Computational Statistics and Statistical Modelling
Example Sheet 1 : Outline solutions

1. The likelihood is

$$f(x|\mu) = \Pi e^{-\mu}\ \mu^{x_i}/x_i!$$

giving the loglikelihood as

$$L = log(f(x|\mu)) = -n\mu + \Sigma x_i\ log(\mu) + \text{constant}.$$

Hence

$$\frac{\partial L}{\partial \mu} = -n + \Sigma x_i/\mu$$

and

$$\frac{\partial^2 L}{\partial \mu^2} = -\Sigma\ x_i/\mu^2\ (\ < 0)$$

so that L is max'd at

$$\hat{\mu} = \Sigma x_i/n.$$

Clearly $E(X_i) = \mu = var(X_i)$. Thus $E(\hat{\mu}) = \mu$

and $var(\hat{\mu}) = \mu/n$, and

$$E - \frac{\partial^2 L}{\partial \mu^2} = n/\mu.$$

The exact distribution of $\hat{\mu}n = \Sigma X_i$ is $Po(n\mu)$.

The asymptotic distribution of $\hat{\mu}$, by the Central Limit Theorem, is $N(\mu, \mu/n)$.

2. (This is the 3-cell multinomial distribution). With

$$f(x, y, z|\theta) = n!\ [{\theta_1}^x {\theta_2}^y {\theta_3}^z/x!y!z!]$$

for $x, y, z = 0, 1, 2..., x + y + z = n$,

we have $X = \Sigma X_i$ say, where

X_i=1 if ith trial results in a Red, $X_i = 0$, otherwise, $Y = \Sigma Y_i$, and $Y_i = 1$ if ith trial results in a White, $Y_i = 0$ otherwise.

Clearly, $P(X_i = 1) = \theta_1$, and X is Bi(n, θ_1)

so that $var(X) = n\theta_1(1 - \theta_1)$, $E(X) = n\theta_1$.

Further

$$cov(X,Y) = \Sigma cov(X_i, Y_i) = n(E(X_1Y_1) - E(X_1)E(Y_1)).$$

Clearly, $E(X_1Y_1) = 0$, so that $cov(X,Y) = -n\theta_1\theta_2$.
Now L $= log f(x,y|\theta) = xlog(\theta_1) + ylog(\theta_2) + zlog(\theta_3) +$ constant
which is maximized subject to $\theta_1 + \theta_2 + \theta_3 = 1$ (use Lagrange multiplier)
by $\hat{\theta}_1 = x/n, \hat{\theta}_2 = y/n, \hat{\theta}_3 = z/n$.
Hence $E(\hat{\theta}_i) = \theta_i$ for $i = 1,2$.
Now

$$\frac{\partial L(\theta)}{\partial \theta_1} = (x/\theta_1) - (z/\theta_3)$$

$$\frac{\partial L(\theta)}{\partial \theta_2} = (y/\theta_2) - (z/\theta_3).$$

Hence, - matrix of 2nd derivatives of L is

$$\begin{pmatrix} x/\theta_1^2 + z/\theta_3^2 & z/\theta_3^2 \\ z/\theta_3^2 & y/\theta_2^2 + z/\theta_3^2 \end{pmatrix}$$

Substituting for $E(X), E(Y), E(Z)$, we see that the expectation of the above matrix is

$$\begin{pmatrix} n(1-\theta_2)/\theta_1\theta_3 & n/\theta_3 \\ n/\theta_3 & n(1-\theta_1)/\theta_2\theta_3 \end{pmatrix}$$

It now remains for you to check that the inverse of this two-by-two matrix is

$$\begin{pmatrix} \theta_1(1-\theta_1)/n & -\theta_1\theta_2/n \\ -\theta_1\theta_2/n & \theta_2(1-\theta_2)/n \end{pmatrix}$$

This is what the general formula for the *asymptotic* covariance matrix gives us. In this case, it agrees exactly with the *exact* covariance matrix.

3.i) With $f(y_i|\mu_i)$ proportional to $e^{-\mu_i}{\mu_i}^{y_i}$
and $\mu_i = e^{(\alpha+\beta x_i)}$ we see that the likelihood for (α, β) is proportional to

$$[exp - \Sigma e^{\alpha+\beta x_i}] exp\ [\alpha\ t_1 + \beta\ t_2]$$

where t_1 is defined as Σy_i , and t_2 as $\Sigma x_i y_i$.
Hence, by the factorisation theorem, (t_1, t_2) are sufficient for (α, β).
The log likelihood is

$$L(\alpha, \beta) = -\Sigma e^{\alpha+\beta x_i} + \alpha\ t_1 + \beta\ t_2 + \text{constant}.$$

$ii)$Thus

$$\frac{\partial L}{\partial \alpha} = 0 \text{ for } t_1 = \Sigma e^{\alpha+\beta x_i}$$

$$\frac{\partial L}{\partial \beta} = 0 \text{ for } t_2 = \Sigma x_i e^{\alpha+\beta x_i}.$$

These are the equations for $(\hat{\alpha}, \hat{\beta})$ (but to verify that this is indeed *the maximum*, we should check that [- matrix of 2nd derivatives] is positive- definite at $(\hat{\alpha}, \hat{\beta})$. These equations do not have an explicit solution, but we could solve them iteratively to find $(\hat{\alpha}, \hat{\beta})$, and hence we could evaluate the maximum of L.

$iii)$ Now, if $\beta = 0, L(\alpha, \beta) = -\Sigma e^{\alpha} + \alpha\ t_1$. It is easily seen that this is maximised with respect to α by α^* say, where $\alpha^* = log(t_1/n)$.

We know, by Wilks' theorem, that to test $H_0 : \beta = 0$ against $H_1 : \beta$ arbitrary, we should refer

$$2[L(\hat{\alpha}, \hat{\beta}) - L(\alpha^*, 0)] \text{ to } {\chi_1}^2.$$

4. (An example of how things can be tricky when we are not in the glm family).

$$log f(y|\beta) = -\sum |y_i - \beta| + \text{constant} = -g(\beta) + \text{constant, say}$$

Defining $y_{(1)}, ..., y_{(n)}$ as the ordered sample values, as instructed in the question, we see that

$$g(\beta) = \sum (y_{(i)} - \beta) \text{for } \beta < y_{(1)}$$

(this is a straight line of slope $-n$)

$$g(\beta) = -(y_{(1)} - \beta) + \sum_2^n (y_{(i)} - \beta) \text{ for } y_{(1)} < \beta < y_{(2)}$$

(this is a straight line of slope $-(n+2)$)

and so on....

Finally,

$$g(\beta) = \sum_1^n (\beta - y_{(i)}), \text{for } \beta > y_{(n)}$$

(a straight line of slope n)

Thus, sketching $-g(\beta)$ we see that the log-likelihood function is concave in shape, consisting of straight line segments, and

if n is odd, say $n = 2m + 1$, then $\hat{\beta} = y_{(m+1)}$,

if n is even, say $n = 2m$, then $\hat{\beta}$ is anywhere between $y_{(m)}$ and $y_{(m+1)}$.

$\partial L/\partial \beta$ is of course not defined at $y_{(1)}, ..., y_{(n)}$.

Otherwise it is the slope of the appropriate linear segment.

In this example we could find the asymptotic distribution of $\hat{\beta}$ by going back to first principles (it would make for quite a tough exercise). But we CANNOT find it by quoting the general theorem for the asymptotic distribution of mle's : the appropriate regularity conditions do not hold.

5. $f(y|\theta) = \theta^n exp{-}\theta\ \Sigma y_i$
which is of exponential family form, with
$t(y) = \Sigma y_i$ as our sufficient statistic, having gamma(n, θ) distribution,
and loglikelihood $L = n\ log\theta - \theta\Sigma y_i$. Thus
$\partial L/\partial\theta = n/\theta - t(y)$.
Hence $\hat{\theta} = n/t(y)$, and $\hat{\pi} = -n/t(y)$.

6.

$$log f(y|\alpha,\beta) = \ -\Sigma(y_i - \alpha - \beta\ x_i)^2/2\sigma^2 + \text{constant} = L_n(\alpha,\beta)$$

say .

Now find [- matrix of 2nd derivatives] : show that it is positive-definite. Hence $L_n(\alpha,\beta)$ is a concave function.

7. Here

$$L_n(\theta) = -(n/2)log(\theta) - \Sigma y_i^2/(2\theta) + \text{constant}$$

Thus

$$\partial L/\partial\theta = -(n/2\theta) + \Sigma y_i^2/(2\theta^2)$$

Hence

$$\hat{\theta} = \Sigma y_i^2/n$$

and

$$n\hat{\theta}/\theta = \Sigma y_i^2/\theta,$$

where $Y_i/\sqrt{\theta}$ are $NID(0,1)$.

Hence $n\hat{\theta}/\theta$ is distributed as χ^2_n and hence has mean n, variance 2n.
Thus $E(\hat{\theta}) = \theta$, as we would hope.
Now,

$$\partial^2 L/\partial\theta^2 = n/(2\theta^2) - \Sigma(y_i^2/\theta^3)$$

giving

$$E(-\partial^2 L/\partial\theta^2) = n/(2\theta^2).$$

So the asymptotic variance of $\hat{\theta}$ is $(2\theta^2)/n$, which is the same as the exact variance.

8. With Y_i independent $Bi(1,\pi_i)$ as given, we see that

$$f(y|\alpha,\beta) = \Pi {\pi_i}^{y_i}(1-\pi_i)^{1-y_i}$$

Hence,

$$L_n(\alpha,\beta) = \Sigma y_i \; log(\pi_i/(1-\pi_i)) - \Sigma log(1-\pi_i)$$

Substituting for (π_i) gives

$$L_n(\alpha,\beta) = \Sigma(\alpha + \beta \; x_i)y_i - \Sigma log(1 + e^{\alpha+\beta \; x_i})$$

thus,

$$L_n(\alpha,\beta) = \alpha \; y_+ + \beta \; \Sigma x_i y_i - \Sigma log(1 + e^{\alpha+\beta \; x_i})$$

Hence we can write down the equations

$$\partial L_n/\partial\alpha = 0, \;\; \partial L_n/\partial\beta = 0.$$

Observe that there is no closed form solution to these equations. We can only find the mle's by an iterative solution.

9. With Y_i dist'd as $NID(\beta_1 + \beta_2 x_i + \beta_3 P_2(x_i), \sigma^2)$
where $\Sigma x_i = 0, \Sigma P_2(x_i) = 0, \Sigma x_i P_2(x_i) = 0$
we see that the loglikelihood is $l(\beta)$ + constant, where

$$l(\beta) = -\Sigma(y_i - \beta_1 - \beta_2 \; x_i - \beta_3 \; P_2(x_i))^2/2\sigma^2.$$

This gives

$$\partial l/\partial\beta_1 = \Sigma(y_i - \beta_1 - \beta_2 \; x_i - \beta_3 \; P_2(x_i))/\sigma^2 = \Sigma(y_i - \beta_1)/\sigma^2.$$

Similarly,

$$\partial l/\partial\beta_2 = \Sigma x_i(y_i - \beta_2 x_i)/\sigma^2$$

$$\partial l/\partial\beta_3 = \Sigma P_2(x_i) \; (y_i - \beta_3 P_2(x_i))/\sigma^2.$$

Hence

$$\begin{aligned}\partial^2 l/\partial\beta\partial\beta^T &= (-1/\sigma^2) \; diag(n, \Sigma x_i^2, \Sigma(P_2(x_i))^2) \\ &= E(\partial^2 l/\partial\beta\partial\beta^T).\end{aligned}$$

Solving $\partial l/\partial\beta = 0$, gives $\hat{\beta}$, in particular

$$\hat{\beta}_3 = \Sigma y_i \; P_2(x_i)/\Sigma(P_2(x_i))^2.$$

Now, since $E(Y_i) = \beta_1 + \beta_2 x_i + \beta_3 P_2(x_i)$, we can see that
$E(\hat{\beta}_3) = \beta_3$.
Further, Y_i are NID, each with variance σ^2, hence
$\hat{\beta}_3$ has variance $\sigma^2/\Sigma(P_2(x_i))^2$, and is normally distributed.
We can test $H_0 : \beta_3 = 0$, by referring
$\hat{\beta}_3/\sqrt{(\text{its variance})}$ to $N(0,1)$.

Given, $P_2(x_i) = {x_i}^2 + a\ x_i + b$, we find a, b by solving the pair of equations

$$\Sigma({x_i}^2 + a\ x_i + b) = 0, \Sigma({x_i}^2 + a\ x_i + b)x_i = 0$$

giving

$$\Sigma x_i^2 + n\ b = 0, \Sigma x_i^3 + a\ \Sigma x_i^2 = 0.$$

The advantage of parametrising the model in ths way is that the parameters $\beta_1, \beta_2, \beta_3$ are *orthogonal*. Thus for example, if we want to fit

$E(Y_i) = \beta_1 + \beta_2\ x_i$

we find that the lse's for β_1, β_2 are the same as in the full model, ie the same as they were when we included the term β_3 .

9. We may write

$$l(\alpha, \beta) = \Sigma y_i log(p_i/(1 - p_i)) + \Sigma n_i log(1 - p_i)$$

from which we may find the expressions for the partial derivatives; for general link function $g(\ )$ these do not simplify.

i) The sufficient statistics for (α, β) for the logit link are

$$(\Sigma y_i, \Sigma y_i x_i).$$

The logit link is of course the *canonical* link for the binomial distribution.

ii) For the complementary log log link, there is no reduction in dimensionality for the sufficient statistics: we will still need the original data (y_i, x_i) to construct the likelihood function.