# Assignment sheet 13

Spring 2025

**Assignment 1.**

(i). Show that the binomial density

$$f(y;\pi) = \binom{m}{y}\pi^y(1-\pi)^{m-y}, \quad 0 < \pi < 1, \quad y = 0, \dots, m.$$

may be written as

$$\exp\left[y\phi + \gamma(\phi) + S(y)\right]$$

and express $\phi$, $\gamma$ and $S(y)$ in terms of the usual parameter $\pi$.

(ii). Deduce the mean and variance function for $Y$.

**Assignment 2.** If $X$ is a Poisson variable with mean $\mu = \exp(x^T\beta)$ and $Y$ is a binary variable indicating the event $X > 0$, find the link function between $\mathbb{E}(Y)$ and $x^T\beta$.

**Assignment 3.** 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 $\widehat{\pi}_j = \exp(x_j^T\widehat{\beta})/\{1 + \exp(x_j^T\widehat{\beta}\}$. Show that the likelihood equation is $X^Ty = X^T\widehat{\pi}$

(ii). Show that the deviance is

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

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

**Assignment 4.**

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 5.** 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 $\nu_1 = tr(S)$, $\nu_2 = tr(S^TS)$.
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 $\sigma^2$. Under what circumstances is it unbiased?

**Assignment 6.** (Natural cubic splines)
Let $n \geq 2$ and $a < x_1 < x_2 < \cdots < x_n < b$. Denote by $N(x_1, x_2, \ldots, x_n)$ the space of natural cubic splines with knots $x_1, x_2, \ldots, x_n$. The goal of this exercise is to show that the solution to the problem

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

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

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

(i). Let $g$ the natural cubic spline interpolating the points $(x_i, z_i)$, $i = 1, \ldots, n$, 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$\hat{f} \in N(x_1, x_2, \ldots, x_n)$.