Answer sheet 13

Assignment 1.

(i). We have

$$\begin{aligned} f(y;\pi) &= \binom{m}{y}\pi^y(1-\pi)^{m-y} \\ &= \exp\left[\log\binom{m}{y} + y\log\left(\frac{\pi}{1-\pi}\right) + m\log(1-\pi)\right] \\ &= \exp\left[y\phi + \gamma(\phi) + S(y)\right] \end{aligned}$$

with

$$\phi = \log\left(\frac{\pi}{1-\pi}\right) \quad \Leftrightarrow \quad \pi = \frac{e^\phi}{1+e^\phi},$$

$$\gamma(\phi) = -m\log(1-\pi) = -m\log\left(\frac{1}{1+e^\phi}\right) = m\log(1+e^\phi),$$

$$S(y) = \log\binom{m}{y}.$$

(ii). From the course we have that

$$\mathbb{E}(Y) = \mu = \gamma'(\phi) = m\frac{e^\phi}{1+e^\phi} = m\pi$$

$$\mathrm{Var}(Y) = \gamma''(\phi) = \frac{me^\phi}{(1+e^\phi)^2} = \frac{me^\phi}{1+e^\phi}\frac{1}{1+e^\phi} = \mu\left(1-\frac{\mu}{m}\right).$$

Assignment 2.

$$\mathbb{E}(Y) = P(X>0) = 1 - P(X=0) = 1-\exp(-\mu) = 1-\exp\{-\exp(x^T\beta)\}.$$

Assignment 3. (i). Let $\eta_j = x_j^T\beta$. The log likelihood function as a function of (η_j) is

$$\ell_\eta(\eta) = \sum_{j=1}^n y_j \log\frac{\exp(\eta_j)}{1+\exp(\eta_j)} + (1-y_j)\log\frac{1}{1+\exp(\eta_j)} = \sum_{j=1}^n y_j\eta_j - \log(1+\exp(\eta_j))$$

and as a function of β

$$\ell(\beta) = \sum_{j=1}^n y_j x_j^T\beta - \log[1+\exp(x_j^T\beta)].$$

To obtain the likelihood equation we equate to zero the derivative of ℓ with respect to β :

$$\frac{\partial\ell(\beta)}{\partial\beta_i} = \sum_{j=1}^n y_j X_{ji} - \pi_j X_{ji} = (y_j - \pi_j)X_{ji}.$$

The likelihood equation says that this should equal 0 for all i, which in matrix form can be written $y^T X = \widehat{\pi}^T X$.

(ii). To calculate the deviance we need to maximise with respect to (η_j) and with respect to β and compare the optimal objective value. Notice that ℓ_η is decreasing in η_j if $y_j = 0$ and increasing if $y_j = 1$. Therefore the supremum is "attained" when $\eta_j = -\infty$ if $y_j = 0$ and $\eta_j = \infty$ if $y_j = 1$ with objective value zero.
The optimal value of $\ell(\beta)$ is

$$\ell(\beta) = \sum_{j=1}^{n} y_j x_j^T \widehat{\beta} - \log[1 + \exp(x_j^T \widehat{\beta})] = y^T X \widehat{\beta} + \sum_{j=1}^{n} \log(1 - \widehat{\pi}_j).$$

The deviance is twice the negative of this expression, since the optimal value in the saturated model was shown to vanish.

(iii). If we plug in $\eta_j = \exp(x_j^T \beta)$ in the first expression of ℓ_η we get

$$D = -2 \sum_{j=1}^{n} y_j \log \widehat{\pi}_j + (1 - y_j) \log(1 - \widehat{\pi}_j)$$

and this depends only on $(\widehat{\pi}_j)$.

Assignment 4. The log-likelihood for a sample of size n for the saturated model is given by

$$\ell(\hat{\pi}_{max}, y) = \ell(\eta, y) = \sum_{i=1}^{n} \{y_i \log(\eta_i) - \eta_i - \log(y_i!)\}.$$

Thus we have $\frac{\partial \ell}{\partial \eta_i} = \frac{y_i}{\eta_i} - 1$, d'où $\eta_i = y_i$. Finally

$$\begin{aligned}
D &= 2 \sum_{j=1}^{n} \{\log f(y_j; \hat{\eta}_{max}) - \log f(y_j; \hat{\eta})\} \\
&= 2 \sum_{j=1}^{n} \{y_j \log(y_j) - y_j - \log(y_j!) - y_j \log(\hat{\eta}_j) + \hat{\eta}_j + \log(y_j!)\} \\
&= 2 \sum_{j=1}^{n} \left\{ y_j \log\left(\frac{y_j}{\hat{\eta}_j}\right) - y_j + \hat{\eta}_j \right\}.
\end{aligned}$$

Assignment 5. Write

$$\begin{aligned}
(y - \hat{g})^T (y - \hat{g}) &= (g + \epsilon - Sg - S\epsilon)^T (g + \epsilon - Sg - S\epsilon) \\
&= \{(I - S)g + (I - S)\epsilon\}^T \{(I - S)g + (I - S)\epsilon\} \\
&= g^T (I - S)^T (I - S) g + 2 g^T (I - S)^T (I - S)\epsilon + \epsilon^T (I - S)^T (I - S)\epsilon.
\end{aligned}$$

The first terms is deterministic, and the second has mean zero. Thus

$$\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^{n} \{y_j - \hat{g}(t_j)\}^2\right] &= g^T (I - S)^T (I - S) g + \mathbb{E}\{\epsilon^T \epsilon\} - 2\,\mathbb{E}\{\epsilon^T S \epsilon\} + \mathbb{E}\{\epsilon^T S^T S \epsilon\} \\
&= g^T (I - S)^T (I - S) g + \sum_{i=1}^{n} \{\mathbb{E}(\epsilon_i^2) - 2 s_{ii}\, \mathbb{E}(\epsilon_i^2) + s s_{ii}\, \mathbb{E}(\epsilon_i^2)\} \quad (\text{as } \mathbb{E}(\epsilon_i \epsilon_j) = 0). \\
&= g^T (I - S)^T (I - S) g + \sigma^2 (n - 2\nu_1 + \nu_2).
\end{aligned}$$

(s_{ij}, ss_{ij} are the ij-th elements of S and $S^T S$ respectively.)

$$\mathbb{E}(s^2) = \sigma^2 + \frac{g^T(I-S)^T(I-S)g}{n - 2\nu_1 + \nu_2}$$

so s^2 can be considered an estimator of σ^2. It is unbiased if $(I-S)g = 0$; equivalently, $Sg = g$.

Assignment 6.

(i). Using integration by parts, we obtain that

$$\begin{aligned}
\int_a^b g''(x)h''(x)dx &= \underbrace{g''(x)h'(x)\Big|_a^b}_{=0,\ \text{car } g''(a)=g''(b)=0} - \int_a^b g'''(x)h'(x)dx \\
&= -\sum_{i=1}^{n-1} g'''(x_i^+) \int_{x_i}^{x_{i+1}} h'(x)dx \\
&= -\sum_{i=1}^{n-1} g'''(x_i^+)\{h(x_{i+1}) - h(x_i)\} = 0.
\end{aligned}$$

Here, the second equality comes from the fact that $g'''(x) = 0$ inside the intervals (a, x_1) and (x_n, b) and that $g'''(x)$ equals to the constant $\lim_{x\to x_i^+} g'''(x) = g'''(x_i^+)$ inside the interval (x_i, x_{i+1}). To obtain the last equality finally, observe that $\tilde{g}(x_i) = g(x_i) = z_i$ hence $h(x_i) = 0$ for every i.

(ii). By direct computation we obtain that

$$\begin{aligned}
\int_a^b \{\tilde{g}''(x)\}^2 dx &= \int_a^b \{g''(x) + h''(x)\}^2 dx \\
&= \int_a^b \{g''(x)\}^2 dx + 2\int_a^b g''(x)h''(x)dx + \int_a^b \{h''(x)\}^2 dx \\
&= \int_a^b \{g''(x)\}^2 dx + \int_a^b \{h''(x)\}^2 dx \geq \int_a^b \{g''(x)\}^2 dx.
\end{aligned}$$

where we have equality if and only if $h''(x) \equiv 0$, so we must have $h(x) = kx + c$. But since $h(x_i) = 0$ for every i, it must be that$h(x) \equiv 0$. In particular we have equality if and only if $\tilde{g} = g$.

(iii). Let $\tilde{f} \in C^2[a,b]\backslash N(x_1, \ldots, x_n)$ and let $f \in N(x_1, \ldots, x_n)$ the spline which is interpolating the points $(x_i, \tilde{f}(x_i))$,
$i = 1, \ldots, n$. The existance of f is guaranteed by the theorems seen in class. By point (2)

$$\int_a^b \{\tilde{f}''(x)\}^2 dx > \int_a^b \{f''(x)\}^2.$$

Moreover

$$\sum_{i=1}^n (y_i - \tilde{f}(x_i))^2 = \sum_{i=1}^n (y_i - f(x_i))^2.$$

Hence, $L(\tilde{f}) > L(f)$ and we notice that if the minimum exists, it must belong to $N(x_1, \ldots, x_n)$.

Remark. Using the properties of splines, it it possible to show that a minimum always exists and is unique. Hence the problem $\min_{f \in C^2[a,b]} L(f)$ admits always a unique solution and this solution is a natural cubic spline.