STAT 512 - Assignment 4 - due date is on course outline

1. (a) Define a function $f : \mathbb{R}^n \to \mathbb{R}^1$ by $f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a symmetric matrix. Derive (i) the Jacobian $J_f(\mathbf{x})$, (ii) the gradient $\nabla_f(\mathbf{x})$, (iii) the Hessian $H_f(\mathbf{x})$.

   (b) In a regression problem, with $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ and $\mathbf{X}$ the $n \times p$ matrix with rows $\mathbf{x}_i'$, the 'weighted least squares' (WLS) estimate is the minimizer $\hat{\boldsymbol{\theta}}$ of the sum of squares function $S(\boldsymbol{\theta}) = \sum_{i=1}^n w_i (y_i - \mathbf{x}_i'\boldsymbol{\theta})^2$, for suitable positive 'weights' $\{w_i\}$. Show that, if $\mathbf{X}$ has full rank p, then $S(\boldsymbol{\theta})$ is a convex function of $\boldsymbol{\theta}$ (i.e. its Hessian is positive definite) which is minimized at $\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$, where $\mathbf{W} = diag(w_1, ..., w_n)$.

2. In part (b) of the previous question, suppose that the observations are independent, so that the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{y}$ is diagonal, but that the diagonal elements $\sigma_1^2, ..., \sigma_n^2$ are not all equal. We say the observations are 'heteroscedastic'.

   (a) Show that the covariance matrix of $\hat{\boldsymbol{\theta}}$ is

   $$\text{COV}\left[\hat{\boldsymbol{\theta}}\right] = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1},$$

   so that the variance of a linear combination $\mathbf{a}'\hat{\boldsymbol{\theta}}$ is

   $$\text{VAR}\left[\mathbf{a}'\hat{\boldsymbol{\theta}}\right] = \mathbf{a}'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{a}.$$

   (b) Show that, for *any* such linear combination, $\text{VAR}\left[\mathbf{a}'\hat{\boldsymbol{\theta}}\right]$ is minimized by the choice of weights $w_i = 1/\sigma_i^2$, i.e. by $\mathbf{W} = \boldsymbol{\Sigma}^{-1}$.

3. Suppose that an experiment, with k possible outcomes, is repeated n times. The ith outcome occurs with (unknown) probability $p_i > 0$, and X_i represents the number of times that this outcome occurs. The likelihood function is

   $$\mathcal{L}(\mathbf{p}; \mathbf{x}) = \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i}$$

   for $x_i \in \{0, 1, ..., n\}$ satisfying $\sum_{i=1}^k x_i = n$. Assuming that all $x_i > 0$, obtain the MLEs $\hat{p}_1, ..., \hat{p}_k$ of $p_1, ..., p_k$.

4. Discuss the computation of the least squares estimates in the nonlinear regression model with Michaelis-Menten response:

$$Y_i = \frac{\alpha x_i}{\beta + x_i} + \varepsilon_i.$$

Suggest appropriate starting values for your iterative scheme, and write down the scheme explicitly. (There is a matrix inverse involved; write out the matrix but not its inverse.) Explain how you would estimate the variance of $\hat{\beta}$. *Hint* for starting values: a common approach, and one which works here, is to ignore the random error and transform the mean response so that it becomes linear in (transformed) parameters, which are then estimated by linear regression. For instance $Y = \alpha e^{-\beta x} + \varepsilon$ could be transformed as $\log Y \approx \log \alpha - \beta x$. Then $\log \alpha$ and $-\beta$ become the intercept and slope in a linear regression of $\log Y$ on x. From the output of this regression, estimates of the original parameters can be obtained and used as starting values.

5. Similar to what was done in class for the 2-parameter gamma distribution, discuss ML estimation for the 2-parameter *logistic* distribution, for which the distribution function is

$$P(X \le x) = \frac{1}{1 + e^{-\left(\frac{x-\mu}{\sigma}\right)}}, \quad -\infty < x < \infty, \ \sigma > 0.$$

Assume that the data are n independent observations from this distribution. Discuss an appropriate method of solving the likelihood equations numerically (including starting values). Show that the asymptotic approximation to the distribution of the MLEs $\hat{\mu}, \hat{\sigma}$ is that of uncorrelated and normal (hence independent) r.v.s, with

$$\sqrt{n}\,(\hat{\mu} - \mu) \xrightarrow{\mathcal{L}} N\left(0, 3\sigma^2\right), \quad \sqrt{n}\,(\hat{\sigma} - \sigma) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2/c\right).$$

You should express c as an expected value, but need not evaluate it explicitly (let me know if you do!).

Hints: (i) The various terms, beginning with the density of X, are more easily expressed in terms of hyperbolic functions. (ii) You might introduce the notation $T_i = \frac{X_i - \mu}{\sigma}$ and $\psi(t) = -f'(t)/f(t) \stackrel{how?}{=} \tanh\left(\frac{t}{2}\right)$, and write things in terms of these, rather than using the more complicated explicit expressions, as much as possible. Once (i) and (ii) are employed, you might establish and employ the identity $2\psi'(t) + \psi^2(t) = 1$.

6. A density of an n-dimensional random vector $\mathbf{X}$ is said to be *spherically symmetric* if it is of the form $g\left(\|\mathbf{x}\|\right)$ for some function $g\left(\cdot\right)$. An example is the multivariate normal density, for which $g(t) = (2\pi)^{-n/2}\, e^{-t^2/2}$. Here you will derive the density of $T = \|\mathbf{X}\|$, when $\mathbf{X}$ has a spherically symmetric distribution. Start by making a transformation from $\mathbf{x} \in \mathbb{R}^n$ to spherical coordinates $\mathbf{y} = \left(\theta_1, ..., \theta_{n-1}, t\right)'$ by defining

$$
\begin{aligned}
x_1 &= t\sin\theta_1, \\
x_2 &= t\cos\theta_1\sin\theta_2, \\
x_3 &= t\cos\theta_1\cos\theta_2\sin\theta_3, \\
&\cdots \\
x_{n-1} &= t\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2}\sin\theta_{n-1}, \\
x_n &= t\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2}\cos\theta_{n-1};
\end{aligned}
$$

for $-\frac{\pi}{2} < \theta_i \le \frac{\pi}{2}$ $(i = 1, ..., n-2)$, $-\pi < \theta_{n-1} \le \pi$, and $t \ge 0$.

(a) Show that $t = \|\mathbf{x}\|$.

(b) Show that

$$
\left|\frac{\partial\mathbf{x}}{\partial\mathbf{y}}\right| = t^{n-1}\cos^{n-2}\theta_1\cos^{n-3}\theta_2\cdots\cos\theta_{n-2}.
$$

For this, first verify to your own satisfaction (I don't want to look at it!) that

$$
\left(\frac{\partial\mathbf{x}}{\partial\mathbf{y}}\right)\cdot
\begin{pmatrix}
\cos\theta_1 & 0 & 0 & \cdots & 0 \\
0 & \cos\theta_2 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cos\theta_{n-1} & 0 \\
t\sin\theta_1 & t\sin\theta_2 & \cdots & t\sin\theta_{n-1} & 1
\end{pmatrix}
$$

$$
=
\begin{pmatrix}
t & * & * & \cdots & & * \\
0 & t\cos\theta_1 & * & \cdots & & * \\
\vdots & \vdots & \ddots & & & \vdots \\
0 & 0 & 0 & t\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2} & & * \\
0 & 0 & \cdots & 0 & & \cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2}\cos\theta_{n-1}
\end{pmatrix},
$$

where '$*$' denotes elements whose values are not needed.

(c) Show that $\int_{-\pi/2}^{\pi/2}\cos^{h-1}\theta\,d\theta = \beta\left(h/2, 1/2\right)$.

(d) Using these results, show that the density of $\mathbf{y}$ is

$$\psi(\mathbf{y}) = t^{n-1} g(t) \cos^{n-2} \theta_1 \cos^{n-3} \theta_2 \cdots \cos \theta_{n-2},$$

and that the components of $\mathbf{y}$ are independently distributed, with densities

$$
\begin{aligned}
\psi_i(\theta_i) &= \frac{\cos^{n-i-1} \theta_i}{\beta\left(\frac{n-i}{2}, \frac{1}{2}\right)}, -\frac{\pi}{2} < \theta_i \le \frac{\pi}{2} \ (i = 1, ..., n-2), \\
\psi_{n-1}(\theta_{n-1}) &= \frac{1}{2\pi}, -\pi < \theta_{n-1} \le \pi, \text{ and} \\
\psi_n(t) &= \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)} t^{n-1} g(t), \ t \ge 0.
\end{aligned}
$$

(e) By choosing an appropriate spherical density for $\mathbf{X}$, show that the volume $V = \int_{\|\mathbf{x}\| \le 1} d\mathbf{x}$ of the unit sphere in $\mathbb{R}^n$ is $V = \pi^{n/2} / \Gamma\left(\frac{n}{2} + 1\right)$.

(f) Use (d), and an appropriate spherical density for $\mathbf{X}$, to derive the χ_n^2 density.

7. Define a function on the real line by

$$f_0(x) = \begin{cases} (1 - \varepsilon)\,\phi(x), & |x| \le a, \\ (1 - \varepsilon)\,\phi(a)e^{-a(|x|-a)}, & |x| \ge a; \end{cases}$$

where $\phi(x) = \left(1/\sqrt{2\pi}\right)\exp\left(-x^2/2\right)$ is the $N(0,1)$ density. Recall that this was asserted, in class, to be the least favourable density for M-estimation in the "gross-errors" class $\mathcal{F} = \{F \mid f(x) = (1 - \varepsilon)\,\phi(x) + \varepsilon g(x)\}$. To complete the verification of this assertion, show that:

(a) For each $\varepsilon \in [0, 1)$ there is a *unique* $a = a(\varepsilon) \ge 0$ satisfying $\int f_0(x)dx = 1$; thus f_0 is a probability density.

(b) There is a *non-negative* function $g_0(x)$ such that

$$f_0(x) = (1 - \varepsilon)\,\phi(x) + \varepsilon g_0(x);$$

thus g_0 is a probability density.

(c) The MLE for location, based on a sample $X_1, ..., X_n$ with a density $f_0(x - \theta)$, is the solution to $\sum \psi_0(X_i - \hat{\theta}) = 0$ where

$$\psi_0(x) = sign(x) \cdot \min\left(|x|, a\right)$$

is "Huber's" minimax ψ-function obtained in Lecture 24.