STATISTICAL MODELLING

Example Sheet 1

(This is meant to be a very easy sheet, to get you started.)

1. Suppose $X_1, \ldots, X_n$ are i.i.d. Poisson random variables with parameter μ.
   Show that $\hat{\mu} = \Sigma X_i/n$, and $\mathrm{var}\,(\hat{\mu}) = \frac{1}{n}\mu$.

   What is $\mathbb{E}\left(-\frac{\partial^2 L}{\partial \mu^2}\right)$? What is the exact distribution of $(n\hat{\mu})$? What is the asymptotic distribution of $\hat{\mu}$?

2. Suppose we have n i.i.d. trials, and the outcome of each trial is

$$\left.\begin{array}{rl} & \text{Red with probability } \theta_1, \\ & \text{White with probability } \theta_2, \\ \text{or} & \text{Blue with probability } \theta_3, \end{array}\right\} \quad \text{thus } \theta_1 + \theta_2 + \theta_3 = 1.$$

   Let (X, Y, Z) be the total number of (Red, White, Blue) trials in the sequence of n; write $X = \sum_1^n X_i$, $Y = \sum_1^n Y_i$ for suitably defined (X_i, Y_i).

   Find $\mathbb{E}(X)$, $\mathrm{var}\,(X)$, and show that

$$\mathrm{cov}\,(X, Y) = -n\theta_1\theta_2.$$

   Find $\binom{\hat{\theta}_1}{\hat{\theta}_2}$, and find the mean vector, and covariance matrix, of its asymptotic distribution (which is of course bivariate normal).

3. Suppose Y_i independent Poisson, mean μ_i, and our model is

$$H : \log(\mu_i) = \alpha + \beta x_i$$

   where (x_i) are given.

   Write down the log likelihood $\log f(\mathbf{y}|\alpha, \beta)$ and hence find

   (i) the sufficient statistics for (α, β);
   (ii) equations for $(\hat{\alpha}, \hat{\beta})$, the mle, and
   (iii) an expression for

$$\max_{\beta=0} f(\mathbf{y}|\alpha, \beta).$$

   Show how you would use this, together with Wilks' theorem, to test $H_0 : \beta = 0$.

4. Suppose $f(y_i|\beta) = \frac{1}{2}\exp -|y_i - \beta|$, $-\infty < y_i < \infty$. Let $L(\beta) = \sum_1^n \log f(y_i|\beta)$. Sketch this as a function of β, and hence find an expression for $\hat{\beta}$, the mle, given that the ordered observations are $y_{(1)} < \ldots < y_{(n)}$.

   What is $\frac{\partial L}{\partial \beta}$?

5. Suppose $f(y_i|\theta) = \theta e^{-\theta y_i}$, $y_i > 0$, $1 \le i \le n$.

   Show that this is an exponential family form distribution, with natural parameter $\pi = -\theta$. Find the sufficient statistic and its distribution, and find the mle for each of π, θ.

6. Suppose $f(y_i|\alpha, \beta)$ is the pdf of $N(\alpha + \beta x_i, \sigma^2)$, where σ^2 is known, and $x_1, \ldots, x_n$ are known.

   Show that the loglikelihood function $L_n(\alpha, \beta)$ is a concave function of (α, β).

7. Suppose $f(y_i|\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp -y_i^2/2\theta$, for $i = 1, \ldots, n$. Show that the mle is $\hat{\theta} = \frac{1}{n}\sum_1^n y_i^2$, and that $\frac{n\hat{\theta}}{\theta} \sim \chi_n^2$.

   Find the exact mean and variance of $\hat{\theta}$, and compare these with the asymptotic mean and variance obtained from general likelihood theory.

8. Suppose $Y_i \sim Bi(1, \pi_i)$ independent, $(i = 1, \ldots, n)$ (i.e. $P(Y_i = 1) = \pi_i$, $P(Y_i = 0) = 1 - \pi_i$). Suppose also $\log(\pi_i/(1 - \pi_i)) = \alpha + \beta x_i$, where $x_1, \ldots, x_n$ are given.

   Write down $L_n(\alpha, \beta)$, the loglikelihood function, and find equations for $(\hat{\alpha}, \hat{\beta})$ the mle of (α, β).

9. Suppose $Y_i \sim NID(\beta_1 + \beta_2 x_i + \beta_3 P_2(x_i), \sigma^2)$, $1 \le i \le n$, where $\Sigma x_i = 0$ and $P_2(x_i)$ is a given quadratic function of x_i such that

   $$\Sigma P_2(x_i) = \Sigma x_i P_2(x_i) = 0.$$

   Find

   $$\frac{\partial \ell}{\partial \beta}, \qquad \frac{\partial^2 \ell}{\partial \beta \partial \beta^T} \qquad \text{and} \qquad \mathbb{E}\left(\frac{\partial^2 \ell}{\partial \beta \partial \beta^T}\right).$$

   Find $\hat{\beta}_3$ and var $(\hat{\beta}_3)$, and show how you would test $H_0 : \beta_3 = 0$ if σ^2 is known. If $P_2(x_i) = x_i^2 + a x_i + b$, find expressions for a, b in terms of $\mathbf{x}$.

   Can you see any advantages in fitting the model with

   $$\mathbb{E}(Y_i) = \beta_1 + \beta_2 x_i + \beta_3 P_2(x_i)$$

   as above rather than the model written as

   $$\mathbb{E}(Y_i) = \beta_1 + \beta_2 x_i + \gamma x_i^2$$

   say?

   [The purpose of this question is to introduce you to ORTHOGONAL polynomials.]

10. Suppose $Y_i \sim Bi(n_i, p_i)$, $1 \le i \le k$, independent, and $g(p_i) = \alpha + \beta x_i$ where $g(\cdot)$ is a known link function. Write down

    $$\frac{\partial \ell}{\partial \alpha}, \quad \frac{\partial \ell}{\partial \beta}.$$

    What are the sufficient statistics for (α, β)

    (i) if $g(p) = \log(p/(1-p))$? (link=logit)
    (ii) if $g(p) = \log(-\log(1-p))$? (link= cloglog)?