Associated exponential families and elliptic functions

Gérard Letac, Toulouse

June 18 th, 2015, Aarhus

Meeting Ole, meeting the exponential families

I remember the very long young man who came to Montreal 1968 for talking about the not yet fashionable subject of exponential families in Statistics. A real meeting took place in Saint Flour 1986 for the Ole's lectures. Seshadri had converted me to statistics and the exponential families before, and I had already fought with the Ole's book, with the steepness and with the cuts.

Ole, Toulouse and the students

After Saint Flour, invitations to Aarhus came. Over the years, Ole created several European networks. We were meeting regularly in Toulouse, and many doctorate students from my university undertook a fruitful pilgrimage to Aarhus: Evelyne Bernadac, Célestin Kokonendji, Dhafer Malouche, Muriel Casalis, Abdelhamid Hassairi, Angelo Koudou. I have always been admirative for the time and the interest that Ole has taken in the training of young people in research. A doctorate honoris causa for Ole at Toulouse in 1995 was a recognition of his involvement, and I cannot tell how grateful all of us are of these Aarhus days, which were also giving the opportunity to meet all these wonderful Danish statisticians: Jens Jensen, Swante Jensen, Eva Jensen, Praeben Blaesild, Steffen Lauritzen, S. Asmussen, M. Sørensen, or Bent Jørgensen to which I am specially indebted.

The natural exponential families

A real natural exponential family (NEF) is the data of a non Dirac positive measure μ on $\mathbb{R}$ such that the Laplace transform $L_\mu(\theta) = \int_{\mathbb{R}} e^{\theta x}\mu(dx)$ is finite on an interval $D(\mu)$ with non empty interior $\Theta(\mu)$. Call $\mathcal{M}(\mathbb{R})$ the set of these measures. The natural exponential family generated by $\mu \in \mathcal{M}(\mathbb{R})$ is the set of probabilities

$$F(\mu) = \{P(\theta, \mu)(dx) = e^{\theta x}\frac{\mu(dx)}{L_\mu(\theta)}; \theta \in \Theta(\mu)\}$$

Mean and variance function of a NEF

If $k_\mu = \log L_\mu$ is the cumulant function, then $\theta \mapsto k_\mu(\theta)$ is strictly convex on the open interval $\Theta(\mu)$ and

$$\theta \mapsto m = k'_\mu(\theta) = \int_{\mathbb{R}} xP(\theta, \mu)(dx)$$

is one-to-one from $\Theta(\mu)$ to some interval M_F called the domain of the means of $F = F(\mu)$. Its inverse fonction $m \mapsto \theta = \psi_\mu(m)$ from M_F to $\Theta(\mu)$ is quite useful. The variance of $P(\theta, \mu)(dx)$ is

$$k''_\mu(\theta) = k''_\mu(\psi_\mu(m)) = \frac{1}{\psi'(m)} = V_F(m).$$

Ole does not like that much the variance functions

The function $m \mapsto V_F(m)$ defined on M_F characterises F. This elegant fact prompted us to consider the study of the variance function as one of the fine arts, while Ole with some reason thought it is lacking of practical applications.
But, in a provocative way, I am going to talk about the variance functions of the form

$$V_F(m) = (\alpha m + \beta)\sqrt{P(m)}$$

when P is a polynomial of degree ≤ 4.

One needs to know a bit on the transformations of variance functions

If V_F is a variance function, then the image $h(F)$ of F by $x \mapsto h(x) = ax + b$ is a NEF such that

$$\boxed{V_{h(F)}(m) = a^2 V_F(\frac{m-b}{a})}$$

If $\lambda > 0$ the function L_μ^λ may be or may be not the Laplace transform of some $\mu_\lambda \in \mathcal{M}(\mathbb{R})$. The set $\Lambda(\mu)$ of the λ's such that μ_λ does exist is called the Jørgensen set of μ and we have for $\lambda \in \Lambda(\mu)$

$$\boxed{V_{F(\mu_\lambda)}(m) = \lambda V_F(\frac{m}{\lambda})}$$

Reciprocity and Moebius transformations

A more complicated transformation is reciprocity, involving a new exponential family F_1 built from F in a fascinating probabilistic way and satisfying on a proper interval

$$V_{F_1}(m) = m^3 V_F(\frac{1}{m}).$$

Combining with the affine transformations of the previous slide with the Moebius transform $m \mapsto 1/m$ we arrive at transformations $h(m) = \frac{am+b}{cm+d}$ and new NEF F_1 related by h to F by

$$\boxed{V_{F_1}(m) = (cm+d)^3 V_F(\frac{am+b}{cm+d})}.$$

Geometric interpretation

While a probabilistic interpretation of these transformations $F \mapsto F_1$ via a Moebius transformation $h(m) = \frac{am+b}{cm+d}$ is generally possible, the following geometric interpretation is easier:

- Draw the convex curve C of equation $m = k_\mu(\theta)$ in the (θ, m) plane.
- Make an affine transformation (linked to the Moebius transformation) of this (θ, m) plane, obtaining a new convex curve C_1.
- Select a segment I of the axis and an arc of C_1 which is the graph of some convex function f defined on I.
- In the favorable cases there exists $\mu_1 \in \mathcal{M}(\mathbb{R})$ such that $k_{\mu_1}(\theta) = f(\theta)$ for all $\theta \in I$.

Example: hitting time of a drifted Brownian motion

The most famous example is $k_\mu(\theta) = \theta^2/2$; the Moebius transform $m \mapsto 1/m$ linked with the affine transformation of the plane is the symmetry with respect to the line $m + \theta = 0$. We get $I = (-\infty, 0)$ and the cumulant transform $k_{\mu_1}(\theta) = -\sqrt{-2\theta}$ of the positive stable law of parameter $1/2$. This leads to show that the hitting time of $x > 0$ by $B(t) - at$ is an inverse Gaussian distribution, ie a member of the exponential family generated by this μ_1.

Families preserved by Moebius: cubic and elliptic

If you apply these Moebius transformations $V_{F_1}(m) = (cm+d)^3 V_F(\frac{am+b}{cm+d})$ to the family $\mathcal{F}_3$ of all variance functions which are polynomials of degree ≤ 3 you observe that $\mathcal{F}_3$ is preserved. It is not difficult to see that $\mathcal{F}_3$ is splitted in 4 orbits: the Normal -Inverse Gaussian orbit, the Poisson Gamma orbit, the hyperbolic orbit and the Kendall-Ressel orbit. If $\mathcal{F}_e$ is the elliptic family of all variance functions of the form $(\alpha m + \beta)\sqrt{P(m)}$ where $\deg P \leq 4$, this is also a stable family by Moebius. Isolating its orbits is not difficult, but deciding of the content of the orbits is a more difficult task.

Only $V(m) = \sqrt{Am^4 + Bm^2 + C}$

(But $V_F(m) = m\sqrt{Am^4 + Bm^2 + C}$ is less difficult). We have no time to discuss the easy cases $A = 0, C = 0$ and $B^2 - 4AC = 0$ and to explain why $Ax^2 + Bx + C$ cannot have simple positive roots if $\sqrt{Am^4 + Bm^2 + C}$ is a variance function. Without loss of generality we may assume that $C = 1$ (by doing a dilation).

The three cases to discuss

By high school algebra $Ax^2 + Bx + 1$ is positive on $(0, \infty)$ if and only if there exists $a \in \mathbb{R}$ and $b > 0$ such that

$$Ax^2 + Bx + 1 = (1 + ax)^2 + 2b^2$$

Finally we introduce a complex number k such that

$$k^2 = 1 + \frac{2a}{b^2}$$

(the square for k is for fitting to the traditional notations of elliptic functions)

1. Either $-1 \leq k^2 < 0$,
2. or $k^2 < -1$,
3. or $0 < k^2$.

The case $-1 \leq k^2 < 0$

We write $k^2 = -1 + p$ with $0 \leq p < 1$ and we introduce the following two constants:

$$\begin{aligned} K &= \int_0^1 (1-x^2)^{-1/2}(2-p-x^2)^{-1/2}dx \\ K' &= \int_0^1 (1-x^2)^{-1/2}(1+(1-p)x^2)^{-1/2}dx. \end{aligned} \tag{1}$$

Using a Jorgensen transformation we assume $b^2 = 2$

$\sqrt{(1+a\frac{m^2}{t^2})^2+4\frac{m^2}{t^2}}$ (case $-1 \leq k^2 < 0$ continued)

Theorem. There exists a natural exponential family G_t with domain of the means $\mathbb{R}$ and variance function

$$t\sqrt{(1+a\frac{m^2}{t^2})^2+4\frac{m^2}{t^2}}$$

when t is a multiple of a. It is concentrated on $\frac{\pi}{2K}\mathbb{Z}$.
The family $G_{|a|}$ is generated by a symmetric probability measure $\mu_{|a|}$ which is the convolution of the Bernoulli distribution $\frac{1}{2}(\delta_{-\frac{\pi}{2K}}+\delta_{\frac{\pi}{2K}})$ by an infinitely divisible distribution $\alpha_{|a|}$ concentrated on $\frac{\pi}{K}\mathbb{Z}$.

A description of α (case $-1 \leq k^2 < 0$ continued)

We denote $q = e^{-\pi K'/K}$ and for a positive integer ν we denote

$$c_\nu = c_{-\nu} = \frac{q^\nu - (-1)^\nu q^{2\nu}}{1 - q^{2\nu}} > 0.$$

Then the Laplace transform of α_t is

$$\int_{-\infty}^{\infty} e^{\theta x} \alpha_t(dx) = \exp\left(\frac{t}{|a|} \sum_{\nu \in \mathbb{Z} \setminus \{0\}} c_\nu (e^{\frac{\nu \pi \theta}{K}} - 1) \right).$$

The characteristic function of μ (case $-1 \leq k^2 < 0$ continued)

Finally the characteristic function of $\mu_{2|a|}$ is $\frac{1}{\wp(s+K)-\frac{p}{3}}$ where $\wp$ is the elliptic Weierstrass function satisfying

$$\wp'^2 = 4(\wp - 1 + \frac{2p}{3})(\wp - \frac{p}{3})(\wp + 1 - \frac{p}{3})$$

which is doubly periodic with primitive periods $2K$ and $2iK'$. In particular it has zeros and G_t cannot be infinitely divisible.

$$\begin{tikzcd}
2iK' \arrow[r, dash] \arrow[d, dash] & K+2iK' \arrow[r, dash] \arrow[d, dash] & 2K+2iK' \arrow[d, dash] \\
iK' \arrow[r, dash] \arrow[d, dash] & K+iK' \arrow[r, dash] \arrow[d, dash] & 2K+iK' \arrow[d, dash] \\
0 \arrow[r, dash] & K \arrow[r, dash] & 2K
\end{tikzcd}$$

Case $k^2 < -1$: using associated NEF

These variance functions do not exist. The proof uses the idea of pair of associated families. Let μ and μ_1 be two symmetric probabilities in $\mathcal{M}(\mathbb{R})$. They are associated if

$$\int_{\mathbb{R}} e^{i\theta x}\mu_1(dx) = \frac{1}{L_\mu(\theta)}.$$

The most celebrated case is $\mu(dx) = \frac{dx}{2\cosh\frac{\pi x}{2}}$ and $\mu_1 = \frac{1}{2}(\delta_{-1} + \delta_1)$.
If μ and μ_1 are an associated pair, the two NEF F and F_1 that they generate are said to be associated, and we have

$$V_F(m) = V_{F_1}(im)$$

in a proper neighborhood of $0 \in \mathbb{C}$ (Variance functions are real analytic and have extensions to parts of $\mathbb{C}$.

Continuation $k^2 < -1$

If F is such that $V_F(m) = \sqrt{Am^4 + Bm^2 + 1}$ with $-1 < k^2 < 0$ and if F_1 is associated to F then $V_{F_1}(m) = \sqrt{Am^4 - Bm^2 + 1}$ with $k_1^2 = 1/k^2 < -1$. Analysis then shows that if μ_1 such that $F_1 = F(\mu_1)$ exists then L_{μ_1} is both a Laplace transform and the Fourier transform of something else: a contradiction.

The case $k^2 > 0$

This case corresponds to the study of $V_F(m) = \sqrt{(m^2+a^2)(m^2+b^2)}$ where $0 < a < b$. Here again, it is concentrated on a multiple of the relative integers. The study is close to the case $k^2 \in (-1, 0)$.

Something about $m\sqrt{1+Bm^2+Am^4}$

This NEF F is the reciprocal NEF of the NEF with variance function $\sqrt{m^4+Bm^2+A}$ which is concentrated on a multiple of $\mathbb{Z}$. Since it is analytic in 0 and equivalent to m it is concentrated on the set of non negative integers $\mathbb{N}$. Let us give details in a particular case:

Theorem Let $x > 0$. The NEF F_x with domain of the means $(0, \infty)$ and variance function

$$V_{F_x}(m) = m(1 + \frac{4m^4}{x^4})^{1/2}$$

is generated by a positive measure on $\mathbb{N}$ which is $\nu_x(dt) = \sum_{n=0}^{\infty} \frac{p_n(x)}{n!}\delta_n(dt)$ with generating function

$$f_x(z) = \sum_{n=0}^{\infty} \frac{p_n(x)}{n!} z^n = e^{x \int_0^z \frac{dw}{(1-w^4)^{1/2}}}$$

which satisfies

$$(1 - z^4) f_x''(z) - 2z^3 f_x'(z) - x^2 f_x(z) = 0.$$

The total mass of ν_x is $\exp(x\frac{1}{4}B(\frac{1}{2}, \frac{1}{4}))$.

Explicitation of the polynomials)

When

$$f_x(z) = \sum_{n=0}^{\infty} \frac{p_n(x)}{n!} z^n = e^{x \int_0^z \frac{dw}{(1-w^4)^{1/2}}}$$

then the polynomials p_n are given by $p_n(x) = x^n$ for $n = 0, 1, 2, 3, 4,\ p_5(x) = x^5 + 12x$ and for $n \geq 2$

$$p_{n+2}(x) = x^2 p_n(x) + n(n-1)^2(n-2)p_{n-2}(x).$$