Q1 / Maximum Likelihood Estimation

(a) Consider i.i.d random variables, $X_1, \dots, X_n$ ~~from~~ Gamma distribution with pdf

$$f(x;\alpha,\lambda) = \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} x^{\alpha-1} \exp(-\lambda x), \quad x \geq 0$$

Suppose the parameter $\alpha > 0$ is known, find the MLE of λ.

We start with the function $f(x;\alpha,\lambda) = \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} x^{\alpha-1} \exp(-\lambda x)$

Then we calculate the likelihood function:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} x_i^{\alpha-1} \exp(-\lambda x_i)$$

Then we calculate the log likelihood as:

$$l(\theta) = \sum_{i=1}^{n} \log\left(\frac{1}{\Gamma(\alpha)} \lambda^{\alpha} x_i^{\alpha-1} \exp(-\lambda x_i)\right) \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} \log\left(\frac{1}{\Gamma(\alpha)} x_i^{\alpha-1}\right) + \sum_{i=1}^{n} \log\left(\lambda^{\alpha} \exp(-\lambda x_i)\right) \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} \log\left(\frac{1}{\Gamma(\alpha)} x_i^{\alpha-1}\right) + \sum_{i=1}^{n} \log(\lambda^{\alpha}) + \sum_{i=1}^{n} \log(\exp(-\lambda x_i)) \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} \log\left(\frac{1}{\Gamma(\alpha)} x_i^{\alpha-1}\right) + n\alpha \log(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

Then we take the first derivative of $l(\theta)$ with respect to λ:

$$\frac{\partial l(\theta)}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} x_i$$

Then we take the second derivative to get:

$$\frac{\partial^2 l(\theta)}{\partial \lambda^2} = \frac{-n\alpha}{\lambda^2} < 0.$$

Once the second derivative is always negative, the log likelihood function is concave, therefore the maximum is where $\frac{\partial l(\theta)}{\partial \lambda} = 0$.

Q1) Maximum Likelihood Estimation continued

(a) cont.

Therefore we solve the equation $\frac{n\alpha}{\lambda} - \sum_{i=1}^{n} x_i = 0 = \frac{\partial \ell(\theta)}{\partial \lambda}$

to get $\lambda = \frac{n\alpha}{\sum_{i=1}^{n} x_i}$ recall $\frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$ therefore

$$\boxed{\hat{\lambda}_{MLE} = \frac{\alpha}{\bar{x}}}$$

(b) Let $x_1, \dots, x_n$ be i.i.d d-dimensional Gaussian random variables distributed according to $\mathcal{N}(\mu, \Sigma)$. That is,

$$f(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

find the MLE for vector $\vec{\mu}$.

We start with the function $\frac{1}{\sqrt{2\pi^d |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$

we then get the likelihood function to be:

$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi^d |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}_i-\vec{\mu}_i)^T \Sigma^{-1} (\vec{x}_i-\vec{\mu}_i)\right)$ Then we calculate

log likelihood:

$\ell(\theta) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi^d |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}_i-\vec{\mu}_i)^T \Sigma^{-1} (\vec{x}_i-\vec{\mu}_i)\right)\right)$ then we use log

rules to get:

$$\ell(\theta) = -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T \Sigma^{-1} (\vec{x}_i - \vec{\mu})$$

Then we take the first derivative of $\ell(\theta)$ with respect to μ:

$$\frac{\partial}{\partial \mu}\ell(\theta) = \underbrace{\frac{\partial}{\partial \mu}\left(-n\frac{d}{2}\log(2\pi)\right)}_{0} - \underbrace{\frac{\partial}{\partial \mu}\left(n\frac{1}{2}\log|\Sigma|\right)}_{0} - \frac{\partial}{\partial \mu}\left(\frac{1}{2}\sum_{i=1}^{n}(\bar{x}_i-\mu)^T \Sigma^{-1}(\bar{x}_i-\mu)\right)$$

Recall that $\frac{\partial}{\partial a} a^T b = \frac{\partial}{\partial a} b^T a = b$, therefore $\frac{\partial}{\partial \mu}\ell(\theta)$ reduces to...

$$\frac{\partial}{\partial \mu}\ell(\theta) = \sum_{i=1}^{n} \Sigma^{-1}(x_i - \mu).$$

we then take the second derivative to get

$$\frac{\partial^2}{\partial \mu^2}\ell(\theta) \underbrace{\frac{\partial}{\partial \mu}\sum_{i=1}^{n}\Sigma^{-1}\bar{x}_i}_{0} - \frac{\partial}{\partial \mu}\Sigma^{-1}n\mu \Rightarrow \frac{\partial^2}{\partial \mu^2} = -n\Sigma^{-1} < 0$$

Therefore, once the second derivative is always negative, the maximum is where $\frac{\partial \ell(\theta)}{\partial \mu} = 0$

Q1 Maximum Likelihood Estimation Continued

(b) cont.

$$\sum_{i=1}^{n} \Sigma^{-1}(x_i - \mu) = 0 \Rightarrow \sum_{i=1}^{n} \Sigma^{-1} x_i - \Sigma^{-1} n \mu = 0 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} \cancel{\Sigma^{-1}} x_i = \cancel{\Sigma^{-1}} n \mu \Rightarrow \frac{1}{n} \sum_{i=1}^{n} x_i = \mu \quad \text{recall } \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}$$

Therefore, $\boxed{\hat{\mu}_{MLE} = \overline{X}}$