salesmen are averaging no more than 15 sales contacts per week. As a check on her claim $n = 36$ salesmen are selected at random and the number of contacts is recorded for a single randomly selected week.

   (a) The sample mean is 17 and the sample variance is 9. Does the evidence contradicts the vice president's claim? Use $\alpha = 0.05$.

   (b) The vice-president wants to be able to detect a difference equal to one call in the mean number of customer calls per week. That is, she is interested in testing $H_0 : \mu = 15$ versus the alternative of $H_1 : \mu = 16$. Describe the simple test this hypothesis corresponds to and calculate the value of β, the probability of making a type II error, for the given value of $n = 36$. What value of n is necessary to take to obtain $\beta < 0.01$?

5. As in class assume that you have a random sample of size n from a Binomial random variable, with parameter (m, θ). Assume that the prior on Θ is

$$p_\Theta(\theta) = Beta(\alpha, \beta)$$

   where α and β are known.

(a) Find the likelihood (see notes).

(b) Find the posterior distribution $p_{\Theta|\underline{X}}(\theta|\underline{x})$.

(c) Find the posterior mean.

(d) Find the posterior mode.

(e) Find the posterior median.

(f) Construct a two sided 95% credible interval for $\Theta|\underline{X}$, denoting the percentiles appropriately.

6. Assume that you have a random sample of size n from a Exponential random variable, with parameter θ. Assume that the prior on Θ is

$$p_{\Theta}(\theta) = Gamma(\alpha, \beta)$$

where α and β are known.

(a) Find the likelihood (see notes).

(b) Find the posterior distribution $p_{\Theta|\underline{X}}(\theta|\underline{x})$.

(c) Find the posterior mean.

(d) Find the posterior mode.

(e) Find the posterior median.

(f) Construct a two sided 95% credible interval for $\Theta|\underline{X}$, denoting the percentiles appropriately.

7. In a simple linear regression model satisfying SOA we have that

$$E_{Y_i|\beta}\left(Y_i|\beta\right) = \beta_0 + \beta_1 x_i$$

Assuming that at least 2 of the x_i are distinct, find $Var(\hat{Y}\left(x\right))$, where as usual

$$\hat{Y}\left(x\right) = \hat{\beta}_0 + \hat{\beta}_1 x$$

is the estimated response at an arbitrary point x. Find the value, x^* of x for which this variance is a minimum and find $Var(\hat{Y}\left(x^*\right))$.

8. A full-rank linear model

$$E_{Y|\beta}\left(\mathbf{Y}|\beta\right) = \mathbf{X}\beta$$

with SOA is fitted and the least square estimate $\hat{\beta}$ is found. The same model, but with an extra term added, is fitted to the data: i.e. the new model is

$$E_{Y|\beta,\gamma}\left(\mathbf{Y}|\beta,\gamma\right) = \mathbf{X}\beta + \gamma\underline{x}$$

where $\underline{x}$ is a known vector linearly independent of the columns of $\mathbf{X}$ and γ is a new unknown parameter. The least square estimate of β in this new model is $\hat{\beta}_N$, and show that

$$\hat{\beta}_N = \hat{\beta} - \hat{\gamma}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\underline{x}$$

where

$$\hat{\gamma} = \frac{\left(\underline{x}^T\mathbf{A}\mathbf{Y}\right)}{\left(\underline{x}^T\mathbf{A}\underline{x}\right)}$$

and

$$\mathbf{A} = \mathbf{I} - \mathbf{P}.$$

9. Let $\mathbf{X}$ be an $n \times p$ matrix of rank $p < n$. Show by direct multiplication that for any $p \times 1$ vector $\underline{x}$

$$\left(\mathbf{X}^T\mathbf{X} + \underline{x}\underline{x}^T\right)^{-1} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\mathbf{I}_p - \frac{\underline{x}\underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}}{1 + \underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}\right).$$

A new observation Y_{n+1} is added to the full rank model of

$$E_{Y|\beta}(\mathbf{Y}|\beta) = \mathbf{X}\beta$$

If

$$E_{Y_{n+1}|\beta}(Y_{n+1}|\beta) = \underline{x}^T\beta$$

and if SOA apply, show that the least squares equations for the model involving all of the observations are

$$\left(\mathbf{X}^T\mathbf{X} + \underline{x}\underline{x}^T\right)\tilde{\beta} = \mathbf{X}^T\mathbf{Y} + Y_{n+1}\underline{x}$$

Show that then

$$\tilde{\beta} = \hat{\beta} + \frac{\left(Y_{n+1} - \underline{x}^T\hat{\beta}\right)\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}{1 + \underline{x}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{x}}$$

where $\tilde{\beta}$ and $\hat{\beta}$ are the least squares estimates β, from the augmented and first given models respectively. We would you *expect* $\tilde{\beta} = \hat{\beta}$ when $Y_{n+1} = \underline{x}^T\hat{\beta}$?

10. The independent rv $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ are normally distributed and both have unknown variance $\sigma^2 > 0$. In addition we assume that

$$E_{Y_i|\beta}(Y_i|\beta) = \beta_0 + \beta_1 x_i$$

and

$$E_{Z_i|\beta}(Z_i|\beta) = \alpha_0 + \alpha_1 x_i$$

where α_i, β_i are unknown but x_i are known. Show how to write this as a single linear model for the $2n$ observations $Y_1, \ldots, Y_n, Z_1, \ldots, Z_n$. Assume that the model is of full rank and construct a $1-\gamma$ confidence intervals for $\beta_1 - \alpha_1$, where $\gamma \in (0, 1)$.

11. Find $\mathbf{P}$ when $E(Y_i) = \mu$.

12. Find $\mathbf{P}$ when $E(Y_i) = \mu + \beta x_i^2$

13. Find the MLE of the parameters in the model

$$E_{Y|\beta}(Y|\beta) = \beta_0 + \beta_1\sqrt{x_i}$$

when

$$Var(Y_i) = \begin{cases} \sigma^2 & \text{if} \quad |x_i| < c \\ 2\sigma^2 & \text{if} \quad |x_i| \geq c \end{cases},$$

where $|x_i| \geq c \forall i \geq n/2 + 1$, and n is even.

14. (Hard) Say as in class we let

$$\mathbf{W} = \mathcal{W}\mathbf{Y} = \mathcal{W}\mathbf{f} + \mathcal{W}\eta$$

we let

$$\mathbf{U} = \mathcal{W}\mathbf{f}$$

and

$$\epsilon = \mathcal{W}\eta$$

Consider now implementing this in a *Bayesian* framework.

(a) Assume that $\epsilon_j \sim N\left(0, \sigma^2\right)$. Write down $f_{W_j|U_j}(w|u)$.

(b) Assume that we let

$$U_j \sim \pi_j N(0, \tau_j) + (1 - \pi_j)\,\delta(u),$$

i.e. a mixture of a discrete and continuous probability density, where U_j with probability $1-\pi_j$ takes the value 0, and otherwise is normally distributed, where the normal pdf is adjusted so that the total probability equals 1. Show that

$$F_{U_j|W_j}(u|w) = \frac{1}{1+\xi_j}\Phi\left(\frac{u - w\tau_j^2/\Omega_j^2}{\sigma\tau_j/\Omega_j}\right) + \frac{\xi_j}{1+\xi_j}I(w \geq 0)$$

where

$$\xi_j = \frac{1-\pi_j}{\pi_j}\frac{\Omega_j}{\sigma}\exp\left(-\frac{\tau_j^2 w_j^2}{2\sigma^2\Omega_j^2}\right)$$

and

$$\Omega_j^2 = \tau_j^2 + \sigma^2.$$

(Hint: Consider separately the two cases $U_j = 0$ and $U_j \neq 0$.

(c) Find the posterior mean, and compare with the thresholding rule given in class.

(d) As an alternative, consider testing the hypothesis

$$H_0 : U_j = 0$$

versus

$$H_1 : U_j \neq 0$$

If the null hypothesis is rejected we shall estimate U_j by w_j otherwise we set the coefficient to 0. This corresponds to

$$\hat{U}_j = \begin{cases} W_j & \text{if} \quad \chi_j \geq 1 \\ 0 & \text{if} \quad \chi_j < 1 \end{cases}$$

where

$$\chi_j = \frac{P\left(U_j = 0 | W_j = w_j\right)}{P\left(U_j \neq 0 | W_j = w_j\right)}.$$

Find χ_j and comment on its value, in terms of the parameters given.

15. Say as in class we let

$$\mathbf{W} = \mathcal{W}\mathbf{Y} = \mathcal{W}\mathbf{f} + \mathcal{W}\eta$$

we let

$$\mathbf{U} = \mathcal{W}\mathbf{f}$$

and

$$\epsilon = \mathcal{W}\eta$$

Assume that $\epsilon_j \sim N\left(0, \sigma^2\right)$, as in the previous question.

Assume for a given λ we let

$$f_{U_j}(u) = \frac{\lambda}{2\sigma^2}\exp\left(-\frac{\lambda}{\sigma^2}|u|\right),$$

where λ is a fixed constant. Show that the posterior is maximised when

$$\frac{1}{2}(w-u)^2 + \lambda|u|$$

is minimised. Use the fact that w and u will have the same sign, i.e $(u-w)^2 = (|w| - |u|)^2$ to show that the maximum posterior value of u is given by

$$u^{(st)} = \begin{cases} w - \lambda & \text{for} \quad w > \lambda \\ 0 & \text{for} \quad |w| \leq \lambda \\ -(|w| - \lambda) & \text{for} \quad w < -\lambda \end{cases}$$