Laplace transformation and weak convergence with an application to Fluorescence Resonance Energy Transfer (FRET)

Giacomo Aletti

Dipartimento di Matematica, Università di Milano, Via Saldini, 50, 20133 Milan, Italy

Received 9 November 2005; accepted 21 November 2005

Abstract

A new distance $d_A$ on Laplace transformations of probability distributions on $\mathbb{R}_+$ is given. The new complete metric space is shown to be homeomorphic to the topological probability distributions space on $\mathbb{R}_+$ with weak convergence. A new estimator for the fluorescence source distribution is described.

Keywords: Laplace transformation; Weak convergence; Metrics on probability spaces; Fluorescence microscopy

1. Introduction

The study of cellular processes involves experiments with large numbers of cells, where the protein content is subsequently isolated and studied using antibodies to detect changes in protein levels, post-translational modifications, pairing with partner molecules, etc. Although extremely informative in many cases, these mass population analyses often lack the time resolution for studying rapid alterations in protein state, and do not allow the characterization of highly dynamic processes. Moreover, analysis of millions of cells at once obviously shows the average response in the population of cells, thereby obscuring cell-to-cell variation and the dynamic range of a process [4]. Finally, subcellular compartmentalization of reactions is difficult to assess in these whole-cell approaches.

With the availability of microscopic techniques in combination with genetically encoded fluorescent probes, many of the above-described restraints have been overcome. Highly dynamic reactions can now be studied in detail in a relatively easy manner, and in the context of a living cell.

One of these techniques is Fluorescence Resonance Energy Transfer, or FRET (see [5]), a physical phenomenon already described many years ago. FRET is the radiationless transfer of energy from an excited donor fluorophore to a suitable acceptor fluorophore, a physical process that depends on spectral overlap and proper dipole alignment of the two fluorophores. Whereas normally an excited fluorophore returns to the ground state with the emission of a photon, FRET results in the excitation of the nearby acceptor fluorophore that in turn emits a photon when it returns to the ground state. The occurrence of FRET is characterized by a decrease in observed donor emission, and a simultaneously sensitized (increased) acceptor emission.

E-mail address: giacomo.aletti@unimi.it.

0893-9659/S - see front matter © 2005 Elsevier Ltd. All rights reserved.
Fluorescence is monitored by taking images at different times $0 = t_0 < t_1 < \cdots < t_n$. Each image (or single pixel) has a very large number of overlapping exponentially decaying sources. What we observe (intensity) in the $i$-th picture, $I(i)$, is proportional to the number of decayed sources between $t_{i-1}$ and $t_i$. If all the sources have the same exponentially decaying coefficient $r$, then $I(i) \propto \exp(-r(t_{i-1} - t_i))$. Therefore, least squares methods are widely implemented in FRET software to investigate the exponential source coefficient. Sometimes, one is forced to assume that more than one exponential source is involved in our experiment. Least squares methods on mixtures of exponential densities are performed to compute the coefficients. Unfortunately, the least squares method is not equivalent to probability convergence. As a consequence, we may wrongly estimate the distribution of sources.

In order to avoid such an instability, we give here a new “distance” $d_A$ on the probability measure on $\mathbb{R}_+$ which is equivalent to probability convergence.

Let $\mu_R$ be a probability distribution on $\mathbb{R}_+$. We may think of $\mu_R$ as the distribution of the intensities of the rates $R$ of exponentially decaying particles. If we observe the survival function of $T$, the time when the particles decay, we have

$$S(t) := P(T > t) = \int_{\mathbb{R}_+} P(T > t \mid R = r) \mu_R(dr) = \int_{\mathbb{R}_+} \exp(-rt) \mu_R(dr).$$

where $T(\mu_R)$ is the Laplace transformation of $\mu_R(dr)$: our observed data. Given $S(t)$ and a sequence of probability measures $\mu_n$ on $\mathbb{R}_+$ (approximated solutions), we will find that

$$d_A(T(\mu_n)(t), S(t)) \to 0 \iff \mu_n \to \mu_R.$$

2. Levy metric and Laplace transform

Let $\mathcal{P}(\mathbb{R}_+)$ be the set of probability measures on $\mathbb{R}_+$. In what follows, $\mu_F$ will denote the distribution with cumulative distribution function (c.d.f.) $F$. We denote by $(\mathcal{P}(\mathbb{R}_+), \rightarrow)$ the topological space induced by weak convergence. When $d$ is a metric on a space $X$, we denote by $(X, d)$ the metric space. We denote by $d_L$ the Levy metric on $\mathcal{P}(\mathbb{R})$: if $F$ and $G$ are cumulative distribution functions,

$$d_L(F, G) = \inf \left\{ \varepsilon : F(x) \leq G(x + \varepsilon) + \varepsilon, \forall x \in \mathbb{R} \right\}.$$

Now, let $\mathcal{T}(\mathcal{P}(\mathbb{R}_+))$ be the set of the Laplace transformations of $\mathcal{P}(\mathbb{R}_+)$. The domain of $f \in \mathcal{T}(\mathcal{P}(\mathbb{R}_+))$ is restricted to $\mathbb{R}_+$:

$$f \in \mathcal{T}(\mathcal{P}(\mathbb{R}_+)) \iff \exists \mu \in \mathcal{P}(\mathbb{R}_+) : f(s) = \int_{\mathbb{R}_+} \exp(-rs) \mu(dr), \quad s \geq 0.$$

On the set of nonnegative monotone functions on $\mathbb{R}_+$ with values in $[0, 1]$, we define

$$d_A(f, g) = \inf \left\{ \varepsilon > 0 : \begin{array}{ll}
& f(t) - (g(t) + \varepsilon) \exp(\varepsilon t) \leq 0, \forall t \in \mathbb{R}_+, \\
& g(t) - (f(t) + \varepsilon) \exp(\varepsilon t) \leq 0
\end{array} \right\}.$$  

Clearly, $\varepsilon \leq 1$. If $\mu \in \mathcal{P}(\mathbb{R}_+)$, $T(\mu)$ is monotone with values in $[0, 1]$. We define

$$d_A(\mu_1, \mu_2) := d_A(T(\mu_1), T(\mu_2)), \quad \mu_1, \mu_2 \in \mathbb{R}_+.$$

Note that

$$\sup_{t \in [0, 1]} \left| T(\mu_1)(t) - T(\mu_2)(t) \right| \leq 5d_A(\mu_1, \mu_2) \tag{2}$$
since \( \exp(x) \leq 1 + 2x, \forall x \in [0, 1] \). Moreover, if \( d_A(f, g) \leq \varepsilon \), then

\[
\begin{cases}
 f(t) - g(t) \leq g(t)(\exp(t \varepsilon) - 1) + \varepsilon \exp(t \varepsilon) \\
 g(t) - f(t) \leq f(t)(\exp(t \varepsilon) - 1) + \varepsilon \exp(t \varepsilon)
\end{cases}
\]

\[
|g(t) - f(t)| \leq \max(g(t), f(t)) (\exp(t \varepsilon) - 1) + \varepsilon \exp(t \varepsilon) \quad \forall t \varepsilon \leq 1.
\] (3)

An analogue of (2) is widely studied in [3], but the metric it produces is not equivalent to distribution convergence (it depends on the concentration function of the distributions). It is well known that \( (P(\mathbb{R}_+), \rightarrow) = (P(\mathbb{R}_+), d_L) \) as topological spaces (see [2]), i.e.

\[
\mu_{F_n} \rightarrow_{n \rightarrow \infty} \mu_F \iff d_L(\mu_F, \mu_{F_n}) \rightarrow_{n \rightarrow \infty} 0.
\]

**Proposition 1.** \( (P(\mathbb{R}_+), d_A) \) is a complete metric space.

**Proof.** Let \( f, g, h \in T(P(\mathbb{R}_+)), \) i.e. \( f = T(\mu_F) \) and \( g = T(\mu_G) \) and \( g = T(\mu_H) \). We have

- \( d(f, g) \geq 0 \), since \( f, g \geq 0 \), \( d(f, g) = 0 \) implies \( f = g \) and hence \( \mu_F = \mu_G \).
- \( d(f, g) = d(g, f) \) by definition.
- Let \( d(f, g) \leq \varepsilon \) and \( d(g, h) \leq \varepsilon' \) and let \( t \in \mathbb{R}_+ \). We have

\[
f(t) - (h(t) + \varepsilon + \varepsilon') \exp(t(\varepsilon + \varepsilon')) \leq f(t) - (g(t) + \varepsilon' \exp(t \varepsilon) + \varepsilon') \exp(t \varepsilon)
\]

\[
\leq f(t) - (g(t) + \varepsilon \exp(t \varepsilon)) \quad \geq g(t), \text{ since } d(g, h) \leq \varepsilon'
\]

\[
\leq 0.
\]

The same computation gives \( h(t) - (f(t) + \varepsilon + \varepsilon') \exp(t(\varepsilon + \varepsilon')) \leq 0 \). Hence \( d_A(f, h) \leq d_A(f, g) + d_A(g, h) \).

Therefore, \( (P(\mathbb{R}_+), d_A) \) is a metric space. Now, take a Cauchy sequence \( (\mu_{F_n})_n \) in \( (P(\mathbb{R}_+), d_A) \):

\[
\lim_{\min(n, n) \rightarrow \infty} d_A(\mu_{F_n}, \mu_{F_{n'}}) = 0.
\]

By (3), for any \( t \in \mathbb{R}_+ \), \( T(\mu_{F_n})(t) \) is a Cauchy sequence in \([0, 1]\), and hence \( T(\mu_{F_n}) \) converges on \( \mathbb{R}_+ \) to a limit function \( f \). By (2), \( (T(\mu_{F_n}))_n \) is uniformly convergent on \([0, 1]\). Moreover \( T(\mu_{F_n}) (0) = 1 \forall n \). Thus \( f(0) = 1 \) and \( f \) is continuous in 0, which ensures that \( f = T(\mu_F) \), for some \( \mu_F \in P(\mathbb{R}_+) \).

**Theorem 2.** \( d_A \) metrizes weak convergence on \( P(\mathbb{R}_+) \): if \( \mu_F, \mu_{F_1}, \mu_{F_2}, \ldots, \in P(\mathbb{R}_+) \), then

\[
\mu_{F_n} \rightarrow_{n \rightarrow \infty} \mu_F \iff d_A(\mu_F, \mu_{F_n}) \rightarrow_{n \rightarrow \infty} 0.
\]

Moreover, \( d_A(\mu_F, \mu_G) \leq d_L(\mu_F, \mu_G), \forall \mu_F, \mu_G \in P(\mathbb{R}_+) \). Thus, the inclusion

\[
(P(\mathbb{R}_+), d_L) \hookrightarrow (P(\mathbb{R}_+), d_A)
\]

is a bijective, uniformly continuous function with continuous inverse function (it is an homeomorphism).

**Proof.** First, let us prove that \( d_A(\mu_F, \mu_G) \leq d_L(\mu_F, \mu_G) \). Let \( F, G \) with \( F(x) \leq G(x + \varepsilon) + \varepsilon, \forall x \in \mathbb{R} \). Since \( F(0^-) = G(0^-) = 0 \), we have \( F(x) - (G(x + \varepsilon) + \varepsilon \mathbb{1}_{x \in [0, \infty)}(x)) \leq 0, \forall x \in \mathbb{R} \). The Laplace transform’s properties (see [11]) give

\[
0 \leq T(F - (G_\varepsilon + \varepsilon \mathbb{1}_{x \in [0, \infty)})) = T(\mu_F - (\mu_{G_\varepsilon + \varepsilon \mathbb{1}_{x \in [0, \infty)})) = t(T(\mu_F) - (\mu_{G_\varepsilon + \varepsilon \delta_{x \in [0, \infty)})) = t(T(\mu_F) - (\exp(t \varepsilon)T(\mu_G) + \varepsilon \exp(t \varepsilon))).
\]
If \( f = T(\mu_F) \) and \( g = T(\mu_G) \), we have obtained that \( f(t) \leq (g(t) + \varepsilon) \exp(t\varepsilon) \), \( \forall t \geq 0 \). The same calculation yields \( g(t) \leq (f(t) + \varepsilon) \exp(t\varepsilon) \), \( \forall t \geq 0 \), and hence the assertion is proved. As a consequence, \( i \) is a bijective, uniformly continuous function and \( \mu_{F_n} \rightarrow_{n \rightarrow \infty} \mu_F \implies d_A(\mu_F, \mu_{F_n}) \rightarrow_{n \rightarrow \infty} 0 \).

Now, let \( G(x) = \lim \inf_{n} F_{h_n}(x) \) for a subsequence \((h_n)_n\). We have

\[
\sup_{t \in [0,1]} |T(\mu_F)(t) - T(\mu_{F_n})(t)| \leq 5d_A(\mu_F, \mu_{F_n}) \rightarrow_{n \rightarrow \infty} 0
\]

which ensures that \((F_{h_n})_n\) is tight, since its Laplace transformation converges uniformly on \([0, 1]\), and hence \( \mu_G \in \mathcal{P}(\mathbb{R}_+) \). By the triangle inequality, \( \mu_F = \mu_G \). Since \( \lim \inf_{n} F_{h_n}(x) = F(x) \) for any subsequence \((h_n)_n\), we have \( \mu_{F_n} \rightarrow_{n \rightarrow \infty} \mu_F \). \( \square \)

3. Uniform continuity

The main question related to the Theorem 2 is the following:

**Problem 3.** Is there a (strictly increasing?) function \( f : [0, 1] \rightarrow [0, 1] \) s.t.
- \( f(t) = 0 \iff t = 0; \)
- \( d_A(\mu_F, \mu_G) \geq f(d_L(\mu_F, \mu_G)), \forall \mu_F, \mu_G \in \mathcal{P}(\mathbb{R}_+) \)?

Or, equivalently, is the inclusion

\[
(\mathcal{P}(\mathbb{R}_+), d_A) \hookrightarrow (\mathcal{P}(\mathbb{R}_+), d_L)
\]
a uniformly continuous function?

The answer is no, since it is not true (in general, and this is the case) that the inversion of a bijective uniformly continuous function is a uniformly continuous function. In fact, let \( \mu_F = \delta_a, \mu_G = \delta_{a+1/2} \). Hence \( d_L(\mu_F, \mu_G) = 1/2, \forall a \geq 0 \). Let \( \varepsilon > 0 \) be fixed, and take \( a \) sufficiently large. Note that

\[
\frac{T(\mu_F)(t)}{\exp(-ta) - (\exp(-t(a + 1/2)) + \varepsilon) \exp(t\varepsilon)} = \frac{t(\varepsilon + 1/2) - \varepsilon(1 + t(\varepsilon + a))}{1 - \exp(-t(\varepsilon + 1/2) - \varepsilon(1 + t(\varepsilon + a)))}
\]

Since \( \exp(x) \geq (1 + x), \forall x \), we have

\[
1 - \exp(-t(\varepsilon + 1/2) - \varepsilon(1 + t(\varepsilon + a))) \leq 1 - (1 - t(\varepsilon + 1/2) - \varepsilon(1 + t(\varepsilon + a))) \\
= t(\varepsilon + 1/2) - \varepsilon(1 + t(\varepsilon + a)) \\
= -\varepsilon + t\varepsilon \left( \frac{1}{2\varepsilon} + 1 - \varepsilon - a \right)
\]

which implies \( d_A(\mu_F, \mu_G) \leq \varepsilon \) if \( a \) is sufficiently large (say, \( a \geq 1/(2\varepsilon) + 1 - \varepsilon \)).

We restrict Problem 3 to a compact subset of \((\mathcal{P}(\mathbb{R}_+), d_L)\) (or, equivalently, \((\mathcal{P}(\mathbb{R}_+), \rightarrow)\)). This is always the case when \( C = (\mathcal{P}(K), d_L) \), where \( K \) is a compact subset of \( \mathbb{R}_+ \), i.e. for bounded distributions (this assumption is often physically correct).

**Proposition 4.** Let \( C \) be a compact subset of \((\mathcal{P}(\mathbb{R}_+), \rightarrow)\).

There exists a function \( f_C : [0, 1] \rightarrow [0, 1] \) s.t.
- \( f_C(t) = 0 \iff t = 0; \)
- \( d_A(\mu_F, \mu_G) \geq f_C(d_L(\mu_F, \mu_G)), \forall \mu_F, \mu_G \in C. \)

Equivalently, the inclusion

\[
(C, d_A) \hookrightarrow (C, d_L)
\]

is a uniformly continuous function.

**Proof.** \((C, d_A)\) is compact. By Theorem 2, \( i^* \) is uniformly continuous. \( \square \)
4. Application to FRET

Let 0 = t_0 < t_1 < \ldots < t_n = T and \{I(i), i = 1, \ldots, n\} be the image intensities as described in the Introduction. Let

\[ \hat{S}(t) = \frac{\sum_{i: t_i > t} I(i)}{\sum_{i=1}^n I(i)} \]

be the observed survival function for the fluorescence emission. \( \hat{S}(t) \) converges uniformly to the survival function \( S(t) \) when the number of sources increases. The very large number of proteins allows us to assume that \( \hat{S}(t_i) \approx S(t_i) \).

Now, take a sequence \( P^{(1)} \subset P^{(2)} \subset P^{(2)} \subset \ldots \) of closed subspaces of \( \mathcal{P}(\mathbb{R}^+) \) and define

\[ \hat{\mu}^{(t_0, t_1, \ldots, t_n)}_{P^{(m)}} := \arg \min_{\mu \in P^{(m)}} \{ d_A(\hat{S}(t), T(\mu)) \}. \] (4)

Note that this computation may be done on the nodes \( t_i \) since \( \hat{S}(t) \) is a stepped function. Eq. (4) may allow more than one solution; each of them will be a \( d_A \)-estimator of \( \mu_R \). Theorem 2 ensures that

\[ \hat{\mu}^{(t_0, t_1, \ldots, t_n)}_{P^{(m)}} \to \min_{\mu \in \bigcup_m P^{(m)}} \mu_R \]

whenever \( \mu_R \in \bigcup_m P^{(m)} \). We could also estimate \( d_L(\hat{\mu}^{(t_0, t_1, \ldots, t_n)}_{P^{(m)}}, \mu_R) \) if we knew “a priori” that \( \mu_R \) belongs to a suitable compact subspace \( C \) of \( \mathcal{P}(\mathbb{R}^+) \) (see Proposition 4).

5. Final remarks

In this work we have introduced a new distance and a new estimator for inverting the Laplace transformation with possible application to FRET. Computational simulations are under development. Two main issues from the numerical point of view are the following: the choice of suitable sequences of subspaces \( P^{(m)} \), and the sample scheduling time related to the optimization problem (4). Numerical analysis and generalization will appear elsewhere, in a forthcoming paper.

Acknowledgements

I wish to thank Prof. G. Toscani and Prof. G. Naldi for stimulating discussions.

References