Minimization with respect to divergences and applications

Pierre Bertrand, Michel Broniatowski, Jean-François Marcotorchino

Minimization with respect to divergences and applications

Pierre Bertrand[1][0000−0003−0838−8161], Michel Broniatowski[1,2], and Jean-François Marcotorchino[3]

[1] Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, Paris, France
[2] CNRS UMR 8001, Sorbonne Université, Paris, France
[3] Institut de statistique, Sorbonne Université, Paris, France

Abstract. We apply divergences to project a prior guess discrete probability law on pq elements towards a subspace defined by fixed margins constraints μ and ν on p and q elements respectively. We justify why the Kullback-Leibler and the Chi-square divergences are two canonical choices based on a 1991 work of Imre Csiszár. Besides we interpret the so called *indetermination* resulting from the second divergence as a construction to reduce couple matchings. Eventually, we demonstrate how both resulting probabilities arise in two information theory applications: guessing problem and task partitioning where some optimization remains to minimize a divergence projection.

Keywords: Mathematical Relational Analysis · Logical *Indetermination* · Coupling Functions · Task Partitioning Problem · Guessing Problem

1 Introduction

The present document is interested in the poorly studied *indetermination* (also called *indeterminacy*) as a natural second canonical coupling, the first one being the usual independence. We extract both notions from a projection of a prior uniform guess $\mathbb{U}$ to a subset of the set of probabilities defined by fixed margins. The problem is closely linked with contingency table adjustment exposed by Maurice Fréchet in [9] or [10] as well as by Edwards Deming and Frederick Stephan [8] or by Sir Alan G. Wilson [18] or [19]. Although we will limit ourselves to the discrete case it can be transposed in a continuous domain (see for example [16] in the same proceeding as the present paper) leading to the optimal transport theory for which an overview and recent breakthroughs are presented in [17].

When limited to the discrete case, a work of Csiszár ([7]) shows that two projection functions set apart from all the others as they are the only ones to convey "natural" properties. As we shall see, the first one leads to independence while the second one, since canonic, motivates our interest for the *indetermination* coupling it generates.

We gather here a unified construction for both couplings and show *indetermination* comes with a stunning property of couple matchings minimization that one could have expected to happen under independence. Eventually this property is applied to two concrete problems: guessing and task partitioning.

2 Projection using divergences

Consider two given discrete probabilities $\mu = \mu_0, \ldots, \mu_p$ on p elements and $\nu = \nu_0, \ldots, \nu_q$ on q elements with non zero values (it amounts to reduce p and q). We associate to them the subspace $\mathcal{L}_{\mu,\nu}$ of probability laws π on pq elements where :

$$\pi_{u,\cdot} = \sum_{v=1}^{q} \pi_{u,v} = \mu_u, \forall 1 \leq u \leq p,$$

$$\pi_{\cdot,v} = \sum_{u=1}^{p} \pi_{u,v} = \nu_v, \forall 1 \leq v \leq q,$$

namely, μ and ν are the margins of π. $\mathcal{L}_{\mu,\nu}$ is thus defined by a set of linear constraints on the components of π where π is a probability law living inside the simplex S_{pq}.

As spotted in [3], we can interpret discrete transport problems related to contingency table adjustments, as projections of a prior guess, typically the uniform law $\mathbb{U} = \frac{1}{pq}, \ldots, \frac{1}{pq}$, to our just defined subspace $\mathcal{L}_{\mu,\nu}$.

It happens that the projections of a prior guess that we fix equal to $\mathbb{U}$ to a subspace $\mathcal{L}$ of S_{pq} defined by a set of linear constraints was deeply studied in [7]. Formally, we are minimizing a divergence (see Definition 1) between $\mathbb{U}$ and any element of $\mathcal{L}$ as exposed in Problem 1.

Definition 1 (Divergence with functional parameter ϕ also called f-divergence).
Given, a positive function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$, strictly convex at 1 with $\phi(1) = 0$, and given two discrete probabilities m and n on pq elements, we define the divergence D_ϕ as:

$$D_\phi(m|n) = \sum_{u=1}^{p} \sum_{v=1}^{q} n_{u,v} \phi\left(\frac{m_{u,v}}{n_{u,v}}\right) = \mathbb{E}_n\left[\phi\left(\frac{m_{u,v}}{n_{u,v}}\right)\right].$$

Additionally we set $0 * \phi(x) = 0$ for all $x \in \mathbb{R}^+$.

Problem 1 (Generic Projection Problem).
Given a subspace $\mathcal{L}$ of S_{pq} defined by a set of linear constraints, the projection problem of the uniform law $\mathbb{U}$ to $\mathcal{L}$ using a divergence D_ϕ as defined in Definition 1, states as follows:

$$\min_{\pi \in \mathcal{L}} D_\phi\left(\pi \,|\mathbb{U}\right).$$

Remark 1.
As mentioned by an anonymous reviewer, since $\mathcal{L}$ is a $m-$affine subspace and D_ϕ is a $f-$divergence, the unicity of the optimal is stated from information geometry theory (see [14]).

A key result of the paper [7] is that, provided we add some hypotheses on the cost function we can limit ourselves to two divergences. We do not gather here those hypotheses, but one can find them obviously in the quoted paper as well as summarized in [3]. Eventually, we obtain the two eligible divergences by setting $\phi : x \mapsto x \log(x) - x + 1$ leading to the usual Kullback-Leibler divergence:

$$D_{KL}\left(\pi \,|\mathbb{U}\right) = \sum_{u=1}^{p} \sum_{v=1}^{q} \pi_{u,v} \log(pq\pi_{u,v}); \tag{1}$$

and thus to the transport problem using entropy exposed in [18] or with $\phi : x \mapsto (x-1)^2$ leading to the so-called Pearson Chi-square divergence function:

$$D_2\left(\pi \,|\mathbb{U}\right) = pq \sum_{u=1}^{p} \sum_{v=1}^{q} \left(\pi_{u,v} - \frac{1}{pq}\right)^2 \tag{2}$$

hence to the minimal transport problem exposed in [15].

Typical computations, using Karush–Kuhn–Tucker, provide for the two divergences, a closed-form expression of the projections that we summarize in the figure 1 below where $\pi^\times$ is the *independence* coupling of the two margins:

$$\pi^\times(u,v) = \mu_u \nu_v; \tag{3}$$

and π^+ is the so-called *indetermination* (or *indeterminacy*) coupling:

$$\pi^+(u,v) = \frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}. \tag{4}$$

At this point we must notice that π^+ is not always defined since it may contain negative values. We add a hypothesis on the margins μ and ν to get rid of this problem:

$$p \min_u \mu_u + q \min_v \nu_v \geq 1. \tag{5}$$

Regarding Equation 5, a typical modification of any couple of margins to deduce eligible margins can be found in [3] while the probability to draw adapted margins if μ and ν are uniformly drawn on S_p and S_q respectively is provided in [1].

Using Mathematical Relational Analysis notations, one can show that π^+ leads to an equality between a "for" vs "against" notion which justifies the name *indetermination* historically coined in [12]. Besides, the $+$ notation we chose in [2] comes from the Full-Monge property (see for instance the survey [6]) that the associated contingency matrix respects. Eventually, we can also extend the notion of *indeterminacy* in a continuous space and estimate an associated copula as exposed in [4]. In the present paper we shall not detail further the properties π^+ conveys and refer the interested reader to the aforementioned articles.
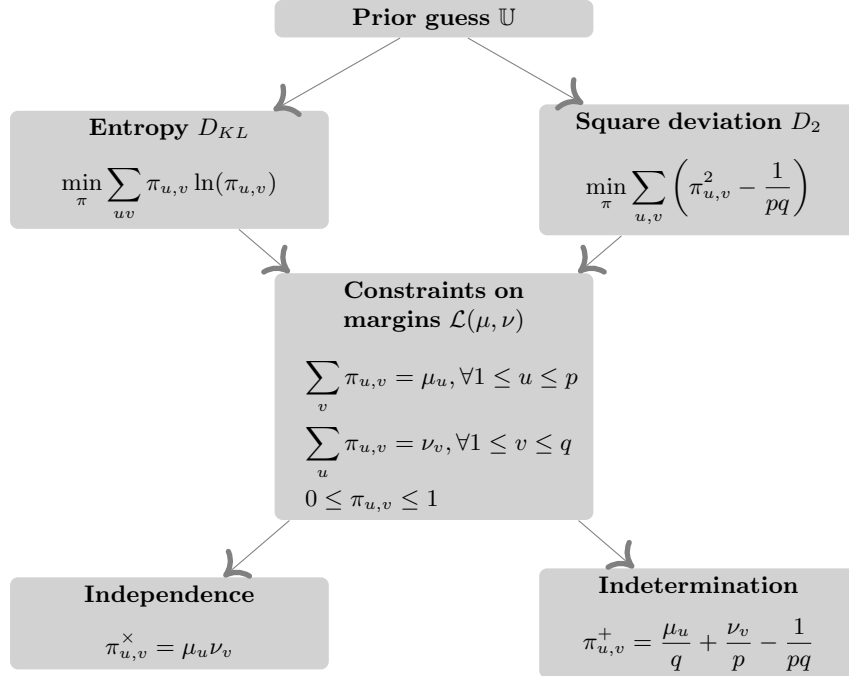

Fig. 1. Summary of a symmetric construction

3 Mimisation of couple matchings

We insist on a property expressed at the very beginning of the *indetermination* construction. Let us say π is a probability law on pq elements. If we independently draw two occurrences of $W = (U, V) \sim \pi$, the probability of getting a matching for a couple, that is to say $U_1 = U_2$ and $V_1 = V_2$ simultaneously, is given by $\pi_{U_1,V_1} \times \pi_{U_2,V_2} = \pi_{U_1,V_1}^2$.

Besides, it turns out that

$$D_2\left(\pi \,|\mathbb{U}\right) = pq \sum_{u=1}^{p} \sum_{v=1}^{q} \left(\pi_{u,v} - \frac{1}{pq}\right)^2 \tag{6}$$

$$= pq \sum_{u=1}^{p} \sum_{v=1}^{q} (\pi_{u,v})^2 - 1 \tag{7}$$

so that the shift introduced by $\mathbb{U}$ vanishes and minimizing $D_2(\pi \,|\mathbb{U})$ amounts to minimize the probability of a couple matching.

This property is illustrated by a decomposition of π^+ whose proof is available in [3] and which leads to a method for drawing under π^+. We first set u_0 such that it corresponds to the smallest line meaning μ_{u0} is minimal. Then, quoting $\forall u, \ \delta_u = \mu_u - \mu_{u_0}$, we can show that the following drawing realizes *indetermination*:

1. We draw u through μ.
2. We roll a loaded dice with the Bernoulli's skew: $I = \mathrm{Be}\left(\frac{\delta_u}{\mu_u}\right)$.
3. If we get 0, we draw v under the distribution $\left(\pi^+_{u_0,1}, \ldots, \pi^+_{u_0,q}\right)$ that we normalize to 1, else, we draw v uniformly hence under $\left(\frac{1}{q}, \ldots, \frac{1}{q}\right)$.

The method this coupling uses to minimize couple matchings reads as follows. If U_1 is frequent then μ_{U_1} is high, hence δ_{U_1} is high, this means that I will often draw 1 and that v will often be uniform preventing $V_1 = V_2$ as much as possible. On the opposite, if U_1 is rare (typically $U_1 = u_0$) then $U_1 = U_2$ is already rare and I will be allowed to often draw 0 so that v will follow $\left(\pi^+_{u_0,v}\right)_v$.

Best option of course would be to stay uniform every time, this is the prior guess $\mathbb{U}$ solution but it does not belong to $\mathcal{L}_{\mu,\nu}$. The trick with indetermination is that it hides the disequilibrium on v that is needed to respect margins inside rare lines.

A stunning result is that couple reduction is not ensured by independence but by indetermination; though the difference between both is small. One can show that in the discrete case, the expected $\mathbb{L}^2$-difference is $\frac{1}{pq}$ (see [2]). Furthermore the probability of a couple matching with a variable under a second law π' only depends on its margins as stated in Proposition 1.

Proposition 1 (Matchings with another variable).
Let us suppose that π' is a second probability law on pq elements with margins μ' and ν'. We give ourselves a first couple $W = (U_1, U_2)$ drawn under π^+ and a second one $W' = (U'_1, U'_2)$ independently drawn under π'. Then the probability of couple matching only depends on the margins and is given by:

$$\mathbb{P}_{\pi^+ \times \pi'}(W = W') = \sum_{u=1}^{p}\left(\frac{\mu_u \mu'_u}{q}\right) + \sum_{v=1}^{q}\left(\frac{\nu_v \nu'_v}{p}\right) - \frac{1}{pq}. \tag{8}$$

Proof. The result is due to the structure of π^+. Formally:

$$\mathbb{P}_{\pi^+ \times \pi'}(W = W') = \sum_{(u,v)} \pi^+_{u,v} \pi'_{u,v}$$

$$= \sum_{(u,v)} \left(\frac{\mu_u}{q} + \frac{\nu_v}{p} - \frac{1}{pq}\right) \pi'_{u,v}$$

$$= \sum_{u=1}^{p}\left(\frac{\mu_u \mu'_u}{q}\right) + \sum_{v=1}^{q}\left(\frac{\nu_v \nu'_v}{p}\right) - \frac{1}{pq}$$

In particular, the last proposition 1 shows that any $\pi' \in \mathcal{L}_{\mu,\nu}$ has the same probability of a couple matching with π^+ than π^+ itself. The next section leverages on that ability to reduce couple matching in two applications.

4 Applications

4.1 Guessing problem

In cryptography, a message u in a finite alphabet $\mathcal{U}$ of size p is typically sent from Alice to Bob while a spy whose name is Charlie tries to decode it. A common strategy for Alice to communicate efficiently and secretly

with Bob consists in encoding the message using a couple of keys (public, private) for each character or a symmetric encryption which only requires one shared key between Alice and Bob. The literature concerned with the encryption method to choose according to the situation is diverse, the most-used standard is Advanced Encryption Standard described in various articles. Possibly, Charlie observes an encrypted message V in a second finite alphabet $\mathcal{V}$ of size q which is a function of the message u. We choose to gather in V any additional data Charlie may observe (the channel used, the sender's location, . . .). In the end U and V are correlated but not always in a deterministic link.

Related to the cryptography situation, the guessing problem quoted hereafter as Problem 2 was first introduced in the article [13] and extended in [11] to take V into account. While in cryptography Charlie tries to decode a sequence of messages, the guessing problem focuses on decoding a unique message.

Let us formalize the situation: Charlie chooses its strategy (see Definition 2) according to the value taken by the observed second variable V: he typically adapts himself to the conveyed encryption. The probability law of the couple (U, V) is quoted $\mathbb{P}_{U,V} = \pi$ while its margins are $\mathbb{P}_U = \mu$ and $\mathbb{P}_V = \nu$ with no zero values.

Problem 2 (Original Guessing Problem or Spy Problem).

When Alice sends a message $U = u$ to Bob, the spy Charlie observes a variable $V = v$ and must find out the value u of the realization. He has access to a sequence of formatted questions for any guess $\tilde{u}$ he may have: "Does u equal $\tilde{u}$?" for which the binary answer is limited to "yes/no".

Definition 2 (Strategy).
For any value v, a strategy $\sigma_v = S|v$ of Charlie is defined by an order on $\mathcal{U}$ representing the first try, the second and so on until number p. It can be deterministic or random: we quote $\mathbb{P}_{S|v}$ its probability law. The global strategy quoted σ or S gathers all the marginal strategies S_v and is a random variable dependent only on V (not U) whose probability law is quoted $\mathbb{P}_S$.

Besides, for a given position $i \in [1, p]$, and an order σ, $\sigma[i]$ is the element in $\mathcal{U}$ corresponding to the i-th try.

The usual performance $G(\sigma, u)$ of a spy consists in counting the number of questions required to eventually obtain a "yes" in Problem 2 when Charlie proposed the order $S = \sigma$ and Alice generated the message $U = u$. For instance, with an alphabet $\mathcal{U} = \{a, b, c, d\}$, if the message is $u = c$ and the strategy σ of the spy consists in the order (b, c, a, d) (meaning he first proposes message b then c,. . .) we have:

$$
\begin{aligned}
G(\sigma, u) &= \sum_{i=1}^{p} i \mathbf{1}_{u=\sigma[i]} \\
&= 1 \cdot \mathbf{1}_{u=b} + 2 \cdot \mathbf{1}_{u=c} + 3 \cdot \mathbf{1}_{u=a} + 4 \cdot \mathbf{1}_{u=d} \\
&= 2 \cdot \mathbf{1}_{u=c} \\
&= 2
\end{aligned}
$$

Eventually in [13], they compute the $\rho-$moment of G as an estimation of the efficiency of a strategy S:

$$
||G(S, U)||_{\rho} = \left[\mathbb{E}_{(S,U,V)\sim\mathbb{P}_{S,U,V}} \left(G(S, U)^{\rho} \right) \right]. \tag{9}
$$

We notice that G actually takes as granted that Charlie has up to p trials which is not adapted in case Alice sends a sequence of messages due to a lack of time. Let us move away from the literature and measure Charlie's performance by its probability to find out after one trial the message u Alice sent. It is a reasonable measure as, if a sequence of messages is sent, he may have to jump from one to the following after only one trial. What actually matters (as mentioned by an anonymous reviewer) is the number of attempts k Charlie can try between two messages from Alice. It could be captured with the one-shot performance M using a trick and supposing that Alice repeats each message k times.

Definition 3 (one-shot performance).
For a given strategy S (which can depend on V with a non deterministic link), we define the following performance measure as the probability to find out the value u after one trial, formally:

$$
M(S, U, V) = M(S, \pi) = \mathbb{P}_{(S,U,V)} (S[1] = U).
$$

We suppose as for the original optimal strategy that the spy has access to the distribution $\pi = \mathbb{P}_{U,V}$. We can imagine he previously observed the non-encrypted messages in a preliminary step.

Two strategies immediately stand out:

1. S_{max}: systematically returns at v fixed (observed by hypothesis), the u associated with the maximal probability on the margin $\mathbb{P}_{U|V=v}$;
2. S_{margin}: returns at v fixed a random realization in $\mathcal{U}$ under the law $\mathbb{P}_{U|V=v}$.

Similarly as in see [13] where they prove S_{max} is the best strategy in case the performance measure is G, we can show it also maximizes the one-shot performance. Actually, M rewrites:

$$M(S,U,V) = \sum_{v=1}^{q} \nu_v \mathbb{P}_{(S_v,U) \sim \mathbb{P}_{S_v} \otimes \mathbb{P}_{U|V=v}} (S_v[1] = U) \tag{10}$$

$$= \sum_{u=1}^{p} \sum_{v=1}^{q} \pi_{u,v} \mathbb{P}_{S_v} (S_v[1] = u). \tag{11}$$

Since $\mathbb{P}_{S_v}$ only depends on v, M is maximal under S_{max}, when $\mathbb{P}_{S_v} = \delta_{u_v}$ where $u_v = \text{argmax}_u \pi_{u,v}$ so that:

$$M(S,U,V) \geq M(S_{max},U,V) = \sum_{v=1}^{q} \pi_{u_v,v}. \tag{12}$$

Eventually, we quote $u_1 = \text{argmax}_u \mu_u$ and notice that $\sum_{v=1}^{q} \pi_{u_v,v} \geq \mu_{u_1}$ leading to the proposition 2.

Proposition 2 (Charlie's best performance).
We suppose that the margins μ and ν are fixed. Then, for any coupling probability π between messages U and ciphers V, the best one-shot performance Charlie can perform always happens under S_{max}. Furthermore it admits a fixed lower-bound μ_{u_1} independent on π; to summarize:

$$M(S,U,V) \geq M(S_{max},U,V) = \sum_{v=1}^{q} \pi_{u_v,v} \geq \mu_{u_1}. \tag{13}$$

Let us suppose, commendable task if any, that Alice wants to minimize Charlie's best one-shot performance. We also suppose that the margins μ on U and ν on V are fixed. It is a common hypothesis: the alphabet $\mathcal{U}$ in which the messages are composed typically respects a distribution on letters; variable V on its own, if it represents frequencies for instance may have to satisfy occupation weights on each channel. Eventually, Alice can only leverage on the coupling between U and V.

Precisely, let us now compute the corresponding value for our two couplings, both are optimal (for Alice).

$$M^{\times} = M(S_{max}, \pi^{\times}) = \mu_{u_1} \qquad M^{+} = M(S_{max}, \pi^{+}) = \mu_{u_1}$$

Regarding the second strategy S_{margin} we know it is less efficient for Charlie in term of one-shot performance. Yet, it is by far harder to cope with for the sender who cannot easily prevent random conflicts. Consequently we come back to the reduction of couple matchings (here a success for Charlie), whose "*indetermination* coupling", we know, prevents us against. Let us unfold this remark hereafter.

Replacing $\mathbb{P}_{S_v}$ by its value under the second strategy in Equation 11 allows us to estimate the one-shot performance of S_{margin} which is given by:

$$M(S_{margin}, U, V) = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \nu_v (\pi_{u|V=v})^2. \tag{14}$$

Concerning the strategy S_{margin} we have the two bounds:

$$\frac{||\pi||_2^2}{\min_{v \in \mathcal{V}} \nu_v} \geq M(S_{margin}, U, V) \geq \frac{||\pi||_2^2}{\max_{v \in \mathcal{V}} \nu_v}, \tag{15}$$

with

$$||\pi||_2 = \sqrt{\sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} (\pi_{u,v})^2}.$$

Depending on ν the bounds are more or less tight. If ν_v goes to the uniform, the situation becomes similar to S_{max}: both couplings are equal and optimal.

In any case, Equation 15 shows that studying the guessing problem brings us back to Problem 1 associated with cost D_2 whose solution is given by the *indetermination* coupling of the margins. It guarantees an efficient reduction of conflicts (see section 3) and eventually a weak one-shot performance under S_{margin} as expressed in Equation (15) and an optimal one under S_{max}.

4.2 Task Partitioning

Task partitioning problem is originally introduced in [5] where the authors provide a lower bound on the moment of the number of tasks to perform. Let us follow the gathering work of [11] where they also coin a generalized task partitioning problem basically adapting it as a special case of the guessing problem.

Formally, we begin with the original problem of tasks partitioning: a finite set $\mathcal{U}$ of p tasks is given together with an integer $q \leq p$. The problem consists in creating a partition $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_q)$ of $\mathcal{U}$ in q classes to minimize the number of tasks to perform, knowing that if one needs to perform a task $u \in \mathcal{A}_i$, it is mandatory to launch simultaneously the whole subset of tasks included within $\mathcal{A}_i$.

Practically, a task $U = u$ to perform is randomly drawn from $\mathcal{U}$ under a probability distribution $\mathbb{P}_U = \mu$ representing the tasks frequencies. As any task, the task u to perform is assigned to a unique class $\mathcal{A}_{i(u)}$ of the arbitrary partition. Hence, $A(u) = |\mathcal{A}_{i(u)}|$ counts the number of tasks to perform. Precisely, one plays on the partition knowledge to perform, in average, as few tasks as possible.

Similarly to the guessing problem, the performance of a partition $\mathcal{A}$ is estimated through the $\rho-$moment of the number of tasks $A(U)$ to perform:

$$\mathbb{E}_{U \sim \mathbb{P}_U} \left[A(U)^\rho \right]. \tag{16}$$

Inspired by the general guessing problem, paper [11] extends the task partitioning problem. Let us introduce here this generalized version, in which we are no longer interested in minimizing the number of tasks to perform but rather in reducing the number of tasks before a selected (or a chosen) task u.

Indeed, in the first version, as soon as u is drawn, an arbitrary rule imposes to perform the whole subset $\mathcal{A}_{i(u)}$ leading to realize $A(u)$ tasks. In the new version, tasks are performed sequentially in $\mathcal{A}_{i(u)}$ according to a global strategy S that can be deterministic or random. Typically, tasks may consist in a flow of signatures which an administration requires while q would be the number of workers dedicated to perform those signatures on incoming documents. A worker can be given the entitlement to perform several signatures, assistants usually do. In that case, the partition encodes the assignments of tasks to workers. When a worker $V = v$ is assigned a document, the depositor waits until the signature. At the same time the worker follows his own strategy S_v to sign his assigned documents, meaning he can always follow the same order leading to a deterministic strategy or change every day leading to a random strategy.

With a global strategy S which gathers the workers' strategies S_v, $\forall 1 \leq v \leq q$ and for a task u to perform, the performance of a partition $\mathcal{A}$ is measured using

$$N_{S,\mathcal{A}}(u) = j \ / \ S_{i(u)}[j] = u,$$

which represents the number of tasks performed before the intended task u (u included) by the worker $i[u]$ it is affected to according to partition $\mathcal{A}$. A lower bound is provided in the paper [11].

Let us now suppose that the keys $1 \leq v \leq q$ are associated with offices that must perform a proportion ν_v of the incoming tasks which still follow a distribution μ. It actually appears as a sensible problem where a manager would have to distribute in advance tasks among teams according to the usual observed distribution of tasks and a list of available teams with their capacities.

Besides, we suppose that each team uses the strategy S_{marge} to perform tasks, meaning they randomly perform one according to their margin theoretical distribution; for a document signing, they randomly sign one.

From now on, we can rewrite our task partitioning problem under the form of a guessing problem:

- $V = v$, formerly corresponds to a worker, now it represents the information the spy has access to;
- $U = u$, formerly represents a task to perform, now it represents a sent message;
- $S = \sigma$, formerly represents the order in which tasks are performed, now it represents the order in which Charlie proposes his guesses.

Under this formalism, we are interested in measuring the probability $M(S, U, V)$ of executing u first as an extended application of the one-shot performance of Definition 3 and we have:

$$M(S,U,V) \geq \frac{||\pi||_2^2}{\max_{1 \leq v \leq q} \nu_v} \geq \frac{||\pi^+||_2^2}{\max_{1 \leq v \leq q} \nu_v}. \tag{17}$$

This inequality provides a lower bound for any distribution of the tasks among the team, no distribution can generate a worst "one-shot probability" of satisfying the intended task. In task partitioning actually, each task u is uniquely associated to a worker $v = i(u)$ so that the random variable representing the worker is deterministic conditionally to U: $\pi_{u,v} = \delta_{v=i(u)}$.

References

1. Bertrand, P.: Transport Optimal, Matrices de Monge et Pont relationnel. Ph.D. thesis, Paris 6 (to be defended 2021)
2. Bertrand, P., Broniatowski, M., Marcotorchino, J.F.: Independence versus Indetermination: basis of two canonical clustering criteria (Jul 2020), https://hal.archives-ouvertes.fr/hal-02901167, working paper or preprint
3. Bertrand, P., Broniatowski, M., Marcotorchino, J.F.: Logical indetermination coupling:a method to minimize drawing matches and its applications (Dec 2020), https://hal.archives-ouvertes.fr/hal-03086553, working paper or preprint
4. Bertrand, P., Broniatowski, M., Marcotorchino, J.F.: Continuous indetermination and average likelihood minimization (May 2021), https://hal.archives-ouvertes.fr/hal-03215096, working paper or preprint
5. Bunte, C., Lapidoth, A.: Encoding tasks and Rényi entropy. IEEE Transactions on Information Theory **60**(9), 5065–5076 (Sep 2014). https://doi.org/10.1109/tit.2014.2329490, http://dx.doi.org/10.1109/TIT.2014.2329490
6. Burkard, R.E., Klinz, B., Rudolf, R.: Perspectives of Monge properties in optimization. Discrete Applied Mathematics **70**, 95–161 (1996)
7. Csiszár, I., et al.: Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. The annals of statistics **19**(4), 2032–2066 (1991)
8. Deming, W.E., Stephan, F.F.: On the least squares adjustment of a sampled frequency table when the expected marginal totals are known. the Annals of Mathematical Statistics **11**, 427–444 (1940)
9. Fréchet, M.: Sur les tableaux de corrélations dont les marges sont données. Annales de l'Université de Lyon, Section. A **14**, 53–77 (1951)
10. Fréchet, M.: Sur les tableaux de corrélations dont les marges et les bornes sont données. Revue de l'Institut de Statistique **28**, 10–32 (1960)
11. Kumar, M.A., Sunny, A., Thakre, A., Kumar, A.: A unified framework for problems on guessing, source coding and task partitioning (2019)
12. Marcotorchino, J.F.: Utilisation des comparaisons par paires en statistique des contingences. Publication du Centre Scientifique IBM de Paris et Cahiers du Séminaire Analyse des Données et Processus Stochastiques Université Libre de Bruxelles pp. 1–57 (1984)
13. Massey, J.L.: Guessing and entropy. IEEE Int. Symp. on Info. Th. p. 204 (1994)
14. Nielsen, F.: What is an information projection. Notices of the AMS **65**(3), 321–324 (2018)
15. Stemmelen, E.: Tableaux d'échanges, description et prévision. Cahiers du Bureau Universitaire de Recherche Opérationnelle **28** (1977)
16. Stummer, W.: Optimal transport with some directed distances. In: International Conference on Geometric Science of Information. Springer (2021)
17. Villani, C.: Optimal Transport, Old and New. Springer –Verlag (2000)
18. Wilson, A.G.: A statistical theory of spatial distribution models. Transportation Research **1**, 253–269 (1967)
19. Wilson, A.G.: Entropy in urban and regional modelling, vol. 1. Routledge (2011)