Tutorial 2 : MLE

1 Theory

Consider a random variable X with probability mass/density function $p(x)$ which has known functional form up to some unknown constant parameters $\theta \in \mathbb{R}^m$, i.e., $p(x) = p(x;\theta)$.

Goal

Given a set of independent and identically distributed (iid) samples of X, denoted by $D = \{x_k\}_{k=1}^{n}$, estimate $p(x)$.

Maximum Likelihood Estimator (MLE)

$$\hat{\theta}_{\text{MLE}} \triangleq \underset{\theta \in \mathbb{R}^m}{\arg\max} \quad p(D;\theta).$$

In most cases we can ease the computation by applying the **log**$(\cdot)$ operation

$$\hat{\theta}_{\text{MLE}} = \underset{\theta \in \mathbb{R}^m}{\arg\max} \quad \log p(D;\theta).$$

Here we use the fact that **log**$(\cdot)$ is monotonically increasing, hence, it does not affect the maximum. Next, we notice that

$$p(D;\theta) = p(x_1, ..., x_n;\theta) = \prod_{k=1}^{n} p(x_k;\theta),$$

since the samples are iid. Therefore, we can write

$$\hat{\theta}_{\text{MLE}} = \underset{\theta \in \mathbb{R}^m}{\arg\max} \quad \sum_{k=1}^{n} \log p(x_k;\theta).$$

For a given set of samples D, we define the **likelihood function** as $L(\theta) \triangleq p(D;\theta)$ and the **log-likelihood function** as $l(\theta) \triangleq \log L(\theta)$.

2 Practice

Question 1

(a) Consider $X \sim N(\mu, \sigma^2)$ where the mean μ and variance σ^2 are unknown.
A set of iid samples of X is given, $D = \{x_k\}_{k=1}^n$.

Prove that $\hat{\mu}_{\text{MLE}} = \frac{1}{n}\sum_{k=1}^n x_k$ and $\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{k=1}^n (x_k - \hat{\mu}_{\text{MLE}})^2$.

(b) Consider $X \sim U[0, \theta]$. Compute $\hat{\theta}_{\text{MLE}}$ given a set of iid samples $D = \{x_k\}_{k=1}^n$.

(c) Consider $X \sim \exp(\lambda)$. Compute $\hat{\lambda}_{\text{MLE}}$ given a set of iid samples $D = \{x_k\}_{k=1}^n$.
You can assume that $\forall k,\ x_k \geq 0$.

Solution

(a) For a given (μ, σ^2) we have

$$p(x = x_k|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{-\frac{(x_k-\mu)^2}{2\sigma^2}\Big\}.$$

Hence,

$$p(D|\mu, \sigma^2) = \Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big)^n \exp\Big\{-\sum_k \frac{(x_k-\mu)^2}{2\sigma^2}\Big\}.$$

According to the definition of MLE

$$\begin{aligned}
\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{MLE}} &= \arg\max_{\mu,\sigma^2} \quad \log p(D|\theta) \\
&= \arg\max_{\mu,\sigma^2} \quad \log\Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big)^n - \sum_k \frac{(x_k-\mu)^2}{2\sigma^2} \\
&= \arg\min_{\mu,\sigma^2} \quad \frac{n}{2}\log\sigma^2 + \frac{1}{2\sigma^2}\sum_k (x_k-\mu)^2.
\end{aligned}$$

Taking the partial derivatives with respect to μ and σ^2 we get

$$\begin{aligned}
\frac{\partial L}{\partial \mu} &= \frac{1}{2\sigma^2}\sum_k (\mu - x_k) = 0, \\
\frac{\partial L}{\partial \sigma^2} &= \frac{n}{2\sigma^2} - \frac{1}{2(\sigma^2)^2}\sum_k (\mu - x_k) = 0.
\end{aligned}$$

From the first equation we get that $\hat{\mu}_{\text{MLE}} = \frac{1}{n}\sum_k x_k$ and from the second we have

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_k (x_k - \hat{\mu}_{\text{MLE}})^2.$$

(b) Notice that for a given θ

$$p(x = x_k|\theta) = \begin{cases} \frac{1}{\theta}, & 0 \leq x_k \leq \theta, \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$p(D|\theta) = \begin{cases} \frac{1}{\theta^n}, & \forall k,\ 0 \leq x_k \leq \theta \\ 0, & \text{otherwise.} \end{cases}$$

The MLE is given by

$$\begin{aligned}
\hat{\theta}_{\text{MLE}} &= \arg\max_{\theta\in\mathbb{R}} \quad \log p(D|\theta) \\
&= \arg\max_{\theta\in\mathbb{R}} \begin{cases} -n\log\theta, & \forall k,\ 0 \leq x_k \leq \theta \\ -\infty, & \text{otherwise.} \end{cases} \\
&= \arg\max_{\forall k,\ 0\leq x_k\leq\theta} \ -n\log\theta \\
&= \arg\min_{\forall k,\ 0\leq x_k\leq\theta} \quad \theta \\
&= \max_{1\leq k\leq n} \quad x_k
\end{aligned}$$

(c) Recall that

$$p(x = x_k|\lambda) = \begin{cases} \lambda \exp(-\lambda x_k), & x_k \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

We assume that $x_k \geq 0$, hence,

$$p(D|\lambda) = \lambda^n \exp(-\lambda \sum_{k=1}^{n} x_k).$$

Therefore,

$$\hat{\lambda}_{\text{MLE}} = \arg\max_{\lambda \geq 0} \quad n \log \lambda - \lambda \sum_k x_k.$$

Computing the derivative with respect to λ we get

$$\frac{dL(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_k x_k = 0.$$

Hence, $\hat{\lambda}_{\text{MLE}} = \frac{n}{\sum_k x_k} = \frac{1}{\frac{1}{n}\sum_k x_k} \geq 0$.

Question 2

Consider the probability density function

$$p(x|a) = \begin{cases} \frac{2}{a^2}x, & 0 < x \leq a, \\ 0, & \text{otherwise.} \end{cases}$$

Find the ML estimators of the mean μ and variance σ^2 given a set of iid samples $D = \{x_k\}_{k=1}^{n}$.

Solution

First notice that for any $a > 0$

$$\int_0^a \frac{2}{a^2} x dx = \frac{2}{a^2} \int_0^a x dx = \frac{2}{a^2} \cdot \frac{a^2}{2} = 1,$$

which implies that $p(x|a)$ is a density function.
Next, we determine the mean and variance of this density function as a function of a:

$$\mu = E[X] = \int_0^a \frac{2}{a^2} x^2 dx = \frac{2}{3} a,$$

$$\sigma^2 = Var(X) = \int_0^a (x - \frac{2}{3}a)^2 \frac{2}{a^2} x dx = \frac{1}{18} a^2.$$

We proceed by computing the MLE of a. Note that a should satisfy

$$a \geq \max_k x_k \triangleq x_{\max}.$$

The MLE of a is given by

$$\begin{aligned}
\hat{a}_{\text{MLE}} &= \arg\max_{a \geq x_{\max}} \quad l(a) \\
&= \arg\max_{a \geq x_{\max}} \quad \log \prod_{k=1}^{n} \frac{2}{a^2} x_k \\
&= \arg\max_{a \geq x_{\max}} - n \log a \\
&= \arg\min_{a \geq x_{\max}} \log a \\
&= x_{\max}.
\end{aligned}$$

Plugging in $\hat{a}_{\text{MLE}}$, the MLE of μ and σ^2 are

$$\hat{\mu}_{\text{MLE}} = \frac{2}{3} \hat{a}_{\text{MLE}} = \frac{2}{3} x_{\max},$$

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{18} \hat{a}^2_{\text{MLE}} = \frac{1}{18} x^2_{\max}.$$

Notice that the estimator is based only on a single sample - the largest one. In addition, the last step is an example of **the invariance principle**:
Consider $f = f(\theta)$ and let $\hat{\theta}_{\text{MLE}}$ be the s the maximum likelihood estimator of θ. Then, , the maximum-likelihood estimator of f is

$$\hat{f}_{\text{MLE}} = f(\hat{\theta}_{\text{MLE}}).$$

Question 3

Consider a set of iid samples $D = \{x_i\}_{i=1}^n$ taken from Gamma distribution

$$p(x|k,\theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)},$$

where $x, k, \theta > 0$ and $\Gamma(k)$ is the gamma function.
Compute the MLE of k and θ.

Solution

The likelihood function for n iid observations is

$$L(k,\theta) = \prod_{i=1}^n \frac{x_i^{k-1}e^{-\frac{x_i}{\theta}}}{\theta^k \Gamma(k)}.$$

The log-likelihood function is

$$l(k,\theta) = (k-1)\sum_{i=1}^n \log x_i - \sum_{i=1}^n \frac{x_i}{\theta} - nk\log\theta - n\log\Gamma(k).$$

Finding the maximum with respect to θ by taking the derivative and setting it equal to zero yields the MLE of θ

$$\hat{\theta}_{\text{MLE}} = \frac{1}{kn}\sum_{i=1}^n x_i$$

Substituting this into the log-likelihood function gives

$$l(k,\hat{\theta}_{\text{MLE}}) = (k-1)\sum_{i=1}^n \log x_i - nk - nk\log\frac{\sum_i x_i}{nk} - n\log\Gamma(k).$$

Finding the maximum with respect to k by taking the derivative and setting it equal to zero yields

$$\log k - \psi(k) = \log\left(\frac{1}{n}\sum_i x_i\right) - \frac{1}{n}\sum_i \log x_i,$$

where $\psi(k) \triangleq \frac{d}{dk}\log\Gamma(k) = \frac{\Gamma'(k)}{\Gamma(k)}$. In this case, there is no closed-form solution for k but it can be solved numerically (e.g. using **gamfit** in MATLAB).