Algebraic Statistics in non-parametric Information Geometry

Giovanni Pistone

POLITECNICO DI TORINO

Thursday December 18, 2008
An example in Statistical Physics

- $\Omega$ is a finite sample space with $N$ points.
- $U: \Omega \rightarrow \mathbb{R}_{\geq 0}$, $U(x) = 0$ for some $x \in \Omega$, $U \neq 0$.

Gibbs model ... 

$$p(x; \beta) = \frac{e^{-\beta U(x)}}{Z(\beta)}, \quad Z(\beta) = \sum_{x \in \Omega} e^{-\beta U(x)}, \quad \beta > 0.$$ 

- $U$ is the energy, $\beta$ is the inverse temperature, $Z$ is the partition function, $e^{-\beta U}$ is the Boltzmann factor.

... and its limits

As $\beta \rightarrow \infty$, 

$$Z(\beta) \rightarrow \#\{x : U(x) = 0\}, \quad e^{-\beta U(x)} \rightarrow (x : U(x) = 0),$$ 

i.e. the weak limit of $p(\beta)$ as $\beta \rightarrow \infty$ is the uniform distribution on the states $x \in \Omega$ with zero energy.
Canonical variable, extended model

- Changing $U \rightarrow V = (\max U - U)$ and $\beta \rightarrow \theta = -\beta \in \mathbb{R}$ we get the same statistical model presented as an exponential model
  
  $$p(x; \theta) \propto e^{\theta V(x)}$$

- There are weak limits as $\theta \rightarrow \pm \infty$, the limits being the uniform distributions on the set of states that minimize or maximize the $U$ function. Such limits are important in a number of applications, e.g. Statistical Physics or simulation methods in optimization. Therefore, the notion of closed or extended exponential model deserve much attention.

- A generic exponential model based on the canonical statistics $V$ can be written
  
  $$p(x; \theta) = e^{\theta V(x) - \psi(\theta)} \cdot p(x)$$

  where the canonical statistics itself is given up to an affine transformation.

- If a canonical variable is integer valued, we obtain a toric model for the likelihood $p_\theta/p$. 
Information geometry

- The exponential model

\[ p(x; \theta) = e^{\theta V(x) - \psi(\theta)} \cdot p(x) \]

has a number of interesting features such as the strict convexity of the cumulant function \( \psi \) or the relation \( \psi'(\theta) = E_\theta [V] \) which do not depend on the parametrization, but are related with the idea of representing the interior of the probability simplex with an affine space.

- In non-parametric Information Geometry the model is presented with respect to a reference density and the canonical variable is centered,

\[ p(x; \theta) = e^{\theta u(x) - \psi(\theta u)} \cdot p(x; 0), \]

with \( u = \theta(V - E_{p_0} [V]) \) and \( \psi(\theta u) = E_{p_0} [e^{u}] \).

- This idea extends to the representation of a generic strictly positive density \( q \) in the form

\[ q = e^{u - \psi(u)} \cdot p(x) \]

where \( u \) is uniquely determined by the reference density \( p \) and by the condition \( E_p [u] = 0 \).
IG is a family of manifolds on $\Delta$

- From Amari work, we know that there are many (differential) geometries on the simplex of probability densities of a given sample space $(\Omega, \mathcal{F}, \mu)$.
- Let $\mathcal{M}_>$ denote the set of all positive densities of $(\Omega, \mathcal{F}, \mu)$. For each $p \in \mathcal{M}_>$ the mapping $s_p : q \mapsto u$ is a chart. The atlas $(s_p)$ defines the **e-manifold**.
- The atlas of the charts $q \mapsto q/p - 1$ defines the **m-manifold**.
- According Amari, in between the e-manifold and the m-manifold there are other differential structures associated with the charts
  \[
  q \mapsto \frac{(q/p)^\lambda - 1}{\lambda}
  \]
  However, $\lambda^{-1}((q/p)^\lambda - 1)$ is bounded below by $-\lambda^{-1}$.
- Here, we discuss the construction of such geometries and their algebraic counterpart in the form of a generalization of the exponential case.

\(\kappa\)-exponential

G. Kaniadakis, based on arguments from Statistical Physics and Special Relativity, has defined the \(\kappa\)-deformed exponential for each \(x \in \mathbb{R}\) and \(-1 < \kappa < 1\) to be

\[
\exp_\kappa(x) = \exp \left( \int_0^x \frac{du}{\sqrt{1 + \kappa^2 u^2}} \right).
\]

Note the special cases

\[
\exp_\kappa(x) = \begin{cases} 
\left( \kappa x + \sqrt{1 + \kappa^2 x^2} \right)^{\frac{1}{\kappa}}, & \text{if } \kappa \neq 0, \\
\exp x, & \text{if } \kappa = 0,
\end{cases}
\]

and the \(\kappa\)-deformed logarithm defined for \(y > 0\) by

\[
\ln_\kappa(y) = \begin{cases} 
\frac{y^\kappa - y^{-\kappa}}{2\kappa}, & \text{if } \kappa \neq 0, \\
\ln y, & \text{if } \kappa = 0.
\end{cases}
\]

Among all possible approximations to $\exp$, this particular one has been selected by Kaniadakis because it is the simplest with the property

$$\exp_\kappa(x) \exp_\kappa(-x) = 1$$

- For $\kappa \neq 0$, the indeterminate $y = (\exp_\kappa(x))^\kappa$ and $x$ are related by the polynomial equation

$$y^2 - 2\kappa xy - 1 = 0$$

- Therefore, the graph of $(\exp_\kappa)^\kappa$ is the upper branch of a hyperbola.
\( \kappa \)-deformed operations

- The function \( \exp_\kappa \) maps \( \mathbb{R} \) unto \( \mathbb{R}_\succ \), it is strictly increasing and it is strictly convex.
- The function \( \ln_\kappa \) maps \( \mathbb{R}_\succ \) unto \( \mathbb{R} \), is strictly increasing and is strictly concave.
- Both the \( \kappa \)-deformed exponential and the \( \kappa \)-deformed functions \( \exp_\kappa \) and \( \ln_\kappa \) reduce to the ordinary \( \exp \) and \( \ln \) functions when \( \kappa \rightarrow 0 \).
- Group operations \((\mathbb{R}, \oplus)\) and \((\mathbb{R}_\succ, \otimes)\) are defined in such a way that \( \exp_\kappa \) is a group isomorphism from \((\mathbb{R}, +)\) onto \((\mathbb{R}_\succ, \otimes)\) and also from \((\mathbb{R}, \oplus)\) onto \((\mathbb{R}_\succ, \times)\):

\[
\exp_\kappa (x_1 + x_2) = \exp_\kappa (x_1) \otimes \exp_\kappa (x_2),
\]
\[
\exp_\kappa \left( x_1 \oplus x_2 \right) = \exp_\kappa (x_1) \exp_\kappa (x_2).
\]
The algebra of $\exp_\kappa$ and $\ln_\kappa$

1. The binary operations $\kappa \oplus$ and $\kappa \otimes$ are defined by
   \[
   x_1 \kappa \oplus x_2 = \ln_\kappa (\exp_\kappa (x_1) \exp_\kappa (x_2))
   \]
   \[
   y_1 \kappa \otimes y_2 = \exp_\kappa (\ln_\kappa (y_1) + \ln_\kappa (y_2))
   \]

2. The operation $\kappa \otimes$ is defined on positive reals. However, $\kappa \otimes$ can be extended by continuity to non-negative reals in such a way that
   \[
   0 \kappa \oplus y = y \kappa \oplus 0 = 0 \kappa \oplus 0 = 0
   \]

3. We want to derive defining relations for the $\kappa$-deformed operations in the form of a polynomial. This is obtained by repeated use of the HYP. Symbolic computations have been done with CoCoA.

We want to find \( x \) such that \( \exp_\kappa(x) = \exp_\kappa(x_1) \exp_\kappa(x_2) \).

From \( y_1 = (\exp_\kappa(x_1))^\kappa \), \( y_2 = (\exp_\kappa(x_2))^\kappa \) and

\[
(\exp_\kappa(x))^\kappa = (\exp_\kappa(x_1) \exp_\kappa(x_2))^\kappa = y_1 y_2,
\]
we have the ideal generated by

\[\text{Eq1} := y[1]^2 - 2\kappa x[1] y[1] - 1;\]

Elimination of \( y_1, y_2 \) gives the polynomial equation

\[
x^4 - 2 \left(2\kappa^2 x_1^2 x_2^2 + x_1^2 + x_2^2\right) x^2 + \left(x_1^2 - x_2^2\right)^2 = 0,
\]

whose solution is

\[
x_1^{\kappa} \oplus x_2 = x_1 \sqrt{1 + \kappa^2 x_2^2} + x_2 \sqrt{1 + \kappa^2 x_1^2}.
\]

Kaniadakis has a relativistic interpretation.
We want to find $z = \left( y_1 \otimes y_2 \right)^\kappa$. Let $y_1 = (\exp_\kappa (x_1))^\kappa$, $y_2 = (\exp_\kappa (x_2))^\kappa$, and $z = (\exp_\kappa (x_1 + x_2))^\kappa$.

Equation HYP gives three quadratic equations in the indeterminates $x_1, x_2, y_1, y_2, z, \kappa$. Elimination of $x_1, x_2$ gives the polynomial equation

$$y_1y_2z^2 + (1 - y_1y_2)(y_1 + y_2)z - y_1y_2 = 0.$$

It is remarkable that this equation does not depend on $\kappa$. An explicit solution is obtained by solving the quadratic equation.

A possibly more suggestive solution is obtained as follows. First, we reduce to the monic equation

$$z^2 + \left( 1 - \frac{1}{y_1y_2} \right)(y_1 + y_2)z - 1 = 0$$

and denote the two solutions by $z > 0$ and $-1/z$. Therefore,

$$z - \frac{1}{z} = \left( y_1 - \frac{1}{y_1} \right) + \left( y_2 - \frac{1}{y_2} \right)$$
The $\kappa$-logarithm is strictly related to a family of transformation which is well known in Statistics under the name of Box-Cox transformation or power transform. For data vector $y_1, \ldots, y_n$ in which each $y_i > 0$, the power transform is:

$$y_i^{(\lambda)} \propto \frac{y_i^\lambda - 1}{\lambda}$$

The same transformation, applied to probability densities, appears in Amari as a device to construct Statistical Manifolds.

Tsallis has applied the transformation in non-extensive thermodynamics.

Naudts discusses the applications of $\ln_\kappa$ and $\exp_\kappa$ in Information Theory and Statistical Physics.

Kaniadakis’s $\kappa$-deformed logarithm $x = \ln_\kappa(y)$ has the extra feature of the symmetry induced by the term $-y^{-\kappa}$.

On a finite state space $\Omega$, equipped with the energy function $U: \Omega \to \mathbb{R}_\geq$, we want to discuss the $\kappa$-deformation of the standard Gibbs model. There are two options, related with two different presentation of the normalizing constant (partition function).

The first option is to consider the statistical model

$$p(x; \theta) = \frac{\exp_\kappa (\theta U(x))}{Z(\theta)}$$

$$= \exp_\kappa \left( \theta U(x) \oplus \ln_\kappa \left( \frac{1}{Z(\theta)} \right) \right)$$

The $\ln_\kappa$-model is, with $\tilde{\psi}_\kappa(\theta) = \ln_\kappa Z(\theta)$,

$$\ln_\kappa p(x; \theta) = \theta U(x) \sqrt{1 + \kappa^2 (\tilde{\psi}_\kappa(\theta))^2} - \tilde{\psi}_\kappa(\theta) \sqrt{1 + \kappa^2 \theta^2 U(x)^2}$$
The second option is to define the model as

\[ p(x; \theta) = \exp_\kappa (\theta U(x) - \psi_\kappa(\theta)) \]

\[ = \exp_\kappa (\theta U(x)) \otimes \exp_\kappa (-\psi_\kappa(\theta)), \]

where \( \psi_\kappa(\theta) \) is the unique solution of the equation

\[ \sum_{x \in \Omega} \exp_\kappa (\theta U(x) - \psi_\kappa(\theta)) = 1. \]

The derivative with respect to \( \theta \) of \( \psi_\kappa \) is given by

\[ E_\theta \left[ \frac{U - \psi_\kappa'(\theta)}{\sqrt{1 + \kappa^2 (\theta U - \psi_\kappa(\theta))^2}} \right] = 0, \]

where \( E_\theta [V] = \sum_x V(x)p(x; \theta) \).
The two one-parameter statistical models are different unless $\kappa = 0$. This fact marks an important difference between the theory of ordinary exponential models and $\kappa$-deformed exponential models.

From the geometrical point of view, the second approach has the advantage of a the linear character of the model describing the $\ln \kappa$-probability.

Let $V = \text{Span} (1, U)$ and $V^\perp$ the orthogonal space, i.e. $v \in V^\perp$ if, and only if, $\sum_x v(x) = 0$ and $\sum_x v(x) U(x) = 0$. Therefore,

$$\sum_{x \in \Omega} v(x) \ln_{\kappa} (p(x; \theta)) = 0, \quad v \in V^\perp$$

Viceversa, if a strictly positive probability density function $p$ is such that $\ln_{\kappa} p$ is orthogonal to $V^\perp$, then $p$ belongs to the $\kappa$-Gibbs model for some $\theta$. 
\( \kappa \)-toric

- For each \( v \in V^\perp \),
  \[
  \sum_{x: \; v(x) > 0} v^+(x) \ln_\kappa (p(x)) = \sum_{x: \; v(x) < 0} v^-(x) \ln_\kappa (p(x)).
  \]

- A (physical) interpretation: a positive density \( p \) belongs to the \( \kappa \)-Gibbs model if, and only if,
  \[
  E_{r_1} [\ln_\kappa (p)] = E_{r_2} [\ln_\kappa (p)]
  \]
  for each couple of densities \( r_1, r_2 \) such that \( r_1 r_2 = 0 \) and \( E_{r_1} [U] = E_{r_2} [U] \).

If \( v \in V^\perp \) happens to be integer valued, using the \( \kappa \)-algebra and the notation \( x \otimes \cdots \otimes x = x \otimes^n \), we can write

\[
\bigotimes_{x: \; v(x) > 0} p(x)^\kappa v^+(x) = \bigotimes_{x: \; v(x) < 0} p(x)^\kappa v^-(x),
\]
The binomial equations are

\[
\begin{cases}
p(1) = p(2) \\
p(4) = p(5) \\
p(1)^\kappa \otimes p(2)^\kappa \otimes p(4)^\kappa \otimes p(5)^\kappa = p(3)^\kappa \otimes 4
\end{cases}
\]

A non strictly positive density that is a solution is either

\[
p(1) = p(2) = p(3) = 0, \quad p(4) = p(5) = 1/2, \quad \text{or} \quad p(1) = p(2) = 1/2, \quad p(3) = p(4) = p(5) = 0.
\]

These two solutions are the uniform distributions on the sets of values that respectively maximize or minimize the energy function.
A further algebraic presentation is available. Consider the new parameters

$$\zeta_0 = \exp_\kappa (-\psi_\kappa (\theta)) , \quad \zeta_1 = \exp_\kappa (\theta) ,$$

so that

$$p(x; \theta) = \exp_\kappa (\theta U(x)) \otimes \exp_\kappa (-\psi_\kappa (\theta)) ,$$

$$= \zeta_0 ^\kappa \otimes \zeta_1 ^\kappa U(x) .$$

The probabilities are $\kappa$-monomials in the parameters $\zeta_0, \zeta_1$, e.g.:

$$\begin{cases} 
  p(1) = p(2) = \zeta_0 \\
  p(3) = \zeta_0 ^\kappa \otimes \zeta_1 \\
  p(4) = p(5) = \zeta_0 ^\kappa \otimes \zeta_1 ^{\kappa 2}
\end{cases}$$

Note that the parameter $\zeta_0$ is required to be strictly positive, while the parameter $\zeta_1$ could be zero, giving rise the uniform distribution on \{1, 2\} = \{x: U(x) = 0\}. The other limit solution is not obtained.
If $\kappa \neq 0$ the last equation of the system

$$\begin{cases} p(1) = p(2) \\ p(4) = p(5) \\ p(1) \kappa \otimes p(2) \kappa \otimes p(4) \kappa \otimes p(5) = p(3) \kappa^4 \end{cases}$$

can be written as

$$\left( p^\kappa(1) - \frac{1}{p^\kappa(1)} \right) + \left( p^\kappa(2) - \frac{1}{p^\kappa(2)} \right) +$$

$$\left( p^\kappa(4) - \frac{1}{p^\kappa(4)} \right) + \left( p^\kappa(5) - \frac{1}{p^\kappa(5)} \right) =$$

$$4 \left( p^\kappa(3) - \frac{1}{p^\kappa(3)} \right)$$

**Question**

Is $\kappa \to 0$ a proper “approximation” of the regular case $\kappa = 0$?
To construct an atlas, we define each chart as associated to a strictly positive probability densities. Such a density $p$ is a reference for each other density $q$ via the notion of likelihood $q/p$.

**Definition**

Fix a $\kappa \in ]0, 1[$. Given positive density functions $q$ and $p$ such that $(q/p), (p/q) \in L^{1/\kappa}(q)$, i.e. $(q/p)^{\kappa}, (p/q)^{\kappa} \in L^1(q)$, the $\kappa$-divergence is

$$D_\kappa(q\|p) = E_q \left[ \ln_\kappa \left( \frac{q}{p} \right) \right] = \frac{1}{2\kappa} E_q \left[ \left( \frac{q}{p} \right)^{\kappa} - \left( \frac{p}{q} \right)^{\kappa} \right].$$

The strict convexity of $-\ln_\kappa$ implies

$$D_\kappa(q\|p) = E_q \left[ -\ln_\kappa \left( \frac{p}{q} \right) \right] \geq -\ln_\kappa \left( E_q \left[ \frac{p}{q} \right] \right) = \ln_\kappa (1) = 0.$$

with equality if, and only if $q = p$. 
exp_\kappa densities

Definition?

\[ \mathcal{E}_p = \left\{ q \in \mathcal{M}_> : \left(\frac{q}{p}\right)^\kappa, \left(\frac{p}{q}\right)^\kappa \in L^{1/\kappa}(p) \right\} \]

\[ = \left\{ q \in \mathcal{M}_> : \frac{q}{p}, \frac{p}{q} \in L^1(p) \right\} = \left\{ q \in \mathcal{M}_> : \frac{p}{q} \in L^1(p) \right\} \]

- The divergence \( D_\kappa(p\|q) \) is defined on \( \mathcal{E}_p \).
- If \( q \in \mathcal{E}_p \), then \( q \) is almost surely positive and we can write it in the form \( q = \exp_\kappa(v) \cdot p \), with

\[ v = \ln_\kappa \left(\frac{q}{p}\right) = \frac{\left(\frac{q}{p}\right)^\kappa - \left(\frac{p}{q}\right)^\kappa}{2\kappa} \in L^{1/\kappa}(p) \]
The expected value at $p$ of $v = \ln_{\kappa} \left( \frac{q}{p} \right)$ is $E_p \left[ \ln_{\kappa} \left( \frac{q}{p} \right) \right] = -D_{\kappa}(p\|q)$ so that we can write every $q \in \mathcal{E}_p$ as

$$q = \exp_{\kappa} \left( u - D_{\kappa}(p\|q) \right) \cdot p$$

where $u$ is a uniquely defined element of the set of $p$-centered $1/\kappa$-integrable random variables $L_{0}^{1/\kappa}(p)$.

Vice versa, given $u \in L_{0}^{1/\kappa}(p)$, the real function $\psi \mapsto E_p [\exp_{\kappa} (u - \psi)]$ is continuous and strictly decreasing from $+\infty$ to 0, therefore there exists a unique $\psi_{\kappa,p}(u)$ such that

$$q = \exp_{\kappa} \left( u - \psi_{\kappa,p}(u) \right) \cdot p \in \mathcal{E}_p \subset \mathcal{M}_>$$
Assume now we want to change of chart, that is we want to change the reference density from $p_1$ to $p_2$ to represent a $q$ that belongs to both $\mathcal{E}_{p_1}$ and $\mathcal{E}_{p_2}$. The formal application of the chart and the patch formulæ gives

$$u_2 = \ln_\kappa \left( \frac{q}{p_2} \right) - \mathbb{E}_{p_2} \left[ \ln_\kappa \left( \frac{q}{p_2} \right) \right]$$

$$= \ln_\kappa \left( \exp_\kappa (u_1 - \psi_{\kappa,p_1}(u_1)) \frac{p_1}{p_2} \right) - \mathbb{E}_{p_2} \left[ \cdots \right]$$

$$= (u_1 - \psi_{\kappa,p_1}(u_1)) \oplus \ln_\kappa \left( \frac{p_1}{p_2} \right) - \mathbb{E}_{p_2} \left[ \cdots \right]$$

- Question: Is the set of $u$'s such that $\exp_\kappa (u - \psi_{\kappa,p_1}) \cdot p_1$ belongs to $\mathcal{E}_{p_1}$ an open set of $L^{1/\kappa}_o(p)$?
- Problem: compute the Fréchet derivative of the change of coordinate.
- Problem: compute the connections.
Tangent vectors

Let $p_\theta, \theta \in ]0, 1[,$ be a curve in $\mathcal{E}_p,$

$$p_\theta = \exp_\kappa (u_\theta - \psi_{\kappa,p}(u_\theta)) \cdot p.$$ 

In the chart at $p$ the velocity vector is given by

$$\dot{u}_\theta \in L^{1/\kappa}_0(p) = T_{\kappa,p}$$

Formal computation gives

$$\frac{\dot{p}_\theta}{p_\theta} = (1 + \kappa^2(u_\theta - \psi_{\kappa,p}(u_\theta)^2)^{-1/2}(u_\theta - D_{u_\theta}\psi_{\kappa,p}(u_\theta))$$

so that

$$\frac{\dot{p}_0}{p_0} = \dot{u}_0$$
Conclusion

- Amari tells us that each probability simplex $\Delta$ supports $\kappa$-statistical manifolds, one for each $\kappa \in [0, 1]$.
- Each $\kappa$ has peculiar algebraic features.
- All $\kappa$-manifolds are possibly deduced from the same template, i.e. the exponential model (work in progress).
- There are domains of application of the algebro-geometric picture not yet explored:
  - Statistical Physics,
  - Optimization,
  - Differential equations for probability densities,
  - Approximation of statistical models.

THANKS