Math 534 Homework 3.1 - 25 points

Mike Palmer
due 2024/02/14

Exercise J-2.2 Write a general function to maximize the following log-likelihood function with respect to parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) = -\frac{1}{2}\left\{ nplog(2\pi) + nlog(|\boldsymbol{\Sigma}|) + trace\left[\boldsymbol{\Sigma}^{-1} c(\boldsymbol{\mu})\right] \right\}$$

where $c(\boldsymbol{\mu}) = \sum_{z=1}^{n} (\boldsymbol{x}_z - \boldsymbol{\mu})(\boldsymbol{x}_z - \boldsymbol{\mu})^T$.

(a) [20 points] Obtain formulas for the elements of the gradient and the Hessian of $\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$d\ell(d\boldsymbol{\mu}) = trace\left(\boldsymbol{\Sigma}^{-1} \sum_{z=1}^{n} (\boldsymbol{x}_z - \boldsymbol{\mu}) \cdot d\boldsymbol{\mu}^T\right)$$

$$dd\ell(d\boldsymbol{\mu}, d\boldsymbol{\mu}) = trace\left[\boldsymbol{\Sigma}^{-1} \cdot (-nd\boldsymbol{\mu}) \cdot d\boldsymbol{\mu}^T\right] = -n \cdot trace\left(\boldsymbol{\Sigma}^{-1} \cdot d\boldsymbol{\mu} \cdot d\boldsymbol{\mu}^T\right)$$

$$d\ell(d\boldsymbol{\Sigma}) = -\frac{n}{2} trace\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{c(\boldsymbol{\mu})}{n}\right)\boldsymbol{\Sigma}^{-1} \cdot d\boldsymbol{\Sigma}\right]$$

$$dd\ell(d\boldsymbol{\Sigma}, d\boldsymbol{\Sigma}) = n \cdot trace\left\{\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{c(\boldsymbol{\mu})}{n}\right) - \frac{1}{2} I\right] \cdot \boldsymbol{\Sigma}^{-1} d\boldsymbol{\Sigma}\ \boldsymbol{\Sigma}^{-1} d\boldsymbol{\Sigma}\right\}$$

$$dd\ell(d\boldsymbol{\mu}, d\boldsymbol{\Sigma}) = trace\left[-\boldsymbol{\Sigma}^{-1} d\boldsymbol{\Sigma}\ \boldsymbol{\Sigma}^{-1} \sum_{z=1}^{n} (\boldsymbol{x}_z - \boldsymbol{\mu}) \cdot d\boldsymbol{\mu}^T\right]$$

For any i,

$$\frac{\partial \ell}{\partial \mu_i} = \left[\Sigma^{-1} \sum_{z=1}^{n} (\boldsymbol{x}_z - \boldsymbol{\mu})\right]_i .$$

When $i = j$,

$$\frac{\partial \ell}{\partial \sigma_{ij}} = -\frac{n}{2} \cdot \left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{c(\boldsymbol{\mu})}{n}\right)\boldsymbol{\Sigma}^{-1}\right]_{ij} .$$

and when $i \neq j$,

$$\frac{\partial \ell}{\partial \sigma_{ij}} = -\frac{n}{2} \cdot \left\{\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{c(\boldsymbol{\mu})}{n}\right)\boldsymbol{\Sigma}^{-1}\right]_{ij} + \left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{c(\boldsymbol{\mu})}{n}\right)\boldsymbol{\Sigma}^{-1}\right]_{ji}\right\} = -n \cdot \left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \frac{c(\boldsymbol{\mu})}{n}\right)\boldsymbol{\Sigma}^{-1}\right]_{ij} .$$

When $i = j$ and $i \neq j$,

$$\frac{\partial^2 \ell}{\partial \mu_i \partial \mu_j} = -n \cdot \left[\boldsymbol{\Sigma}^{-1}\right]_{ij}$$

When $i = j$ and any k,

$$\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \mu_k} = -\sum_{w=1}^{p} \left[\sum_{z=1}^{n} (\boldsymbol{x}_z - \boldsymbol{\mu})\right]_w \cdot \left([\boldsymbol{\Sigma}^{-1}]_{iw} \cdot [\boldsymbol{\Sigma}^{-1}]_{ki}\right)$$

and when $i \neq j$ and any k

$$\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \mu_k} = -\sum_{w=1}^{p} \left[\sum_{z=1}^{n} (\boldsymbol{x}_z - \boldsymbol{\mu}) \right]_w \cdot \left([\boldsymbol{\Sigma}^{-1}]_{iw} \cdot [\boldsymbol{\Sigma}^{-1}]_{kj} + [\boldsymbol{\Sigma}^{-1}]_{jw} \cdot [\boldsymbol{\Sigma}^{-1}]_{ki} \right)$$

Let $A = \left[\boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\Sigma} - \frac{c(\boldsymbol{\mu})}{n} \right) - \frac{1}{2} I \right] \cdot \boldsymbol{\Sigma}^{-1}$

then

$$dd\ell(d\boldsymbol{\Sigma}, d\boldsymbol{\Sigma}) = n \cdot trace \left\{ A \ d\boldsymbol{\Sigma} \ \boldsymbol{\Sigma}^{-1} d\boldsymbol{\Sigma} \right\}.$$

When $i = j$ and $k = l$,

$$\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}} = n \cdot \left\{ [A]_{ik} \cdot [\boldsymbol{\Sigma}^{-1}]_{lj} \right\}$$

and when $i \neq j$ $k = l$

$$\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}} = n \cdot \left\{ [A]_{ik} \cdot [\boldsymbol{\Sigma}^{-1}]_{kj} + [A]_{jk} \cdot [\boldsymbol{\Sigma}^{-1}]_{ki} \right\}$$

and when $i = j$ $k \neq l$

$$\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}} = n \cdot \left\{ [A]_{il} \cdot [\boldsymbol{\Sigma}^{-1}]_{ji} + [A]_{ij} \cdot [\boldsymbol{\Sigma}^{-1}]_{li} \right\}$$

and when $i \neq j$ $k \neq l$

$$\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}} = n \cdot \left\{ [A]_{il} \cdot [\boldsymbol{\Sigma}^{-1}]_{kj} + [A]_{ik} \cdot [\boldsymbol{\Sigma}^{-1}]_{lj} + [A]_{jl} \cdot [\boldsymbol{\Sigma}^{-1}]_{ki} + [A]_{kj} \cdot [\boldsymbol{\Sigma}^{-1}]_{li} \right\}.$$

(b) [5 points] Obtain formulas for the elements of the information matrix.

Using Hint: $E[(x_i - \mu)] = 0$ and $E[(x_i - \mu)(x_i - \mu)^T] = \Sigma$

When $i = j$ and when $i \neq j$,

$$-E \left[\frac{\partial^2 \ell}{\partial \mu_i \partial \mu_j} \right] = n \cdot \left[\boldsymbol{\Sigma}^{-1} \right]_{ij}.$$

When $i = j = k$, and when $i = j \neq k$, and when $i = k \neq j$, and when $i \neq j = k$, and when $i \neq j \neq k$,

$$-E \left[\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \mu_k} \right] = 0.$$

When $i = j$ and $k = l$,

$$-E \left[\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}} \right] = -\frac{n-2}{2} \cdot \left([\boldsymbol{\Sigma}^{-1}]_{ik} \right)^2$$

and when $i \neq j$ and $k \neq l$

$$-E \left[\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}} \right] = \frac{n-2}{2} \cdot \left\{ [\boldsymbol{\Sigma}^{-1}]_{il} \cdot [\boldsymbol{\Sigma}^{-1}]_{kj} + [\boldsymbol{\Sigma}^{-1}]_{ik} \cdot [\boldsymbol{\Sigma}^{-1}]_{lj} + [\boldsymbol{\Sigma}^{-1}]_{jl} \cdot [\boldsymbol{\Sigma}^{-1}]_{ki} + [\boldsymbol{\Sigma}^{-1}]_{kj} \cdot [\boldsymbol{\Sigma}^{-1}]_{li} \right\}$$

and when $i \neq j$ and $k = l$

$$-E\left[\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}}\right] = \frac{n-2}{2} \cdot \left\{[\mathbf{\Sigma}^{-1}]_{ik} \cdot [\mathbf{\Sigma}^{-1}]_{kj} + [\mathbf{\Sigma}^{-1}]_{jk} \cdot [\mathbf{\Sigma}^{-1}]_{ki}\right\}$$

and when $i = j$ and $k \neq l$

$$-E\left[\frac{\partial^2 \ell}{\partial \sigma_{ij} \partial \sigma_{kl}}\right] = \frac{n-2}{2} \cdot \left\{[\mathbf{\Sigma}^{-1}]_{il} \cdot [\mathbf{\Sigma}^{-1}]_{ji} + [\mathbf{\Sigma}^{-1}]_{ij} \cdot [\mathbf{\Sigma}^{-1}]_{li}\right\}.$$