

ASSIGNMENT SHEET 13

December 20, 2017

Assignment 1. Let y_1, \dots, y_n be independent Bernoulli random variables such that $\pi_j = \mathbb{P}(y_j = 1) = \exp(x_j^T \beta) / \{1 + \exp(x_j^T \beta)\}$.

(i). Let $\hat{\pi}_j = \exp(x_j^T \hat{\beta}) / \{1 + \exp(x_j^T \hat{\beta})\}$. Show that the likelihood equation is $X^T y = X^T \hat{\pi}$

(ii). Show that the deviance is

$$D = -2 \left\{ y^T X \hat{\beta} + \sum_{j=1}^n \log(1 - \hat{\pi}_j) \right\}.$$

(iii). Show that the deviance is only a function of $\hat{\pi}_j$.

Assignment 2 (independence test). Consider the 2x2 contingency setup of slide 504.

(i). Write down the likelihood and log likelihood functions.

(ii). Find the maximum likelihood estimators of λ and ψ . (Since this is the end of the semester, we do not ask you to verify that your solution is a maximum; do not expect this sort of kindness in the exam.) *Hint: there is a one-to-one correspondence with π_0 and π_1 and it may be more convenient to work with the π 's.*

(iii). Express the equality $\pi_0 = \pi_1$ in terms of λ and ψ .

(iv). Find the maximum likelihood estimators for the constrained model where $\pi_0 = \pi_1$.

(v). Find the deviance difference between the two models.

(vi). Use the limit distribution result of slide 478 to test the hypothesis $\pi_0 = \pi_1$ in the following case ($n = 1024$)

	success	failure	total
Case	$R_1 = 266$	$m_1 - R_1 = 231$	$m_1 = 497$
Control	$R_0 = 243$	$m_0 - R_0 = 284$	$m_0 = 527$
Total	509	515	1024

Assignment 3. By writing $\sum \{y_j - \hat{g}(t_j)\}^2 = (y - \hat{g})^T (y - \hat{g})$, with $y = g + \epsilon$ and $\hat{g} = Sy$, where S is a smoothing matrix, show that

$$\mathbb{E} \left[\sum_{j=1}^n \{y_j - \hat{g}(t_j)\}^2 \right] = \sigma^2 (n - 2\nu_1 + \nu_2) + g^T (I - S)^T (I - S) g,$$

where $\nu_1 = \text{tr}(S)$, $\nu_2 = \text{tr}(S^T S)$.

Hence explain the use of

$$s^2 = \frac{1}{n - 2\nu_1 + \nu_2} \sum_{j=1}^n \{y_j - \hat{g}(t_j)\}^2$$

as an estimator of σ^2 . Under what circumstances is it unbiased?

Assignment 4.

Show that the contribution to the scaled deviance for a response variable with Binomial density $\binom{m}{r} \pi^r (1 - \pi)^{m-r}$, $0 < \pi < 1$, $r = 0, \dots, m$, $\eta = \pi$ is

$$2\{r \log\{r/(m\hat{\pi})\} + (m - r) \log[(m - r)/\{m(1 - \hat{\pi})\}]\}.$$

Assignment 5.

Show that the contribution to the scaled deviance for a response variable with Poisson density $\eta^y e^{-\eta}/y!$, $\eta > 0$, $y = 0, 1, \dots$, is $2\{y \log(y/\hat{\eta}) - y + \hat{\eta}\}$.

Assignment 6. (Natural cubic splines)

Let $n \geq 2$ and $a < x_1 < x_2 < \dots < x_n < b$. Denote by $N(x_1, x_2, \dots, x_n)$ the space of natural cubic splines with knots x_1, x_2, \dots, x_n . The goal of this exercise is to show that the solution to the problem

$$\min_{f \in C^2[a,b]} L(f), \text{ où } L(f) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_a^b \{f''(x)\}^2 dx, \quad \lambda > 0, \quad (1)$$

must belong to $N(x_1, x_2, \dots, x_n)$. In order to show this, we need the following theorem

Theorem. For every set of points $(x_1, z_1), (x_2, z_2), \dots, (x_n, z_n)$, it exists a natural cubic spline g interpolating those points. In other words, $g(x_i) = z_i$, $i = 1, \dots, n$, for a unique natural cubic spline g . Moreover, the knots of g are x_1, x_2, \dots, x_n .

- (i). Let g the natural cubic spline interpolating the points (x_i, z_i) , $i = 1, \dots, n$, and let $\tilde{g} \in C^2[a, b]$ another function interpolating the same points. Show that

$$\int_a^b g''(x)h''(x)dx = 0,$$

where $h = \tilde{g} - g$. *Hint : integration by parts*

- (ii). Using point (1) show that

$$\int_a^b \{\tilde{g}''(x)\}^2 dx \geq \int_a^b \{g''(x)\}^2 dx$$

when the equality holds if and only if $\tilde{g} = g$.

- (iii). Use point (2) to show that if the problem (1) has a solution \hat{f} , then $\hat{f} \in N(x_1, x_2, \dots, x_n)$.