Tutorial 6

1. **Maximum Likelihood: Hypergeometric** Suppose X has probability function

$$\mathbb{P}(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \qquad k=0,1,\ldots,n \qquad 0\le n\le M\le N$$

where N, M, n are non negative integers. Show that the maximum likelihood estimate of M for N and n fixed is given by

$$\widehat{M}(X)=\left\lfloor \frac{X}{n}(N+1)\right\rfloor$$

if $\frac{X}{n}(N+1)$ is not an integer and

$$\widehat{M}(X)=\frac{X}{n}(N+1) \qquad \text{or} \qquad \frac{X}{n}(N+1)-1$$

otherwise, where $\lfloor x\rfloor$ denotes the integer part of x.

Hint: Consider the ratio $\frac{L(M+1,x)}{L(M,x)}$ as a function of M.

2. **Maximum Likelihood** Suppose $X_1,\ldots,X_n$ is a sample from a population with density

$$f(x;\mu,\sigma^2)=\frac{9}{10\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)+\frac{1}{10}\phi(x-\mu)$$

where ϕ defined as $\phi=\frac{1}{\sqrt{2\pi}}\exp\{-x^2/2\}$ $-\infty<x<+\infty$ is the standard normal density function and the parameter space is $(\mu,\sigma)\in\Theta=\mathbb{R}\times(0,+\infty)$. Show that the maximum likelihood estimator for the pair (μ,σ) does not return a good answer if $\sigma>0$.

3. Let $X_1,\ldots,X_n$ be i.i.d., with parent distribution $U((\theta-\frac{1}{2},\theta+\frac{1}{2}))$ where θ is an unknown parameter. That is, the distribution with density

$$p(x;\theta)=\mathbf{1}_{(\theta-\frac{1}{2},\theta+\frac{1}{2})}(x).$$

Find the maximum likelihood estimator of θ.

4. (a) Let Y be any random variable and let $R(c)=\mathbb{E}[|Y-c|]$ be the *mean absolute prediction error*. Show that either $R(c)\equiv+\infty$ or else $R(c)$ is minimised by any number c_0 such that

$$\mathbb{P}(Y\ge c_0)\ge\frac{1}{2} \qquad \text{and} \qquad \mathbb{P}(Y\le c_0)\ge\frac{1}{2}.$$

A number c satisfying this property is known as the *median*.

Hint: First show that if $c<c_0$ then

$$\mathbb{E}[|Y-c_0|]=\mathbb{E}[|Y-c|]-(c_0-c)(\mathbb{P}(Y\ge c_0)-\mathbb{P}(Y<c_0))-2\mathbb{E}[(Y-c)\mathbf{1}_{(c,c_0)}(Y)]$$

and consider the consequences if c_0 is the median. Consider a symmetric argument for $c>c_0$.

(b) Suppose that $Y_1, \dots, Y_n$ are independent with Y_i having the Laplace density

$$\frac{1}{2\sigma} \exp\left\{-\frac{|y_i - \mu_i|}{\sigma}\right\} \qquad \sigma > 0,$$

where $\mu_i = \sum_{j=1}^{p} z_{ij}\beta_j$. The z_{ij} are fixed and known, the β_j are unknown parameters.

i. Show that the MLE of $(\beta_1, \dots, \beta_p, \sigma)$ is obtained by finding $\widehat{\beta}_1, \dots, \widehat{\beta}_p$ that minimises the *least absolute deviation contrast function* $\sum_{j=1}^{n} |y_j - \mu_j|$ and then setting $\widehat{\sigma} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{\mu}_i|$ where $\widehat{\mu}_i = \sum_{j=1}^{p} z_{ij}\widehat{\beta}_j$.

ii. Suppose $\mu_i = \mu$ for each i. Show that the *sample median* $\widehat{y}$ is the minimiser of $\sum_{i=1}^{n} |y_i - \mu|$.

5. Let $X \sim Poiss(n(\mu_1 + \mu_2))$, $Y \sim Poiss(m\mu_1)$ and $Z \sim Poiss(m\mu_2)$ be independent variables, where n and m are fixed and known. Find the MLE of (μ_1, μ_2) based on (X, Y, Z).

6. Let $X_1, \dots, X_n$ be i.i.d. with density $\frac{1}{\sigma} f_0\left(\frac{x-\mu}{\sigma}\right)$, $\sigma > 0$ and $\mu \in \mathbb{R}$. Let $w = -\log f_0$ and assume that w'' exists and satisfies $w'' > 0$; $w(\pm\infty) = +\infty$.

(a) Show that if $n \geq 2$, the likelihood equations are:

$$\begin{cases} \sum_{i=1}^{n} w'\left(\frac{X_i-\mu}{\sigma}\right) = 0 \\ \sum_{i=1}^{n} \left\{\frac{(X_i-\mu)}{\sigma} w'\left(\frac{X_i-\mu}{\sigma}\right) - 1\right\} = 0 \end{cases}$$

and that they have a unique solution $(\widehat{\mu}, \widehat{\sigma})$.

Hint Show that the function $D(a,b) = \sum_{i=1}^{n} w(aX_i - b) - n\log a$ is strictly convex in the variables (a,b) and $\lim_{(a,b)\to(a_0,b_0)} D(a,b) = +\infty$ if either $a_0 = 0$ or $+\infty$, or $b_0 = \pm\infty$. You may use the following:

- If a strictly convex function has a minimum, then it is unique.
- For a function D of two variables, if $\frac{\partial^2 D}{\partial a^2} > 0$, $\frac{\partial^2 D}{\partial b^2} > 0$ and $\frac{\partial^2 D}{\partial a^2}\frac{\partial^2 D}{\partial b^2} > \left(\frac{\partial^2 D}{\partial a \partial b}\right)^2$ then D is strictly convex.

(b) Suggest an algorithm, using Newton-Raphson techniques applied to the problem of locating the minimum of $D(a,b)$ such that, with initial conditions $\widehat{\mu}^{(0)} = 0$, $\widehat{\sigma}^{(0)} = 1$, $\widehat{\mu}^{(i)} \to \widehat{\mu}$ and $\widehat{\sigma}^{(i)} \to \widehat{\sigma}$.

(c) Show that for the logistic distribution (c.d.f. $F_0(x) = \frac{1}{1+e^{-x}}$ for $-\infty < x < +\infty$), w is strictly convex. Give the likelihood equations in this case for μ and σ.

7. Let $X_1, \dots, X_n$ be i.i.d. random p-vectors, with density

$$f(x;\theta) = c(\alpha)\exp\{-\|x-\theta\|^{\alpha}\} \qquad \theta \in \mathbb{R}^p \qquad \alpha \geq 1$$

where $\frac{1}{c(\alpha)} = \int_{\mathbb{R}^p} \exp\{-\|x\|^{\alpha}\}\, dx$, $\|.\|$ denotes the Euclidean norm.

(a) Show that if $\alpha > 1$, then the MLE $\widehat{\theta}$ exists and is unique.

(b) Show that if $\alpha = 1$ and $p = 1$, then the MLE $\widehat{\theta}$ exists, but is not unique if n is even.

8. Let $X_1 \sim N(\theta_1, 1)$ and $X_2 \sim N(\theta_2, 1)$ be independent. Find the maximum likelihood estimates of θ_1 and θ_2 when it is known that $\theta_1 \le \theta_2$.

9. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ populations respectively. Show that the MLE of $\theta = (\mu_1, \mu_2, \sigma^2)$ is

$$\widehat{\theta} = (\overline{X}, \overline{Y}, \widehat{\sigma^2})$$

where

$$\widehat{\sigma^2} = \frac{1}{m+n}\left(\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2\right)$$

10. Suppose that $T(X)$ is sufficient for θ and that $\widehat{\theta}(X)$ is a maximum likelihood estimator of θ. Show that if $\widehat{\theta}$ is unique, then it depends on X only through $T(X)$. (Use the factorisation theorem)

11. (a) Let $X \sim \mathbb{P}_\theta$, $\theta \in \Theta$ and let $\widehat{\theta}$ denote the MLE of θ. Suppose that h is a one-to-one function from Θ onto $h(\Theta)$. Define $\eta = h(\theta)$ and let $p(x, \eta)$ denote the density or probability mass function in terms of η (i.e. reparametrise the model using η). Show that the MLE of η is $h(\widehat{\theta})$. In other words, the MLE is unaffected by reparametrisation; they are equivalent under one-to-one transformations.

    (b) Let $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^p$, $p \ge 1$ be a family of models for X, with state space $\mathcal{X} \subset \mathbb{R}^d$. Let q be a map from Θ onto Ω, where $\Omega \subset \mathbb{R}^k$, $1 \le k \le p$. Show that if $\widehat{\theta}$ is a MLE of θ, then $q(\widehat{\theta})$ is a MLE of $\omega = q(\theta)$.
    Hint Let $\Theta(\omega) = \{\theta \in \Theta : q(\theta) = \omega\}$, then $\{\Theta(\omega) : \omega \in \Omega\}$ is a partition of Θ and $\widehat{\theta}$ belongs to only one member of this partition, say $\Theta(\widehat{\omega})$. Because q is onto Ω, it follows that for each $\omega \in \Omega$ there is $\theta \in \Omega$ such that $\omega = q(\theta)$. Thus the MLE of ω is by definition

$$\widehat{\omega}_{MLE} = \arg\sup_{\omega \in \Omega} \sup \{L(\theta; X) : \theta \in \Theta(\omega)\}$$

    where arg sup means the value of ω which maximises Now show that $\widehat{\omega}_{MLE} = q(\widehat{\theta})$.

12. **E-M Algorithm**

    (a) Suppose that we only observe $S(X)$, where S is a function of X (the observation). Suppose that X is discrete and let $q(s, \theta)$ denote the mass function for S when θ is the parameter. Show that

$$\frac{q(s,\theta)}{q(s,\theta_0)} = \mathbb{E}_{\theta_0}\left[\left.\frac{p(X,\theta)}{p(X,\theta_0)}\right| S(X) = s\right]$$

    (b) Establish part 2. of Theorem 5.6. Do this by showing that $\{(\theta_m, \theta_{m+1})\}$ has a subsequence converging to $\{(\theta^*, \theta^*)\}$ and that therefore θ^* is a global minimiser.

Short Answers

1.
$$\frac{L(M+1,x)}{L(M,x)} = \frac{(M+1)(N-M-n+x)}{(M+1-x)(N-M)}$$
$$\frac{L(M+1,x)}{L(M,x)} > 1 \Leftrightarrow M < \frac{x(N+1)}{n} - 1$$

It follows that
$$L(M+1,x) \leq L(M,x) \Leftrightarrow M \geq \frac{x(N+1)}{n} - 1$$

so that $L(M,x)$ is the maximum value if and only if

$$M = \begin{cases} \left\lfloor \frac{x(N+1)}{n} \right\rfloor & \frac{x(N+1)}{n} \notin \mathbb{Z}_+ \\ \frac{x(N+1)}{n} - 1 & \frac{x(N+1)}{n} \in \mathbb{Z}_+ \\ \frac{x(N+1)}{n} & \frac{x(N+1)}{n} \in \mathbb{Z}_+ \quad \text{and} \quad \frac{L(M+1,x)}{L(M,x)} = 1 \end{cases}$$

2.
$$L(\mu,\sigma;x_1,\ldots,x_n) = \prod_{j=1}^n \left(\frac{9}{10\sigma}\phi\left(\frac{x_j-\mu}{\sigma}\right) + \frac{1}{10}\phi(x_j-\mu)\right)$$

Clearly, taking $\mu = x_j$ for any $j \in \{1,\ldots,n\}$:

$$\lim_{\sigma\to 0} L(x_1,\ldots,x_n;x_j,\sigma) = +\infty$$

Hence $(\widehat{\mu},\widehat{\sigma}) = (x_j,0)$ for any $j \in \{1,\ldots,n\}$ returns a value of $+\infty$ for the likelihood. For any $\sigma > 0$ and any $\mu \in \mathbb{R}$, $L(\mu,\sigma) < +\infty$, hence $\widehat{\sigma}_{ML} = 0$ irrespective of the true value of σ.

3.
$$L(\theta;x_1,\ldots,x_n) = \prod_{j=1}^n \mathbf{1}_{(x_j-\frac{1}{2},x_j+\frac{1}{2})}(\theta) = \mathbf{1}_{(\max_j x_j - \frac{1}{2}, \min_j + \frac{1}{2})}(\theta)$$

so $\widehat{\theta}_{ML}$ is not unique; any value $\widehat{\theta}_{ML} \in (\max_j x_j - \frac{1}{2}, \min_j x_j + \frac{1}{2})$ maximises the likelihood.

4. (a) For $c < c_0$,

$$\begin{aligned}
\mathbb{E}[|Y-c_0|] &= \mathbb{E}[(Y-c_0)\mathbf{1}_{\{Y>c_0\}}] + \mathbb{E}[(c_0-Y)\mathbf{1}_{\{Y<c_0\}}] \\
&= \mathbb{E}[(Y-c)\mathbf{1}_{\{Y>c\}}] - \mathbb{E}[(Y-c)\mathbf{1}_{\{c<Y\leq c_0\}}] - (c_0-c)\mathbb{P}(Y>c_0) \\
&\quad +\mathbb{E}[(c-Y)\mathbf{1}_{\{Y<c\}}] + \mathbb{E}[(c-Y)\mathbf{1}_{\{c<Y<c_0\}}] + (c_0-c)\mathbb{P}(Y<c_0) \\
&= \mathbb{E}[|Y-c|] - (c_0-c)\left(\mathbb{P}(Y\geq c_0) - \mathbb{P}(Y<c_0)\right) - 2\mathbb{E}[(Y-c)\mathbf{1}_{(c,c_0)}(Y)]
\end{aligned}$$

Now choose c_0 such that $\mathbb{P}(Y \geq c_0) \geq \frac{1}{2}$, $\mathbb{P}(Y \leq c_0) = \frac{1}{2}$, then $R(c_0) \leq R(c)$. The inequality is strict unless both $\mathbb{P}(Y = c_0) = 0$ and $\mathbb{P}(Y \in (c,c_0)) = 0$. It follows that $R(c_0) < R(c)$ unless c also satisfies $\mathbb{P}(Y \geq c) \geq \frac{1}{2}$ and $\mathbb{P}(Y \leq c) \geq \frac{1}{2}$. Similar arguments for $c > c_0$. It follows that a value c minimises if and only if it is a median.

(b) i. Log likelihood

$$\log L(\beta, \sigma; y_1, \dots, y_n) = -n \log 2 - n \log \sigma - \frac{1}{\sigma} \sum_{i=1}^{n} |y_i - \mu_i|$$

Let f denote the maximum of $\sum_{i=1}^{n} |y_i - \mu_i|$. Then, for β that gives f,

$$\frac{\partial}{\partial \sigma} \log L = -\frac{n}{\sigma} + \frac{1}{\sigma^2} f \Rightarrow \widehat{\sigma} = \frac{1}{f}$$

The result follows.

ii. Consider the empirical distribution defined by $Y_{(1)}, \dots, Y_{(n)}$ and apply the result of the first part.

5.

$$\mathbb{P}(X = x, Y = y, Z = z) = \frac{(n(\mu_1 + \mu_2))^x (m\mu_1)^y (m\mu_2)^z}{x!y!z!} e^{-(n(\mu_1+\mu_2))-m\mu_1-m\mu_2)}$$

$$\begin{aligned} \log L(\mu_1, \mu_2; x, y, z) = & \left(x \log \frac{n}{x!y!z!} + x \log(\mu_1 + \mu_2) + y \log m + z \log m \right) \\ & + y \log \mu_1 + z \log \mu_2 - \mu_1(n + m) - \mu_2(n + m) \end{aligned}$$

A critical point, if it is in $(0, \infty) \times (0, \infty)$ (the interior) satisfies

$$\begin{cases} \frac{\partial}{\partial \mu_1} \log L(\mu_1, \mu_2) = \frac{x}{\mu_1+\mu_2} + \frac{y}{\mu_1} - (n + m) = 0 \\ \frac{\partial}{\partial \mu_2} \log L(\mu_1, \mu_2) = \frac{x}{\mu_1+\mu_2} + \frac{z}{\mu_2} - (n + m) = 0 \end{cases}$$

If $x > 0$, $y > 0$ and $z > 0$, there is exactly one solution to these equations. From the equations, $\frac{z}{\mu_2} = \frac{y}{\mu_1}$ so that

$$\frac{x}{\mu_1 + \frac{z\mu_1}{y}} + \frac{y}{\mu_1} = n + m \Rightarrow \frac{xy}{(y + z)\mu_1} + \frac{y}{\mu_1} = n + m \Rightarrow \mu_1 = \frac{y(x + y + z)}{(y + z)(n + m)}$$

giving

$$\mu_2 = \frac{z(x + y + z)}{(n + m)(y + z)} \qquad \mu_1 = \frac{y(x + y + z)}{(n + m)(y + z)}.$$

It turns out that this (μ_1, μ_2) gives a global maximum in $\mathbb{R}_+ \times \mathbb{R}_+$ in all cases. To see this, we consider the boundaries of $\mathbb{R}_+ \times \mathbb{R}_+$ which are: $\mu_1 + \mu_2 \to +\infty$, $\mu_1 \to 0$ for $\mu_2 < +\infty$ and $\mu_2 \to 0$ for $\mu_1 < +\infty$. The different cases are as follows:

1) If $x > 0$, $y > 0$, $z > 0$, then $\log L(\mu_1, \mu_2; x, y, z)$ is strictly concave in (μ_1, μ_2), bounded above, and $\log L(\mu_1, \mu_2; x, y, z) \overset{\mu_1+\mu_2 \to +\infty}{\longrightarrow} -\infty$, $\log L(\mu_1, \mu_2) \to -\infty$ if $\mu_1 \to 0$ (μ_2 fixed) or $\mu_2 \to 0$ (μ_1 fixed).

Therefore, from strict concavity and differentiability, that the maximum is unique and is in the interior of the domain and satisfies $\frac{\partial \log L}{\partial \mu_1} = \frac{\partial \log L}{\partial \mu_2} = 0$.

2) If $x > 0$, $y > 0$, $z = 0$, then $\log L(\mu_1, \mu_2)$ is strictly concave, but there is no solution to the equations $\frac{\partial}{\partial \mu_1} \log L = \frac{\partial}{\partial \mu_2} \log L = 0$ in $(0, +\infty) \times (0, +\infty)$ and hence the maximum is on the boundary. $\mathcal{L}(\mu_1, \mu_2) \overset{\mu_1+\mu_2 \to +\infty}{\longrightarrow} -\infty$, $\log L(\mu_1, \mu_2) \overset{\mu_1 \to 0}{\longrightarrow} -\infty$ for μ_2 fixed.

Therefore, the part of the boundary where the maximum is achieved is $\mu_2 = 0$. The problem now reduces to finding μ_1 that maximises $\log L(\mu_1, 0; x, y, 0)$ which is $\mu_1 = \frac{x+y}{n+m}$.

3) Similarly for $x > 0$, $y = 0$, $z > 0$.

4) $x > 0$, $y = z = 0$ - the only thing that can be estimated is $\mu_1 + \mu_2$, the estimate is

$$\widehat{\mu_1 + \mu_2} = \frac{x}{m+n}$$

5) $x = 0$: This splits into two separate estimation problems, μ_1 which maximises $y \log \mu_1 - \mu_1(n+m)$ and μ_2 which maximises $z \log \mu_2 - \mu_2(n+m)$ which gives

$$\mu_1 = \frac{y}{n+m} \qquad \mu_2 = \frac{z}{n+m}.$$

6. (a) $\log L(\mu, \sigma) = \log \prod_{j=1}^n \left(\frac{1}{\sigma} f_0 \left(\frac{x_j - \mu}{\sigma} \right) \right) = -n \log \sigma - \sum_{j=1}^n w \left(\frac{x_j - \mu}{\sigma} \right)$. Likelihood equations are $\nabla \log L(\mu, \sigma) = 0$ giving

$$\begin{cases} \sum_{j=1}^n w' \left(\frac{x_j - \mu}{\sigma} \right) = 0 \\ \sum_{i=1}^n \left\{ \frac{(x_i - \mu)}{\sigma} w' \left(\frac{X_i - \mu}{\sigma} \right) - 1 \right\} = 0 \end{cases}$$

as required. For uniqueness, consider the function $D(a, b)$. Then

$$\begin{aligned} \nabla D(a,b) = 0 \quad &\Leftrightarrow \quad \left(\sum_{i=1}^n X_i w'(aX_i - b) - \frac{n}{a}, - \sum_{i=1}^n w'(aX_i - b) \right) = 0 \\ &\Leftrightarrow \quad \left(\sum_{i=1}^n \left\{ \frac{(X_i - \mu)}{\sigma} w' \left(\frac{X_i - \mu}{\sigma} \right) - 1 \right\}, \sum_{i=1}^n w' \left(\frac{X_i - \mu}{\sigma} \right) \right) = 0 \end{aligned}$$

using $a = \frac{1}{\sigma}$ and $b = \frac{\mu}{\sigma}$. Now,

$$\frac{\partial^2 D}{\partial a^2} = \sum_{i=1}^n X_i^2 w''(aX_i - b) + \frac{n}{a^2}, \quad \frac{\partial^2 D}{\partial b^2} = \sum_{i=1}^n w''(aX_i - b), \quad \frac{\partial^2 D}{\partial a \partial b} = - \sum_{j=1}^n X_j w''(aX_j - b)$$

$$\left(\frac{\partial^2 D}{\partial a \partial b}\right)^2 = \left(\sum_{j=1}^n X_j w''(aX_j - b)\right)^2$$
$$\leq \sum_{j=1}^n X_j^2 w''(aX_j - b) \sum_{j=1}^n w''(aX_j - b) < \left(\frac{\partial^2 D}{\partial a^2}\right)\left(\frac{\partial^2 D}{\partial b^2}\right)$$

using $|\sum c_i d_i| \leq (\sum c_i^2)^{1/2}(\sum d_i^2)^{1/2}$; $c_i = X_i\sqrt{w''(aX_i - b)}$ and $d_i = \sqrt{w''(aX_i - b)}$ from which convexity follows. We're using $w'' > 0$.

Finally, we have to show that $\lim_{(a,b)\to(a_0,b_0)} D(a,b) = +\infty$ for (a_0, b_0) as described. The only part which requires attention is: $a_0 = +\infty$. But $w'' > 0$ implies that w' is *increasing*. Since $w(\pm\infty) = +\infty$, this implies that there exists an x_0 such that $w(x_0) = \min_x w(x)$, that $\lim_{x\to+\infty}(-w'(x)) = c_1 > 0$ and $\lim_{x\to+\infty} w'(x) = c_2 > 0$ where c_1 and/or c_2 may be $+\infty$. From this, it is clear that unless $X_1 = \ldots = X_n = 0$, $\lim_{a\to+\infty} D(a,b) = +\infty$, since $\frac{d}{da}\log a = \frac{1}{a} \overset{a\to+\infty}{\longrightarrow} 0$.

(b) Minimise $D(a,b)$. The matrix of second derivatives is positive definite and well defined. Call it M and let $U = \nabla D$. Then

$$\begin{pmatrix} a^{(i+1)} \\ b^{(i+1)} \end{pmatrix} = \begin{pmatrix} a^{(i)} \\ b^{(i)} \end{pmatrix} - M^{-1}(a^{(i)}, b^{(i)})U(a^{(i)}, b^{(i)}).$$

(c) $f_0(x) = (1 + e^{-x})^{-2}e^{-x}$ so that

$$w(x) = -\log f_0(x) = 2\log(1 + e^{-x}) + x, \qquad w''(x) = (1 + e^{-x})^{-2}e^{-2x} + (1 + e^{-x})^{-1}e^{-x}$$

so it is strictly convex.

$$w'(x) = -\frac{1}{e^x + 1}$$

Likelihood equations are:

$$\begin{cases} \sum_{i=1}^n \frac{1}{(e^{(x_i-\mu)/\sigma}+1)} = 0 \\ \sum_{i=1}^n \left\{\frac{(x_i-\mu)/\sigma}{(e^{(x_i-\mu)/\sigma}+1)} - 1\right\} = 0 \end{cases}$$

7. (a) For $\alpha > 1$, $\|y\|^\alpha$ is strictly convex in y. This can be seen as follows: for $\alpha > 1$, the function $g : \mathbb{R} \to \mathbb{R}_+$ defined by $g(x) = |x|^\alpha$ is strictly convex. It follows that for $t \in (0,1)$, if $\|x\| \neq \|y\|$, then

$$\begin{aligned} \|tx + (1-t)y\|^\alpha &= \left(t^2\|x\|^2 + 2t(1-t)\langle x, y\rangle + (1-t)^2\|y\|^2\right)^{\alpha/2} \\ &\leq \left(t^2\|x\|^2 + 2t(1-t)\|x\|\|y\| + (1-t)^2\|y\|^2)^{\alpha/2}\right) \\ &= (t\|x\| + (1-t)\|y\|)^\alpha < t\|x\|^\alpha + (1-t)\|y\|^\alpha. \end{aligned}$$

and if $\|x\| = \|y\|$ but $x \neq y$, then $|\langle x, y\rangle| < \|x\|\|y\|$ where the inequality is strict, so that again

$$\|tx + (1-t)y\|^\alpha < t\|x\|^\alpha + (1-t)\|y\|^\alpha.$$

$$\log L(x_1, \ldots, x_n; \theta) = n \log c(\alpha) - \sum_{j=1}^{n} \|x_j - \theta\|^\alpha$$

and the sum of strictly convex functions is again strictly convex. It follows that the likelihood function has a unique maximiser $\widehat{\theta}_{ML}$.

(b)

$$f(x; \theta) = c \exp\{-|x - \theta|\}$$

$$\log L(x_1, \ldots, x_n; \theta) = n \log c - \sum_{j=1}^{n} |x_j - \theta|$$

Problem is therefore to find θ that minimises $\sum_{j=1}^{n} |x_j - \theta|$. It follows from earlier exercise that $\widehat{\theta}$ provides a minimiser where $\widehat{\theta}$ is any sample median. If n is even and $x_{(n/2)} < x_{(n/2)+1}$ then the median is not unique.

8. Minimise

$$(\theta_1 - x_1)^2 + (\theta_2 - x_2)^2$$

subject to the constraint that $\theta_1 \leq \theta_2$. If $x_1 \leq x_2$, then $(\widehat{\theta}_1, \widehat{\theta}_2) = (x_1, x_2)$. If $x_1 > x_2$, then $\widehat{\theta}_1 = \widehat{\theta}_2$ (on the boundary) so that it is the minimiser of

$$2\theta^2 - 2(x_1 + x_2)\theta + (x_1^2 + x_2^2)$$

which is: $\widehat{\theta}_1 = \widehat{\theta}_2 = \frac{x_1 + x_2}{2}$.

9. Minimise:

$$\frac{1}{2\sigma^2}\left(\sum_{j=1}^{m}(x_j - \mu_1)^2 + \sum_{j=1}^{n}(y_j - \mu_2)^2\right) + \frac{(m+n)}{2}\log\sigma^2$$

μ_1 and μ_2 are easy; $\widehat{\mu}_1 = \overline{x}$ and $\widehat{\mu}_2 = \overline{y}$. For σ^2, $\widehat{\sigma^2}$ is the point which satisfies:

$$-\frac{1}{2(\sigma^2)^2}\left(\sum_{j=1}^{m}(x_j - \overline{x})^2 + \sum_{j=1}^{n}(y_j - \overline{y})^2\right) + \frac{m+n}{2\sigma^2}$$

giving the MLE of

$$\widehat{\sigma^2} = \frac{1}{m+n}\left(\sum_{j=1}^{m}(X_j - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2\right)$$

10. Factorisation theorem gives:

$$p(x, \theta) = h(x)g(T(x), \theta);$$

maximising in terms of θ is equivalent to maximising $g(T(x), \theta)$, hence the result follows.

11. (a) $\widehat{\eta}_{ML}$ maximises $p(x, \eta) = p(x, h(\theta))$. The value of θ which maximises this is $\widehat{\theta}_{ML}$, hence if $\widehat{\theta}_{ML}$ is a value of θ which maximises $p(x, \theta)$ then $\eta_{ML} = h(\widehat{\theta}_{ML})$ is a value of η which maximises $p(x, \eta)$. Similarly, if θ does not maximise $p(x, \theta)$, then $\eta = h(\theta)$ does not maximise the reparametrised family $p(x, \eta)$.

    (b) Using the hint, $\widehat{\omega}_{MLE}$ maximises $\sup_{\theta \in \Omega(\omega)} L(\theta; X)$. ω_{ML} is therefore the value of ω that satisfies $\widehat{\theta}_{ML} \in \Omega(\widehat{\omega}_{ML})$ and is therefore (by definition) $\widehat{\omega}_{ML} = q(\widehat{\theta}_{ML})$.

12. (Omitted)