Some considerations on MLEs

By : Thibault Lahire (thibault.lahire@student.isae-supaero.fr)

We consider the following model : z and x are discrete variables taking respectively M and K different values with $p(z = m) = \pi_m$ and $p(x = k|z = m) = \theta_{mk}$. In what follows, we compute the maximum likelihood estimator for π and θ based on an i.i.d. sample of observations.

We consider that we observed the sample : $S = \{(z_1, y_1), ..., (z_n, y_n)\}$.

To make the notations clearer, we note $Z = [Z_1, ..., Z_M]$ and $X = [X_1, ..., X_K]$. We have $\mathbb{P}(Z_m = 1) = \mathbb{P}(Z = m) = \pi_m$ and $\mathbb{P}(X_k = 1|Z_m = 1) = \mathbb{P}(X = k|Z = m) = \theta_{mk}$.

We first focus on the iid sample $\{z_1, ..., z_n\}$. We compute the probability distribution $p(z|\pi)$:

$$p(z|\pi) = \prod_{m=1}^{M} \pi_m^{z_m}$$

Given the iid sample we consider, we can re-write this sample to stick to the notations we have introduced. Indeed, we now consider that $\forall m \in [\![1; M]\!]$, we work on $z_{m,1}, ..., z_{m,n}$. Hence the likelihood is :

$$l_{(z_i)_i}(\pi) = \prod_{i=1}^{n} \prod_{m=1}^{M} \pi_m^{z_{m,i}} = \prod_{m=1}^{M} \pi_m^{\sum_{i=1}^{n} z_{m,i}}$$

We can therefore compute the log-likelihood :

$$L_{(z_i)_i}(\pi) = \sum_{m=1}^{M} \log(\pi_m) \sum_{i=1}^{n} z_{m,i}$$

To find the MLE estimators of the parameters π_m, we maximize the log-likelihood. This optimization process is constrained by $\sum_{m=1}^{M} \pi_m = 1$.

We consider the function h such that :

$$h(\pi, \nu) = L_{(z_i)_i}(\pi) + \nu \left(1 - \sum_{m=1}^{M} \pi_m\right)$$

Note that $\nabla_\pi h(\pi, \nu)$ is of size M. $\forall m \in [\![1; M]\!]$, when we set the gradient to 0, we have component-wise :

$$\frac{1}{\pi_m} \sum_{i=1}^{n} z_{m,i} - \nu = 0 \qquad \longrightarrow \pi_m = \frac{1}{\nu} \sum_{i=1}^{n} z_{m,i}$$

We also have :

$$\nu = \nu \sum_{m=1}^{M} \pi_m = \sum_{m=1}^{M} \sum_{i=1}^{n} z_{m,i} = \sum_{i=1}^{n} \sum_{m=1}^{M} z_{m,i} = \sum_{i=1}^{n} 1 = n$$

If we note N_m the number of times where $Z_m = 1$ (i.e. $Z = m$) in our sample, we obtain :

$$\widehat{\pi}_m = \frac{1}{n} \sum_{i=1}^{n} z_{m,i} = \frac{N_m}{n}$$

We now introduce $\mu_{mk} = \mathbb{P}(X_k = 1, Z_m = 1)$. We have thanks to the Bayes' rule : $\mu_{mk} = \theta_{mk}\pi_m$. Hence, computing an estimator for μ_{mk} is equivalent to computing an estimator for θ_{mk}. Indeed : $\widehat{\theta}_{mk} = \widehat{\mu}_{mk}/\widehat{\pi}_m$.

We compute the joint distribution $p(x, z|\mu)$:

$$p(x, z|\mu) = \prod_{k=1}^{K}\prod_{m=1}^{M} \mu_{mk}^{z_m x_k}$$

Then the likelihood of this probability distribution is :

$$l_{(x_i)_i,(z_i)_i}(\mu) = \prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{m=1}^{M} \mu_{mk}^{z_{m,i}x_{k,i}} = \prod_{k=1}^{K}\prod_{m=1}^{M} \mu_{mk}^{\sum_{i=1}^{n} z_{m,i}x_{k,i}}$$

We derive the log-likelihood :

$$L_{(x_i)_i,(z_i)_i}(\mu) = \sum_{m=1}^{M}\sum_{k=1}^{K} \log(\mu_{mk}) \sum_{i=1}^{n} z_{m,i}x_{k,i}$$

To find the MLE estimators of the parameters μ_{mk}, we maximize the log-likelihood. This optimization process is constrained by $\sum_{m,k} \mu_{mk} = 1$. Indeed, μ defines a probability distribution. We can also see it with : $\sum_{m,k} \mu_{mk} = \sum_m \pi_m \sum_k \theta_{mk} = 1 \times 1 = 1$.

We consider the function h such that :

$$h(\mu, \nu) = L_{(x_i)_i,(z_i)_i}(\mu) + \nu\left(1 - \sum_{m=1}^{M}\sum_{k=1}^{K} \mu_{mk}\right)$$

Note that $\nabla_\mu h(\mu, \nu)$ is of size $M \times K$. $\forall (m, k)$, when we set the gradient to 0, we have component-wise :

$$\frac{1}{\mu_{mk}} \sum_{i=1}^{n} z_{m,i}x_{k,i} - \nu = 0 \qquad \longrightarrow \mu_{mk} = \frac{1}{\nu} \sum_{i=1}^{n} z_{m,i}x_{k,i}$$

We also have :

$$\nu = \nu \sum_{m,k} \mu_{mk} = \sum_{m,k}\sum_{i=1}^{n} z_{m,i}x_{m,i} = \sum_{i=1}^{n}\left(\sum_{m,k} z_{m,i}x_{m,i}\right) = \sum_{i=1}^{n} 1 = n$$

If we note N_{mk} the number of times where $Z_m = 1$ and $X_k = 1$ (i.e. $Z = m$ and $X = k$) in our sample, we obtain the MLE :

$$\widehat{\mu}_{mk} = \frac{1}{n} \sum_{i=1}^{n} z_{m,i}x_{m,i} = \frac{N_{mk}}{n}$$

Finally, thanks to Bayes' rule, we have an estimator for θ_{mk} :

$$\widehat{\theta}_{mk} = \frac{N_{mk}}{N_m}$$
