STATISTICAL MODELLING Part IIC.

Example Sheet 1 (of 4) RN/Lent 2010

1. Let Y be a random variable with density $f(y;\theta)$ for $y \in \mathcal{Y} \subseteq \mathbb{R}^n$ and some $\theta \in \Theta \subseteq \mathbb{R}^d$, and write $\ell(\theta; Y)$ and $U(\theta; Y)$ for the corresponding log-likelihood and score functions. Assume that the order of differentiation with respect to a component of θ and integration over $\mathcal{Y}$ may be interchanged where necessary. Show that, for $r, s = 1, \ldots, d$,
$$\mathrm{Cov}_\theta\{U_r(\theta;Y), U_s(\theta;Y)\} = -\mathbb{E}_\theta\Big\{\frac{\partial^2}{\partial\theta_r\partial\theta_s}\ell(\theta;Y)\Big\}.$$

2. Let $Y_1, \ldots, Y_n$ be independent Poisson random variables with mean θ. Compute the maximum likelihood estimator $\hat{\theta}_n$. By considering $n\hat{\theta}_n$, write down the distribution of $\hat{\theta}_n$ and deduce its asymptotic distribution directly. Verify that this asymptotic distribution agrees with that predicted by the general asymptotic theory for maximum likelihood estimators.

3. Let $Y_1, \ldots, Y_n$ be independent Poisson(θ) random variables. Show that both $\bar{Y} = n^{-1}\sum Y_i$ and $S^2 = (n-1)^{-1}\sum(Y_i - \bar{Y})^2$ are unbiased estimators of θ. Without calculating $\mathrm{Var}_\theta(S^2)$, argue that $\bar{Y}$ is at least as good an estimator as S^2.

4. Let $Y_1, \ldots, Y_n$ be independent $U[0, \theta]$ random variables, for some $\theta \in \Theta = (0, \infty)$. Find the maximum likelihood estimator $\hat{\theta}_n$, as well as its distribution function, mean and variance. What is the asymptotic distribution of $n(\theta - \hat{\theta}_n)/\theta$? Why does the standard theory not apply?

5. Consider the standard linear model $Y = X\beta + \epsilon$, where X is an $n \times p$ matrix of full rank p. Find the distribution of maximum likelihood estimator $\hat{\beta}$ of β. Calculate $i(\beta)$ and argue that the asymptotic distribution of the m.l.e. is exact in this case.

6. Bayesian Inference.

   (a) Find the posterior distribution (up to a normalising constant) for the parameters β, σ^2 under the standard linear model in Question 6. Use the Jeffreys' prior $p(\beta, \sigma^2) \propto \sigma^{-2}$.

   (b) Derive the posterior conditionals $p(\beta|\sigma^2, X, Y)$ and $p(\sigma^2|\beta, X, Y)$, and the posterior marginal $p(\sigma^2|X, Y)$.

(c) Derive the posterior predictive distribution of y^* at x^*, conditional on σ^2: $p(y^*|\sigma^2, x^*, X, Y)$.

7. Recall that in the standard linear model above we may express the fitted values $\hat{Y} = X\hat{\beta}$ as $\hat{Y} = PY$, where $P = X(X^\top X)^{-1}X^\top$.

   (a) Show that P represents an orthogonal projection.

   (b) Show that P and $I - P$ are positive semi–definite, where I is the $n \times n$ identity matrix.

   (c) Show that $I - P$ has rank $n - p$ and P has rank p.

8. In the standard linear model above, find the maximum likelihood estimator $\hat{\sigma}^2$ of σ^2, and use Cochran's theorem to find its distribution. [*Hint: use the results from the previous question.*]

9. Let $Y = X\beta + \epsilon$, where X and β are partitioned as $X = (X_0\, X_1)$ and $\beta^\top = (\beta_0^\top\, \beta_1^\top)$ respectively (where β_0 has p_0 components and β_1 has $p - p_0$ components).

   (a) Show that β_0 and β_1 are orthogonal if and only if the Fisher information matrix is block diagonal. [*This is the appropriate generalisation of parameter orthogonality to more general parametric models.*]

   (b) Use this generalisation to show that β and σ^2 are orthogonal.

10. Consider the model for responses $Y_1, \ldots, Y_n$ given by

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 P_2(x_i) + \epsilon_i,$$

where $\epsilon_1, \ldots, \epsilon_n$ are independent $N(0, \sigma^2)$ random variables, $\sum_{i=1}^n x_i = 0$, and P_2 is a monic quadratic polynomial. Find P_2 to make β_0, β_1 and β_2 mutually orthogonal. For this choice of P_2, compute the maximum likelihood estimator $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^\top$ and write down its distribution.

11. In the balanced, additive two-way ANOVA model, show that the maximum likelihood fitted values are $\tilde{Y}_{ijk} = \bar{Y}_{i++} + \bar{Y}_{+j+} - \bar{Y}$. [*Hint: use the sum-to-zero identifiability constraints.*]