

# Properties of the MDI Estimator with linear equality constraints, and the case of separability

*Miguel Biron*

*August, 2016*

## Description of the framework

Assume we have a distribution  $p$  on a set  $\mathbf{X} \subseteq \mathbb{R}^n$ . We want to find another distribution  $q$  that solves the following Minimum Discrimination Information (MDI) estimation problem (Kullback 1959):

$$q = \arg \min_q D(q||p)$$
$$s.t \int_{\mathbf{X}} q(x) f_i(x) = m_i, i = 0...k$$

where

- $D(q||p) = \int_{\mathbf{X}} q(x) \log \frac{q(x)}{p(x)} dx$  is the Kullback - Leibler Divergence from  $q$  to  $p$
- $\{(f_i(\cdot), m_i)\}$  impose  $k + 1$  constraints on  $q$ .

Usually  $f_0(\cdot) = 1 \wedge m_0 = 1$  so as to ensure that  $q$  is a density. Note that the equality constraints are linear on  $q$ . It can be shown that in this setup, the solution is given by (Shore and Johnson 1981):

$$q(x) = p(x) \exp \left[ - \left( \lambda_0 + \sum_{i=1}^k \lambda_i f_i(x) \right) \right], \forall x \in \mathbf{X}$$

where the  $\lambda_i$  are the Lagrange multipliers that make  $q$  satisfy the given constraints.

## Estimation

It should be clear that the restrictions in the MDI estimation problem can be expressed as expectations with respect to the new measure  $q$ :

$$m_i = \int_{\mathbf{X}} q(x) f_i(x) = \mathbb{E}_q[f_i(x)], i = 0...k$$

Solving this problem as it stands is difficult because we are optimizing with respect to the measure that is used to calculate the expectation. However, this can be avoided by a change of measure:

$$\mathbb{E}_q[f_i(x)] = \int_{\mathbf{X}} q(x) f_i(x) = \int_{\mathbf{X}} p(x) \underbrace{\frac{q(x)}{p(x)}}_{\phi_q(x)} f_i(x) = \mathbb{E}_p[\phi_q(x) f_i(x)]$$

where  $\phi_q(x)$  is the Radon - Nikodym derivative of  $q$  with respect to  $p$ . From the general solution to the MDI with linear equality constraints one obtains

$$\phi_q(x) = \frac{q(x)}{p(x)} = \exp \left[ - \left( \lambda_0 + \sum_{i=1}^k \lambda_i f_i(x) \right) \right], \forall x \in \mathbf{X}$$

Hence, the  $k + 1$  restrictions on  $q$  of the MDI estimation can be rewritten as a system of  $k + 1$  equations involving the Lagrange multipliers:

$$\begin{aligned} 0 &= 1 - e^{-\lambda_0} \mathbb{E}_p \left[ \exp \left( - \sum_j \lambda_j f_j(x) \right) \right] \\ 0 &= m_i - e^{-\lambda_0} \mathbb{E}_p \left[ f_i(x) \exp \left( - \sum_j \lambda_j f_j(x) \right) \right], \forall i \in \{1, \dots, k\} \end{aligned}$$

Hence, we will seek to find the root of the function  $F : \mathbb{R}^{k+1} \rightarrow \mathbb{R}^{k+1}$ :

$$F(\lambda) = \begin{pmatrix} 1 - e^{-\lambda_0} \mathbb{E}_p \left[ \exp \left( - \sum_j \lambda_j f_j(x) \right) \right] \\ m_1 - e^{-\lambda_0} \mathbb{E}_p \left[ f_1(x) \exp \left( - \sum_j \lambda_j f_j(x) \right) \right] \\ \vdots \\ m_i - e^{-\lambda_0} \mathbb{E}_p \left[ f_i(x) \exp \left( - \sum_j \lambda_j f_j(x) \right) \right] \\ \vdots \\ m_k - e^{-\lambda_0} \mathbb{E}_p \left[ f_k(x) \exp \left( - \sum_j \lambda_j f_j(x) \right) \right] \end{pmatrix}$$

It is easy to check that the Jacobian of  $F$  is

$$J(\lambda) = e^{-\lambda_0} \mathbb{E}_p \left[ \exp \left( - \sum_j \lambda_j f_j(x) \right) M(x) \right]$$

where the matrix  $M(x)$  is defined as

$$M(x) \equiv (1, f_1(x), \dots, f_k(x))^T (1, f_1(x), \dots, f_k(x))$$

In turn, if we define the vector  $m \equiv (1, m_1, \dots, m_k)$ , we can write  $F$  as

$$F(\lambda) = m - e^{-\lambda_0} \mathbb{E}_p \left[ \exp \left( - \sum_j \lambda_j f_j(x) \right) M(x)_{:,1} \right]$$

As it was stated before, the expectations will be estimated with a MC method, that works by

- 1) Drawing a sample of size  $S$  under  $p$ ,  $\{x_s\}$ .
- 2) Using this sample, the expectation of any function  $h : \mathbf{X} \rightarrow \mathbb{R}$  can be approximated as

$$\frac{1}{S} \sum_{s=1}^S h(x_s) \xrightarrow{\mathbb{P}} \mathbb{E}_p(h(x)), S \rightarrow \infty$$

Using this approximation,  $F(\cdot)$  and  $J(\cdot)$  (together with the parameters of the MDI problem) can be plugged in a library for solving non-linear systems of equations and thus obtain the optimal  $\lambda$ .

## Separable constraints

### Independence preserving property

This section will study the behaviour of  $q$  when the constraints are ‘separable’ and the components of  $x$  are independent under  $p$ , i.e  $p(x)$  can be written as

$$p(x) = p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i)$$

The following paragraph shows the definition of ‘separable constraints’ that will be used throughout the paper.

**Definition 1.** The constraints imposed on an MDI problem are said to be separable if  $k = n$  and  $\forall i \in \{1, \dots, n\}$ :

$$f_i(x) = f_i(x_i)$$

That is, the  $i$ -th constraint function depends only on the  $i$ -th component of  $x$ .

First, we will need to prove the following lemma<sup>1</sup>:

**Lemma 1.** A distribution  $f$  over a subset  $\mathbf{X} \subseteq \mathbb{R}^n$  can be factored

$$f(x) = \prod_{i=1}^n u_i(x_i), \forall x \in \mathbf{X}$$

where  $u_i(\cdot)$  is integrable  $\forall i$ , iff the components of  $x$  are independent under  $f$ .

*Proof.* The  $\Rightarrow$  part of the equivalence is just the definition of independence.

For the  $\Leftarrow$  implication, note that  $\forall i \in \{1, \dots, n\}$  the cdf of  $x_i$  can be written as

$$\begin{aligned} F_i(t) &= \mathbb{P}(x_i \leq t) \\ &= \underbrace{\left( \prod_{j \neq i} \int_{-\infty}^{\infty} u_j(x_j) \right)}_{c_i} \int_{-\infty}^t u_i(x_i) \\ &= c_i \int_{-\infty}^t u_i(x_i) \end{aligned}$$

where  $c_i$  does not depend on  $t$ . Hence, the pdf of  $x_i$  will be equal to

$$f_i(t) = \frac{dF_i}{dt}(t) = c_i u_i(t)$$

Now, the constant  $c_i$  can be found by imposing

$$1 = c_i \int_{-\infty}^{\infty} u_i(x_i) \Rightarrow c_i = \left( \int_{-\infty}^{\infty} u_i(x_i) \right)^{-1}$$

---

<sup>1</sup>Proof adapted from <http://www.math.uah.edu/stat/dist/Joint.html>.

Also, note that

$$\begin{aligned}
1 &= \int_{\mathbf{X}} f(x) = \prod_{i=1}^n \int_{-\infty}^{\infty} u_i(x_i) = \left( \prod_{i \neq n} \underbrace{\int_{-\infty}^{\infty} u_i(x_i)}_{c_i} \right) \underbrace{\int_{-\infty}^{\infty} u_n(x_n)}_{c_n} \\
&\Rightarrow c_n = \left( \prod_{i \neq n} c_i \right)^{-1}
\end{aligned}$$

Putting everything together back in the original definition of  $f(x)$ , we have that

$$\begin{aligned}
f(x) &= \prod_{i=1}^n u_i(x_i) \\
&= \frac{f(x_n)}{c_n} \prod_{i \neq n} \frac{f(x_i)}{c_i} \\
&= f(x_n) \underbrace{\frac{1}{c_n} \frac{1}{\prod_{i \neq n} c_i}}_1 \prod_{i \neq n} f(x_i) \\
&= \prod_{i=1}^n f_i(x_i)
\end{aligned}$$

which proves the independence of the components of  $x$  under  $f$ .  $\square$

**Theorem 1.** In an MDI estimation problem, if the components of  $x$  are independent under  $p$ , and the constraints imposed are separable, then the components of  $x$  are independent under  $q$ .

*Proof.* Given the independence under  $p$  and the separability of the constraints, from the general solution of the MDI we see that  $q$  can be rewritten as

$$q(x) = c \prod_{i=1}^n p(x_i) \prod_{i=1}^n g_i(x_i), \forall x \in \mathbf{X}$$

where

$$\begin{aligned}
c &= \exp(-\lambda_0) \\
g_i(x_i) &= \exp(-\lambda_i f_i(x_i)), \forall i \in \{1, \dots, n\}
\end{aligned}$$

The above can be rewritten as

$$q(x) = \prod_{i=1}^n \underbrace{c^{1/n} p(x_i) g_i(x_i)}_{u_i(x_i)}, \forall x \in \mathbf{X}$$

In other words, the joint distribution of  $x$  under  $q$  factors into  $n$  integrable functions. Therefore, by Lemma 1, the components of  $x$  are also independent under  $q$ .  $\square$

**Example 1.** One example of MDI estimation with linear equality and separable constraints, in the field of credit risk, is the CIMDO method. In CIMDO, the restriction functions in the MDI estimation take the form

$$f_i(x) = f_i(x_i) = \mathbb{I}(x_i > \chi_i), \forall i = 1, \dots, n$$

where  $\mathbb{I}(\cdot)$  is the indicator function, and  $\chi_i$  are known constants. Hence, it can be seen that the restrictions are separable, and therefore by Theorem 1, we know that when using a prior  $p$  under which the components of  $x$  are independent, then they will also be independent under  $q$ .

## Marginals and conditional distributions

In this section we focus on discovering properties about the MDI distribution  $q$  when using an arbitrary prior.

**Lemma 2.** In an MDI with linear equality constraints, let

$$\mathcal{R}(f, m) \equiv \left\{ v \in \mathcal{P}(\mathbf{X}) : \int_{\mathbf{X}} v(x) f_i(x) = m_i, \forall i = 1, \dots, n \right\}$$

be the set of all distributions that satisfy the constraints, where  $\mathcal{P}(\mathbf{X})$  is the set of all probability measures on  $\mathbf{X}$ . Then, under separability, the distribution  $\pi$  defined by the product of marginals of each of the components of  $x$  under  $q$  satisfies

$$\pi(x) \equiv \left( \prod_{i=1}^n q(x_i) \right) \in \mathcal{R}(f, m)$$

*Proof.* To avoid confusion, we will use  $q_i(x_i)$  (we usually drop the subscript  $i$  on  $q$ ) to denote the marginal of the  $i$ -th component of  $x$  under  $q$ . Then, on the one hand:

$$\begin{aligned} \int_{\mathbf{X}} f_i(x) \prod_{j=1}^n q_j(x_j) &= \int_{\mathbf{X}} f_i(x_i) \prod_{j=1}^n q_j(x_j) \\ &= \int_{x_i} q_i(x_i) f_i(x_i) \underbrace{\int_{x_{-i}} \prod_{j \neq i} q_j(x_j)}_1 \\ &= \int_{x_i} q_i(x_i) f_i(x_i) \end{aligned}$$

But, on the other hand, since  $q$  does satisfy the  $i$ -th constraint:

$$\begin{aligned} m_i &= \int_{\mathbf{X}} f_i(x) q(x) \\ &= \int_{x_i} q_i(x_i) f_i(x_i) \underbrace{\int_{x_{-i}} q(x_{-i} | x_i)}_1 \\ &= \int_{x_i} q_i(x_i) f_i(x_i) \end{aligned}$$

We see then that  $\int_{\mathbf{X}} f_i(x) \pi(x) = m_i, \forall i$  and therefore  $\pi \in \mathcal{R}(f, m)$ . □

The following Lemma does not need separability, but since we just defined  $\mathcal{R}(f, m)$ , we might as well just put it here.

**Lemma 3.** In an MDI with linear equality constraints

$$\int_{\mathbf{X}} v(x) \log \frac{q(x)}{p(x)} = D(q\|p), \forall v \in \mathcal{R}(f, m)$$

*Proof.* We follow the proof of property 10 given by [shore1981properties]. Since

$$\log \frac{q(x)}{p(x)} = -\lambda_0 - \sum_{i=1}^n \lambda_i f_i(x)$$

Then,

$$\begin{aligned} \int_{\mathbf{X}} v(x) \log \frac{q(x)}{p(x)} &= \int_{\mathbf{X}} v(x) \left( -\lambda_0 - \sum_{i=1}^n \lambda_i f_i(x) \right) \\ &= -\lambda_0 - \sum_{i=1}^n \lambda_i m_i \\ &= -\lambda_0 - \sum_{i=1}^n \lambda_i \int_{\mathbf{X}} q(x) f_i(x) \\ &= \int_{\mathbf{X}} q(x) \left( -\lambda_0 - \sum_{i=1}^n \lambda_i f_i(x) \right) \\ &= \int_{\mathbf{X}} q(x) \log \frac{q(x)}{p(x)} \\ &= D(q\|p) \end{aligned}$$

□

Note that

$$\begin{aligned} D(q\|p) &\leq D(\pi\|p) \\ &= \int_{\mathbf{X}} \pi(x) \log \frac{\pi(x)}{p(x)} \\ &= \int_{\mathbf{X}} \pi(x) \log \frac{q(x)}{p(x)} + \int_{\mathbf{X}} \pi(x) \log \frac{\pi(x)}{q(x)} \\ &= D(q\|p) + \int_{\mathbf{X}} \pi(x) \log \frac{\pi(x)}{q(x)} \end{aligned}$$

In the following, it will be useful to note that, even without independence, we can still write  $q$  similarly as we did in the last section

$$q(x) = c p(x) \prod_{i=1}^n g_i(x_i), \forall x \in \mathbf{X}$$

where  $c$  and the  $g_i(\cdot)$  follow the same definitions given before.

**Lemma 4.** For every partition  $x = (x_1, x_2)$ , the marginal distribution of  $x_1$  will have the form

$$q(x_1) = cp(x_1) \prod_{i \in I_1} g_i(x_i) \mathbb{E}_p \left[ \prod_{j \in I_2} g_j(x_j) \middle| x_1 \right]$$

where  $I_1 = \{i \in \{1, \dots, n\} : x_i \in x_1\}$  and  $I_2 = \{1, \dots, n\} \setminus I_1$ .

*Proof.* Note that we can write  $p(x) = p(x_1)p(x_2|x_1)$ . Then,

$$\begin{aligned} q(x_1) &= \int_{x_2} q(x) \\ &= c \int_{x_2} p(x) \prod_{j=1}^n g_j(x_j) \\ &= cp(x_1) \prod_{i \in I_1} g_i(x_i) \int_{x_2} p(x_2|x_1) \prod_{j \in I_2} g_j(x_j) \\ &= cp(x_1) \prod_{i \in I_1} g_i(x_i) \mathbb{E}_p \left[ \prod_{j \in I_2} g_j(x_j) \middle| x_1 \right] \end{aligned}$$

□

**Lemma 5.** For every partition  $x = (x_1, x_2)$ , the conditional distribution of  $x_2$  given  $x_1$  will be

$$q(x_2|x_1) = p(x_2|x_1) \frac{\prod_{i \in I_2} g_i(x_i)}{\mathbb{E}_p \left[ \prod_{i \in I_2} g_i(x_i) \middle| x_1 \right]}$$

*Proof.*

$$\begin{aligned} q(x_2|x_1) &= \frac{q(x_1, x_2)}{q(x_1)} \\ &= \frac{cp(x) \prod_{i=1}^n g_i(x_i)}{cp(x_1) \prod_{i \in I_1} g_i(x_i) \mathbb{E}_p \left[ \prod_{j \in I_2} g_j(x_j) \middle| x_1 \right]} \\ &= \frac{p(x_2|x_1) \prod_{i \in I_2} g_i(x_i)}{\mathbb{E}_p \left[ \prod_{j \in I_2} g_j(x_j) \middle| x_1 \right]} \end{aligned}$$

□

## References

Kullback, Solomon. 1959. *Information Theory and Statistics*.

Segoviano, Miguel. 2006. "Consistent Information Multivariate Density Optimizing Methodology."

Shore, John, and Rodney Johnson. 1981. "Properties of Cross-Entropy Minimization." *IEEE Transactions on Information Theory* 27 (4). IEEE: 472–82.